Intervention evaluation and structural identifiability in compartmental models of infectious disease transmission



A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Trinity Term 2022

Emmanuelle Afaribea Dankwa

St Peter's College

Acknowledgements

I am very grateful to my supervisor, Prof. Christl Donnelly, for coaching and supporting me wonderfully these three-and-half years, not only on my DPhil projects but also on career development and more. It's been a true pleasure to know and work with Christl.

I thank Prof. Deirdre Hollingsworth, Prof. David Cox (of blessed memory), Prof. Seth Flaxman, Dr. Alex Villaverde, Dr. Ben Lambert and anonymous referees for feedback and useful discussions which helped to improve this work.

I am also grateful to the researchers I have had the privilege of collaborating with, and from whom I have learned much: Prof. Natasha Martin (UC San Diego), Dr. Andrew Brouwer (Univ. of Michigan), Rui Zhao (Louisville Dept. of Public Health), Dr. Martha Montgomery, Dr. Mark Weng (US Centers for Disease Control), Dr. Robin Thompson (Univ. of Warwick), Dr. Sarah Hayes (Univ. of Oxford) and Dr. Sébastien Lambert (Univ. of Toulouse).

I am especially grateful to Dr. Sébastien Lambert and the staff at the Department of Statistics for generous technical support.

For their great peer support, I thank my research group – Team Donnelly – and the Room 2.08 office squad: Ruth, Jessie and Zoe.

My time at St Peter's College has been rewarding and I am grateful to fellow Peterites and staff for making St Peter's the warm community it is.

I appreciate the Rhodes West Africa Selection Committee, Mr. Robert Afulimi, Dr. Ralph Twum and Madam Gloria Botchway for their immense support on this journey. The Rhodes Trust has been phenomenal not only in providing financial support but also care, a wonderful scholar community and many opportunities for growth. I also thank the Department of Statistics for the studentship award which sponsored my fourth year.

I am thankful to family at Oxford Community Church for their encouragement and love. I am especially grateful to the youth team, Nana Amoasi, Ree Masih and Sam Masih for keeping the doors of their homes and hearts open to me.

I thank my family, mentors and friends for their continuous support and care throughout the DPhil, and for many of them, despite a separation of at least 5000 miles.

Finally, to my beloved Friend and King, Jesus – I am grateful for all these.

Table of Contents

Chapter 1 Introduction 1
1.1 Background 1
1.2 Motivation for this work 4
1.3 Objectives and contributions 6
1.4 Reproducibility 9
1.5 Thesis outline 9
Chapter 2 Basic compartmental models for infectious disease transmission 11
2.1 Basic compartmental models 11
2.1.1 A deterministic SIR model 11
2.1.2 A stochastic SIR model 12
2.1.3 Deterministic versus stochastic compartmental models 13
2.1.4 Extensions to the SIR model 14
2.2 Key epidemiological quantities 16
2.2.1 The basic reproduction number 16
2.2.2 Herd immunity threshold and critical vaccination coverage 18
2.3 Parameter estimation 19
2.3.1 Maximum likelihood estimation 19
2.3.2 Bayesian inference 20
2.4 Approaches to structural identifiability analysis in ODE models 22
2.4.1 General modelling framework and definitions 23
2.4.2 Approaches to testing structural global identifiability 25
2.4.3 Software toolboxes 30
Chapter 3 Estimating vaccination threshold and impact in the 2017–2019 hepatitis A virus outbreak among persons experiencing homelessness or who use drugs in Louisville, Kentucky, United States 31
Abstract 32
3.1 Introduction 34
3.2 Methods 35
3.2.1 Data 35
3.2.2 Analysis of risks among sub-populations 36
3.2.3 Epidemic modelling 37
3.3 Results 43
3.3.1 Analysis of risks among subpopulations 43
3.3.2 Transmission model results 44
3.4 Discussion 45
Tables 50
Figures 54

Supplementary material 57
S1: Relative risk analysis 57
S2: The model 61
S3: Estimation of population size of PEH/PWUD in Louisville 64
S4: Derivation of R_0 65
S5: Sensitivity analysis 66
Chapter 4 Stochastic modelling of African swine fever in wild boar and domestic pigs: epidemic forecasting and comparison of disease management strategies 71
Abstract 71
4.1 Introduction 72
4.1.1 Background to ASF 73
4.1.2 Current situation in Europe 74
4.1.3 Mathematical Models 75
4.1.4 Challenge overview and objectives 76
4.2 Materials and methods 78
4.2.1 Synthetic data provided by the challenge coordinators 78
4.2.2 Model 80
4.2.3 Analysis considerations 91
4.2.4 Parameter estimation 91
4.2.5 Simulations and outputs 92
4.2.6 Sensitivity analysis 93
4.3 Results 94
4.3.1 Parameter estimation and model fit 94
4.3.2 ASF management strategies 95
4.3.3 Probability of epidemic fade-out by day 230 96
4.3.4 Sensitivity analysis 97
4.4 Discussion 98
4.4.1 What do our results show and what do they mean? 99
4.4.2 What were the main drivers of model dynamics? 104
4.4.3 What challenges did we face? 106
4.4.4 How could the modelling approach/choices be improved? 106
4.4.5 How can the projections be improved? 108
4.4.6 Comparison to previous modelling studies 108
4.5 Conclusions 110
Tables 112
Figures 118
4.6 Appendix A: Detailed model description 124
4.6.1 Model description 124
4.7 Appendix B: Model fit 139

4.8 Appendix C: Sensitivity analysis 141
Chapter 5 Structural identifiability of compartmental models for infectious disease transmission is influenced by data type 142
Abstract 143
5.1 Introduction 144
5.2 Methods 148
5.2.1 General modelling framework and formal definitions 148
5.2.2 Model structures 149
5.2.3 Toolbox employed 153
5.2.4 Structural identifiability assessments 154
5.3 Results 156
5.4 Discussion 158
5.5 Conclusions 163
Chapter 6 Discussion 170
6.1 Recap of contributions 170
6.2 Strengths of thesis 171
6.3 Limitations of thesis and potential future directions 174
Chapter 7 Conclusion 177
References 179

Abstract

This thesis concerns the principled use of mathematical models for informing control interventions in the context of an infectious disease outbreak context. We address two main related questions. First, how can interventions be evaluated and compared in a principled manner using a dynamic (time-varying), mathematical model? The second question pertains to structural identifiability, a prerequisite for successful parameter inference in dynamic models: how does the type of model output influence the structural identifiability of dynamic, mathematical models for infectious disease transmission?

We address the first question using two outbreaks of infectious disease – hepatitis A and African swine fever (ASF) – as illustrative examples, and the second question by conducting structural identifiability analysis of various ordinary differential equation (ODE) model versions. Links to all code and files needed to reproduce the results in this thesis have been provided. Our contributions are detailed in three papers: a published article and two manuscripts, now summarized.

Chapter 3: Dankwa et al. (2021) developed a dynamic, deterministic model to explain transmission in the 2017-2019 hepatitis A virus (HAV) outbreak in Louisville, Kentucky, US, among persons experiencing homelessness or who use drugs, known to be at a high risk of HAV infection. With the model, alternative vaccination scenarios were examined for effectiveness and cost, and an estimate for the critical vaccination threshold required for herd immunity in this population was derived. Prior to this study, no such estimate had been obtained for this population anywhere in the US.

Chapter 4: Dankwa et al. (2022b) developed a dynamic, stochastic, spatial model to explain ASF virus (ASFV) transmission among wild boar and domestic pig herds, and to evaluate alternative outbreak management measures. The model was developed such that it could be refined to account for more outbreak data and hence it is suitable for real-time outbreak analysis. This study is relevant, given the ongoing ASF epidemic in Europe, and the scarcity of ASFV transmission models which account for transmission at the wildlife-livestock interface, despite the evidence of such transmission in Europe.

Chapter 5: Dankwa et al. (2022a) conducted structural identifiability analysis of unknown parameters, including initial conditions, of 26 ODE model versions to demonstrate the influence of the type of model output(s) on models' structural identifiability. Data types such as incidence and prevalence, typically encountered in disease surveillance, were studied as model outputs. This analysis emphasizes the importance of a careful consideration of model outputs prior to performing inference with transmission models.

List of abbreviations and acronyms

The abbreviations and acronyms used in the thesis are explained below.

Abbreviation/acronym	Meaning
ABC	approximate Bayesian computation
ASF	African swine fever
ASFV	African swine fever virus
CI	confidence interval
COVID-19	Coronavirus disease 2019
CrI	credible interval
CSTE	Council of State and Territorial Epidemiologists
cullPZ	preventive culling of all herds in a protection zone
cullTR	preventive culling of all at-risk herds that have traded pigs with an infected farm less than three weeks before detection
cullWB	preventive culling of all pig herds located at less than 3 km from positive wild boar carcasses
DFE	disease-free equilibrium
ERGM	exponential random graph model
FRP	free-roaming domestic pig
GLM	generalized linear model
HAV	hepatitis A virus
IC	initial condition
IgM	immunoglobulin M
incrSZ	preventive measure involving increasing the size of a surveillance zone from 10 km to 15 km
LMPHW	Louisville Metro Department of Public Health and Wellness
MCMC	Markov chain Monte Carlo
MIR	maximum infection range
MLE	maximum likelihood estimate
MSM	men who have sex with men
NPIs	non-pharmaceutical interventions
NSDUH	National Survey on Drug Use and Health
ODE	ordinary differential equation

Abbreviation/acronym	Meaning
OR	odds ratio
OIE	World Organisation for Animal Health
PEH	persons experiencing homelessness or unstable housing
PWUD	persons who use drugs
RR	relative risk
s.g.i.	structurally globally identifiable
s.l.i.	structurally locally identifiable
s.n.i.	structurally non-identifiable
SE	standard error
SIP	Susceptible-Infectious-Post-infectious
SIR	Susceptible-Infectious-Recovered
SLI	Susceptible-Latent-Infectious
SLIRD	Susceptible-Latent-Infectious-Recovered-Deceased
SLIRQ	Susceptible-Latent-Infectious-Recovered-Remission
SLIS	Susceptible-Latent-Infectious-Susceptible
SMC	sequential Monte Carlo
US	United States
WHO	World Health Organization

This page is intentionally left blank.

Chapter 1 Introduction

Wave after wave, the nations race
Two years on, they race, searching for answers
The deadly, yet unrelenting pathogen has no plans of surrender
The year? 2042, not 2022
The situation is all too familiar, yet challenging as new
Dose after dose, booster after booster, measure after measure, they race
Is all hope lost?

1.1 BACKGROUND

The adverse social and economic impacts associated with infectious disease outbreaks remain a constant source of concern globally (Global Burden of Disease 2019 Diseases and Injuries Collaborators, 2020; World Health Organization (WHO), 2020). In 2019 alone, infectious diseases in humans accounted for about 7.7 million deaths worldwide, with tuberculosis and lower respiratory infections together accounting for about half of these deaths (WHO, 2020). Over the past two decades, infectious disease outbreaks have led to significant economic losses globally. Aggregating the results from economic impact studies on the 2003 severe acute respiratory syndrome epidemic (Lee and McKibbin, 2004), the 2009 H1N1 influenza pandemic (Fan et al., 2016) and the 2014-2016 West Africa Ebola outbreak (Huber et al., 2018) reveal a total minimum cost of US$138 billion. The devastating impacts of the ongoing Coronavirus disease 2019 (COVID-19) pandemic are yet to be fully quantified; nevertheless, this pandemic is yet another demonstration of infectious disease outbreaks as a bane to society.

Infectious diseases occurring among livestock pose significant threats to animal health, welfare and conservation, food security, economic growth and public health (Tomley and Shirley, 2009). For example, the current epidemic of African swine fever (ASF) in parts of Europe, Asia and the Pacific has resulted in the loss of at least seven million pigs in a

series of more than 12,000 outbreaks (World Animal Health Information System, 2020; World Organisation for Animal Health (OIE), 2020). These events present a threat to farmers' livelihood, global food security and international trade (Guinat et al., 2016a), as pork accounts for more than 37% of global meat intake (Food and Agriculture Organization, 2020).

Across a range of academic fields, considerable research efforts have been invested towards the understanding, control, prevention and proper quantification of the impact of infectious disease epidemics on both public health and animal health. In mathematics and statistics, research into the spread of infectious diseases is often conducted as a modelling analysis in which a mathematical model is used as an approximation of the disease transmission process and conclusions about the outbreak dynamics are drawn based on inferences made on model *parameters* (Anderson and May, 1991; Andersson and Britton, 2000; Diekmann and Heesterbeek, 2000; Keeling and Rohani, 2011; Vynnycky and White, 2010).

Methods for infectious disease modelling have been employed for the analysis of outbreaks in real time (Camacho et al., 2015; Finger et al., 2019), the evaluation of likely impacts of a range of possible intervention measures such as vaccination (Brisson et al., 2020; Brouwer et al., 2020; Dankwa et al., 2021) and non-pharmaceutical interventions (NPIs) (Flaxman et al., 2020) and for forecasting disease trends (Johansson et al., 2019; Reich et al., 2019; Viboud et al., 2018). Epidemic models have thus served to inform public health responses to several outbreaks in human and animal populations; examples include the spread of bovine tuberculosis in cattle in Great Britain (Donnelly et al., 2007), the 2017 diphtheria outbreak in Bangladesh (Finger et al., 2019) and the West Africa Ebola epidemic (Camacho et al., 2015).

Epidemic models are particularly desired for their flexibility as well as their capability to be used in the evaluation of hypothetical disease control scenarios thus eliminating the need for real-life experiments, which, in an infectious disease context, are likely to prove challenging or present serious ethical concerns, such as in the evaluation of vaccination

interventions (Hendrix et al., 2016). Furthermore, recent advances in computational methods for statistical inference, particularly driven by innovations in Bayesian statistics, have contributed to widening the appeal of mathematical models for infectious disease modelling.

Epidemic models are not without drawbacks, however. There is the potential for models to produce misleading results, especially when necessary checks are not performed, both on the model structure and on the data to which the model is fitted. The subject of *identifiability* (Bellman and Åström, 1970; Chis et al., 2011a; Cobelli and Distefano III, 1980; Massonis et al., 2021a; Miao et al., 2011; Raue et al., 2014; Wieland et al., 2021) is particularly important in this regard, as it concerns whether it is possible to obtain unique solutions for all unknown model parameters given the model structure and available data. Identifiability is thus a necessary condition required for a successful estimation of parameters (Eisenberg and Hayashi, 2014). When a model is non-identifiable (i.e., not identifiable), associated inferences may be unreliable (Kao and Eisenberg, 2018; Roda et al., 2020; Roosa and Chowell, 2019) and resulting recommendations may be misleading. A manifestation of the potential dangers of conducting inferences with a non-identifiable model was illustrated by Kao and Eisenberg with a dengue transmission model: they show two sets of parameters which produced similar fits to incidence data but which yielded very dissimilar incidence predictions given the same intervention (Kao and Eisenberg, 2018).

Two types of identifiability have been described: *structural identifiability* and *practical identifiability*. Structural identifiability analysis focuses attention on the **model structure** (including the measurement function, inputs (if any), initial conditions (ICs; i.e., initial values of model states) and constraints), seeking to determine whether it is possible to find distinct solutions for the unknown model parameters given the model structure (Bellman and Åström, 1970; Chis et al., 2011a; Cobelli and Distefano III, 1980; Miao et al., 2011; Raue et al., 2014; Wieland et al., 2021). Since the purpose of structural identifiability analysis is to determine identifiability issues arising as a result of the model

structure alone, the analysis thrives on the assumption of an ideal data scenario, where neither measurement errors nor noise is present in the observed data. Some authors have therefore preferred to include the term *"a priori*" in structural identifiability discourses (see (Audoly et al., 2001; Cobelli and Distefano III, 1980; Distefano and Cobelli, 1980)), emphasizing the fact that the structural identifiability of a model is a property that is independent of the observed data and completely dependent on the model equations (Massonis and Villaverde, 2020). This is in contrast to practical identifiability which concerns whether the **available data** contains enough information to uniquely determine the values of all unknown parameters (Brouwer et al., 2017; Chis et al., 2011b; Kao and Eisenberg, 2018; Miao et al., 2011; Raue et al., 2009; Tuncer et al., 2016; Tuncer and Le, 2018; Wieland et al., 2021). Structural identifiability is a necessary, but not sufficient, condition for practical identifiability; that is, a structurally non-identifiable parameter cannot be practically identifiable, and a structurally identifiable parameter could be practically non-identifiable depending on the data available (Cobelli and Distefano III, 1980; Eisenberg et al., 2013). Our focus here is on structural identifiability.

1.2 MOTIVATION FOR THIS WORK

What types of interventions will be most efficient for successfully managing an outbreak? Which interventions are most likely to optimize both cost and public health outcomes? How long should interventions be implemented to realize a set of desired outcomes? Policy makers are often confronted with these kinds of questions in choosing among a range of possible interventions.

As noted above, mathematical models are being increasingly employed to assist in this decision-making process. Mathematical models are indeed useful for answering these types of questions, as they are capable of providing insights on both epidemiological and economic outcomes given a range of alternative intervention scenarios. A major part of

my work (Chapter 3 and Chapter 4) is dedicated to applying mathematical models to answering such real policy questions in the contexts of recent outbreaks of hepatitis A in the United States (US) (Centers for Disease Control and Prevention, 2020) and ASF in Europe (Department for Environment, Food and Rural Affairs, 2022).

Due to the important place models tend to occupy in the public health policymaking process, it is crucial that model inferences are reliable. As already mentioned, an important property which affects the reliability of model inferences is structural identifiability. Although structural identifiability is a necessary condition for successful parameter estimation and has accordingly received considerable attention in the systems biology and control literature (e.g., Audoly et al., 2001; Balsa-Canto et al., 2010; Castro and de Boer, 2020; Chis et al., 2011a, 2011b; Hong et al., 2020; Massonis and Villaverde, 2020; Meshkat et al., 2014; Pohjanpalo, 1978; Raue et al., 2013, 2009; Vajda et al., 1989; Villaverde et al., 2016; Wieland et al., 2021), it is inconsistently applied in the infectious disease modelling literature. Relatively few studies exist on identifiability analysis focused on between-host infectious disease transmission models (Eisenberg et al., 2013; Evans et al., 2005; Kao and Eisenberg, 2018; Massonis et al., 2021a; Tuncer et al., 2016; Tuncer and Le, 2018) and the practice of routinely checking the identifiability of these models before parameter estimation in infectious disease models is not widespread.

The limited application of structural identifiability analysis in the infectious disease modelling literature may be attributed to two factors: 1) limited accessibility of the literature on structural identifiability analysis methods; and 2) the scarcity of studies which demonstrate the application of these methods to infectious disease transmission models. The vast majority of the literature on approaches for testing structural identifiability is mathematically heavy (Castro and de Boer, 2020), thus not very accessible to a wider audience. (The problem of testing structural identifiability has often been described as a challenging one (Audoly et al., 2001; Chis et al., 2011b; Meshkat et al., 2014; Muñoz-Tamayo et al., 2018; Villaverde et al., 2016).) This limited accessibility

presents potential challenges for understanding and accurately interpreting results from structural identifiability analysis software. Also, the standard benchmark models used for illustrating the implementation of such software are mostly systems biology models (see for example, Hong et al. (2019), Ligon et al. (2018), and Villaverde et al. (2016)). Although some studies exist which apply such software to infectious disease transmission models (Eisenberg et al., 2013; Hong et al., 2019; Kao and Eisenberg, 2018; Massonis et al., 2021a; Tuncer et al., 2016; Tuncer and Le, 2018), implementation details and associated code are seldom provided. Given the current surge in infectious disease modelling publications, stimulated by the COVID-19 pandemic, it is critical that the barriers to such an important facet of modelling as structural identifiability analysis be urgently tackled. Chapter 5 of my work is a contribution in this direction.

1.3 OBJECTIVES AND CONTRIBUTIONS

My thesis has three objectives. The first two objectives concern the principled comparison of disease control interventions using mathematical transmission models, while the third objective concerns the structural identifiability of such models. Contributions on these objectives are presented in detail in three main papers (chapters) which make up the body of this *integrated* thesis. My specific contributions are provided in an authorship statement at the end of each main chapter, except in Chapter 4 where a statement is included as part of the manuscript. The thesis objectives and main contributions are now outlined.

- **Objective 1**: To develop a mathematical model to evaluate the impacts of alternative vaccination interventions on hepatitis A virus (HAV) transmission

Our main contributions to the achievement of Objective 1 are now summarized.

1. We develop a dynamic (time-varying), deterministic model of HAV transmission among a high-risk group – persons experiencing homelessness and unstable housing or who use drugs (PEH/PWUD) – in

a US city. Ours is the first model of HAV transmission among this group in the US, as far as we know.

2. To inform relevant vaccination policies, we apply the model to estimate:
    - the effectiveness of various vaccination interventions within this group, in terms of number of cases averted and savings in hospitalization costs, demonstrating that earlier and faster vaccination rollouts are most effective for control; and
    - the critical vaccination threshold needed to achieve herd immunity within this group, the first such estimate for this group as far as we know.

These contributions are detailed in the following publication presented in Chapter 3.

> **Dankwa, E.A.**, Donnelly, C.A., Brouwer, A.F., Zhao, R., Montgomery, M.P., Weng, M.K. and Martin, N.K., 2021. Estimating vaccination threshold and impact in the 2017–2019 hepatitis A virus outbreak among persons experiencing homelessness or who use drugs in Louisville, Kentucky, United States. Vaccine *39*(49), 7182-7190. doi: https://doi.org/10.1016/j.vaccine.2021.10.001

- **Objective 2**: To develop a mathematical model to evaluate impact of alternative control interventions on ASF virus (ASFV) transmission in a multi-host context involving wild boar and domestic pigs

Our main contributions to the achievement of Objective 2 are now summarized.

1. We develop a stochastic, spatiotemporal model for ASFV transmission among wild boar and domestic pig herds. Although cross-species transmission has been shown to be critical to the spread of ASFV in the current epidemic in Europe (Department for Environment, Food and Rural

Affairs, 2022), only a small number of studies (Pietschmann et al., 2015; Pollock et al., 2021; Taylor et al., 2021; Yoo et al., 2021) have captured this kind of transmission; the majority of studies have been limited to a single species (Hayes et al., 2021). Our multi-host model is therefore relevant, and a valuable contribution to the diversity of the literature.

2. We demonstrate how our ASF model may be used to evaluate a range of control interventions, including regulations stipulated for managing ASF in the European Union (European Commission, 2002, 2020a).
3. Our model projections reveal the importance of timescale considerations when evaluating interventions: projections show that interventions which appear to have limited impact in the short-term may yield significant long-term benefits.

These contributions are detailed in the following manuscript presented in Chapter 4. The manuscript is currently undergoing peer review.

> **Dankwa, E.A.**, Lambert, S., Hayes, S., Thompson, R.N., Donnelly, C.A. Stochastic modelling of African swine fever in wild boar and domestic pigs: epidemic forecasting and comparison of disease management strategies.

- **Objective 3**: To demonstrate the influence of the type(s) of data considered as model output(s) on structural identifiability of typical deterministic models for infectious disease transmission

Our main contributions to the achievement of Objective 3 are now summarized.

1. We perform structural identifiability analysis on unknown parameters, including ICs, of 26 *ordinary differential equation* (ODE) model versions, each defined to have a distinct structure-output combination.

2. Analysis results show that,
    - structural identifiability of a parameter may vary depending on the type of model output; and
    - in general, models with multiple data types as outputs tend to have more structurally identifiable parameters, than models with a single data type as output.
3. By providing code and outputs used in the analysis, our work serves as a tutorial for practitioners seeking to learn and apply structural identifiability techniques, thus promoting uptake of structural identifiability analysis in infectious disease modelling studies.

These contributions are detailed in the following manuscript presented in Chapter 5.

> **Dankwa, E.A.**, Brouwer, A.F., Donnelly, C.A. Structural identifiability of compartmental models for infectious disease transmission is influenced by data type.

1.4 REPRODUCIBILITY

In keeping with principles of open science and best practices for reproducibility, we provide well-documented code and relevant files for all results presented in this thesis. Links to the associated repositories are provided in each main chapter (Chapter 3, Chapter 4 and Chapter 5).

1.5 THESIS OUTLINE

The outline of the rest of the thesis is as follows.

In Chapter 2, I provide an overview of basic compartmental models for infectious disease transmission. I highlight some distinctions between deterministic and stochastic

compartmental models, discuss some extensions to the basic Susceptible-Infectious-Recovered (SIR) model structure and briefly discuss key epidemiological quantities – the basic reproduction number, the herd immunity threshold and the critical vaccination threshold – employed later in this thesis (i.e., in Chapter 3). I also briefly discuss common parameter estimation techniques applicable to these model types. This overview covers Sections 2.1-2.3. In Section 2.4, I review structural identifiability in the context of ODE models, focusing on methods and associated software tools for testing structural identifiability.

Chapter 3, Chapter 4, and Chapter 5 are dedicated to Objectives 1, 2, and 3, respectively, as explained above. In Chapter 6, I summarize the main contributions of this thesis, discuss its strengths and limitations and propose directions for further work. In Chapter 7, I present concluding remarks.

Chapter 2 Basic compartmental models for infectious disease transmission

2.1 BASIC COMPARTMENTAL MODELS

Compartmental models derive their name from their design: in a compartmental model, individuals in the population are grouped into compartments based on their infection status; individuals who share a common status belong to the same compartment. We focus on the basic so-called SIR compartmental model, based on the work of Kermack and Mckendrick (1927). In the SIR model, at any given time during the epidemic, an individual is either susceptible to the disease (that is, capable of being infected but not yet so), infectious (that is, infected and capable of transmitting the infection) or recovered from the disease and immune, or deceased.

2.1.1 A deterministic SIR model

We now describe a basic deterministic SIR model. Suppose an infectious individual is introduced into a homogeneously mixing population. Assume the population now has a constant size, N and occupies a constant area. Let $S(t)$, $I(t)$ and $R(t)$ represent the number of susceptible, infectious and removed individuals respectively within this population at any given time, t; for example, at $t = 0$, $S(0) = N - 1,\ I(0) = 1$ and $R(0) = 0$.

New infections arise as infectious individuals contact susceptible individuals. The number of new infections per unit time is $\beta SI/N$, where β is the per-capita transmission rate and is calculated as the product of c, the constant contact rate and ν, the probability that a contact is a successful transmission event (Begon et al., 2002). Infectious individuals recover at the rate γ and once recovered, become permanently immune to

the infectious pathogen. The dynamics of this infection process can be described by the following set of ODEs:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{2.1}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I(t) \quad , \tag{2.2}$$

$$\frac{dR}{dt} = \gamma I(t) \tag{2.3}$$

equipped with ICs $S(0)$, $I(0)$ and $R(0)$. Typically, no explicit form exists for the solution to the system of equations (2.1)-(2.3) due to the nonlinearity of $\beta SI/N$; hence, numerical methods are often employed to obtain approximate solutions (Keeling and Rohani, 2011).

2.1.2 A stochastic SIR model

The stochastic SIR model discussed in this section is as described in (Bartlett, 1949). Let $S(t)$, $I(t)$, $R(t)$, N, β and γ be defined as before. Consider a time interval $(t, t+\tau)$ within which either an infection or recovery may occur. Inter-event times are assumed to be independent and identically distributed (i.i.d.) exponential random variables with mean τ. The probability of an infection during this interval is given by $\beta S(t)I(t)\tau/N + o(\tau)$. Similarly, the probability of a recovery is $\gamma I(t)\tau + o(\tau)$. The equations below show the state transitions and probabilities for an infection and a recovery, respectively (Greenwood and Gordillo, 2009).

$$P\Big(\big(S(t+\tau), I(t+\tau)\big) - \big(S(t), I(t)\big) = (-1,1)\Big) = \beta S(t)I(t)\tau/N + o(\tau) \tag{2.4}$$

$$P\Big(\big(S(t+\tau), I(t+\tau)\big) - \big(S(t), I(t)\big) = (0,-1)\Big) = \gamma I(t)\tau + o(\tau) \tag{2.5}$$

This model is the so-called *general stochastic epidemic*.

Realisations under this model are usually obtained by implementing the Gillespie's stochastic simulation algorithm (Gillespie, 1977, 1976), also known as Gillespie's direct method, originally developed in the context of chemical kinetics. The algorithm simulates the time to the next event (infection or recovery) from an exponential distribution with rate equal to the sum of event rates at the previous iteration, and thereafter determines the exact event to occur by a sampling procedure whose probability weights are based on event rates at the previous iteration. The number of individuals in each class is then updated based on the sampled event.

2.1.3 Deterministic versus stochastic compartmental models

The deterministic and stochastic compartmental models discussed in the preceding sections are now compared. We note that the deterministic-stochastic distinction is typically not black-and-white in practice, as some deterministic models are defined to include stochastic components, and vice versa (Vynnycky and White, 2010).

As hinted in Section 2.1.2, the main difference between the stochastic and deterministic models is the factor of randomness which is present in the stochastic model but absent from the deterministic model. In their book, Andersson and Britton (2000) suggest that the stochastic framework is 'the most natural way' of analysing the spread of disease. Indeed, real-world infection processes possess inherent randomness due to variations in environmental and demographic factors (Duncan et al., 2013; Lambert et al., 2018).

It is well known that the law of mass action obeyed by the deterministic model relies on the law of large numbers. Consequently, the deterministic model does not accurately model disease spread in small communities. Since random effects tend to be more pronounced when numbers are small (for instance, in the early stages of an epidemic (Bartlett, 1949)), stochastic models are better suited to such contexts. The two models are however linked in this regard, in that the deterministic model is the large-population limit of the stochastic model (Greenwood and Gordillo, 2009).

Stochastic models are preferred when the aim is to report uncertainty estimates for a desired model outcome (e.g., final size). Due to the factor of randomness in the stochastic model, different runs of the model with the same set of input values may lead to different realisations which can be combined to give a distribution for the outcome of interest (Vynnycky and White, 2010). For a deterministic model, a fixed set of input values would always produce a fixed outcome. However, it is possible, with deterministic models, to obtain a distribution for the outcome of interest by varying input parameter values. Methods for obtaining such distributions include prediction profile likelihoods (Kreutz et al., 2012), Bayesian approaches, ensembles of models with different parameterizations and error propagation using the Fisher information matrix. For a comparison of the last three methods, see Villaverde et al. (2019).

Another difference between the deterministic and stochastic models is found in the area of disease extinction (Britton, 2010). In the deterministic SIR, the number of infected individuals decreases at an exponential rate and thus only gets to zero in infinite time (Allen, 2017; Lloyd, 2001). In the stochastic model however, it is possible for the number of infections to reach zero in finite time, enabling the estimation of extinction probability (Allen, 2017).

2.1.4 Extensions to the SIR model

The SIR models discussed in the foregoing sections rely heavily on simplifying assumptions and as such are not reflective of most real-world epidemics. Several modifications have therefore been proposed to allow for the study of more complex transmission processes.

For example, the SLIR model is an extension to the SIR model to include a latent (L) (or, exposed, E)[1] class comprising individuals who have been infected but are not yet

[1] Although the term "exposed" is more commonly used in the literature to describe individuals who have been infected but not yet infectious, we prefer to use the term "latent" as it a more descriptive

infectious. The SLIR model is thus suitable for modelling diseases which have a considerable latent period, for example, measles and mumps (Keeling and Rohani, 2011; Porter and Oleson, 2016). In cases where the disease-causing pathogen has a primary host other than the main population under consideration, as in malaria, models can be defined to include a separate set of compartments to represent the dynamics of the host population; for example, a Susceptible-Latent-Infectious-Susceptible/Susceptible-Latent-Infectious (SLIS/SLI) model defined for malaria transmission, with the SLIS and SLI structures representing dynamics among humans and mosquitoes respectively (Anderson and May, 1991).

Compartmental models have been defined to account for age heterogeneity amongst hosts (Anderson and May, 1991; Keeling and Rohani, 2011) via defining age-specific transmission rates, infectious periods or latent periods, the latter being prominent in models for HIV as its mean incubation period is known to be age-dependent (Medley et al., 1987). Non-uniform infection risks (especially relevant for sexually transmitted infections) (Keeling and Rohani, 2011) and non-uniform mixing rates (Isham, 2005) have also been modelled in a similar fashion.

Extensions have also incorporated spatial structure (Brooks-Pollock et al., 2014; Keeling and Rohani, 2011; Porter and Oleson, 2016). Spatial epidemic models have enabled insight into spatiotemporal patterns of disease spread and have facilitated the evaluation of possible interventions at various spatial scales (Chowell and Rothenberg, 2018). Such insights cannot be obtained under the simple SIR framework.

Deaths, births and migration can be easily incorporated into the SIR structure (Anderson and May, 1991; Greenwood and Gordillo, 2009; Keeling and Rohani, 2011). The need to account for potential changes in population structure is particularly relevant when modelling an endemic disease (e.g., malaria) or a disease that has a seasonal

medical term and avoids ambiguity: individuals who are exposed to a pathogen may not necessarily become infected as a result of the exposure.

transmission cycle (e.g., influenza) as significant fluctuations in population size are expected in such scenarios.

It must be noted that model complexity tends to increase with the addition of more realistic features. In designing an epidemic model, the general advice is to begin with the use of parsimonious models with simple assumptions to help build capacity for the use of more realistic albeit less tractable models (Bailey, 1975; Greenwood and Gordillo, 2009).

2.2 KEY EPIDEMIOLOGICAL QUANTITIES

2.2.1 The basic reproduction number

A key measure of the transmissibility of an infection is the basic reproduction number, typically denoted by R_0, and generally regarded as one of the most important quantities in epidemiology (Anderson, 1991; Heffernan et al., 2005; Keeling and Rohani, 2011; Li et al., 2011; Nishiura and Chowell, 2009; Vynnycky and White, 2010). R_0 is defined as the average number of secondary infections caused by a typical infected individual throughout their infectious period, in a wholly susceptible population. The consequence of this definition is immediately evident. If $R_0 < 1$, only a minor outbreak occurs since each infected individual infects less than one other individual on average. Conversely, if $R_0 > 1$, each infected individual transmits the infection to more than one individual, on average, hence an increase in the number of cases. (This increase is guaranteed if the model is deterministic. If it is stochastic, there is a possibility that the epidemic might die out early by chance.) This result is the so-called *threshold property* of R_0 (Anderson and May, 1991; Heesterbeek, 2002). An expression for R_0 for the basic SIR model is obtained by combining the threshold criterion and the result of Kermack-McKendrick's threshold theorem (Kermack and McKendrick, 1927), which broadly states that under the model assumptions described in Section 2.1.1, an epidemic will only take off in a population if the number of susceptible individuals exceeds γ/β. The resulting expression is (Anderson, 1991; Kermack and McKendrick, 1927):

$$R_0 = \frac{\beta}{\gamma}. \tag{2.6}$$

From equation (2.6), we can deduce that R_0 depends on the contact rate c, the probability of infection given contact v, and the duration of infectiousness $1/\gamma$ (recall that $\beta = cv$).

Due to its threshold property, R_0 is often used as a measure of the potential risk of epidemics as well as for the determination of the type and scale of interventions required to control their spread (Heffernan et al., 2005; Li et al., 2011).

As real epidemics are hardly observed at the moment when there is only one infected individual, it is challenging, in practice, to calculate R_0 based on its definition (Li et al., 2011). Secondary methods are therefore employed for the estimation of R_0: these methods either involve the direct estimation of R_0 from epidemiological data (e.g., incidence data (Riley et al., 2003) or seroprevalence data (Keeling and Rohani, 2011)) or its derivation in terms of model parameters. The survival function method (Heesterbeek and Dietz, 1996) and the next generation method (Diekmann et al., 1990; Diekmann and Heesterbeek, 2000) are examples of methods commonly used for the derivation of R_0 given a transmission model. The next generation method is of particular interest to the current work (employed in Chapter 3), as it is can be applied to models with more than one infected class (for example, in the SLIR model) (Heffernan et al., 2005).

In a population with both susceptible and immune hosts, or in the presence of NPIs and behavioural changes which influence the rate of transmission, the effective (also, instantaneous) reproduction number $R(t)$ (Nishiura and Chowell, 2009) is the analogue of R_0 at time t. See Cori et al. (2013) and Nishiura and Chowell (2009) for some methods for estimating $R(t)$.

For a given infectious pathogen, the value of R_0 may differ across different populations, due to its partial dependence on the rate of contact within a population (captured by β)

(Anderson and May, 1982). It is worth noting the existence of surrogate threshold parameters which possess a similar threshold property as R_0 but are not suitable for comparing the risks of different diseases since the biological interpretation of these surrogates may differ distinctly from that of R_0; see Heffernan et al. (2005) for examples.

2.2.2 Herd immunity threshold and critical vaccination coverage

The herd immunity threshold (ψ) is the population proportion that should be immune to ensure a decreasing or stable disease incidence (Fine et al., 2011). The concept dates back to the work by (Topley and Wilson, 1923) who studied the spread of a bacterial infection among groups of immunized and susceptible mice. They found that both mortality and the rate of spread of the infection were inversely related to the proportion of initially immune mice. Although not explicitly noted in their paper, the existence of an immunity threshold was hinted by the authors and their study---particularly, the questions posed therein---provided the basis for later, more detailed work on the subject (Fine, 2005). It was Dietz (1975) who proved the now well-known expression for ψ:

$$\psi = 1 - \frac{1}{R_0}. \tag{2.7}$$

Where a vaccine confers only a partial degree of protection, the critical vaccine coverage level (V_c), given by

$$V_c = \frac{\psi}{\omega}, \tag{2.8}$$

where ω is the proportion of vaccinated individuals that are completely protected, is used; to possibly control an outbreak through vaccination, ω must be greater than ψ (Fine et al., 2011). Other extensions to ψ accounting for waning immunity and non-uniform mixing have been proposed (Fine et al., 2011) but are outside the scope of this work.

2.3 PARAMETER ESTIMATION

Given a transmission model, it is often of interest to determine which values (or distributions) of model parameters are most likely to have produced a given set of observations (for example, incidence data).

Inference with compartmental transmission models are usually performed using standard statistical parameter estimation techniques (Cox, 2006; Robert, 2007); these typically involve the computation or approximation of a *likelihood function* – an expression of the likelihood of a parameter value for a given observation.

Assume a set of observations, $\boldsymbol{D} = \{D_1, D_2, \dots, D_n\}$, from a parameterized probability distribution $f(\boldsymbol{D}; \boldsymbol{\theta})$ where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_k\}$ is a vector of unknown model parameters. The likelihood function is written as

$$L(\boldsymbol{\theta}|\boldsymbol{D}) = f(\boldsymbol{D}; \boldsymbol{\theta}). \tag{2.9}$$

2.3.1 Maximum likelihood estimation

The method of parameter estimation which involves the determination of the value of $\boldsymbol{\theta}$ that maximizes $L(\boldsymbol{\theta}|\boldsymbol{D})$ is known as *maximum likelihood estimation*. See Andersson and Britton (2000), Brouwer et al. (2020), Dankwa et al. (2021), Guinat et al. (2016b) and Riley et al. (2003) for examples of applications of maximum likelihood estimation for inference in compartmental models.

By definition, the maximum likelihood estimate (MLE) is that which satisfies

$$\boldsymbol{\theta}_{MLE} = \arg\max_{\theta \in \Theta} L(\boldsymbol{\theta}|\boldsymbol{D}), \tag{2.10}$$

where Θ is the parameter space. The maximum likelihood estimator is mostly preferred for its intuitive and ideal statistical properties: *consistency* (estimate tends to the true parameter value as sample size increases), *efficiency* (the variance of the estimator is the lowest among a set of candidates when the sample size is large), *parameterization*

invariance (for any function g, the MLE of $g(\boldsymbol{\theta})$ is $g(\boldsymbol{\theta}_{MLE})$) and *asymptotic normality* (the distribution of the estimator tends to the normal distribution as the sample size increases).

2.3.2 Bayesian inference

In the foregoing discussions, it has been assumed that the parameter of interest, $\boldsymbol{\theta}$, is a fixed quantity. This assumption underlies the *classical* (*frequentist*) approach to inference. In the *Bayesian* paradigm, parameters are thought to be random variables with probability distributions; see Cox (2006) for a comparison of frequentist and Bayesian approaches. Prior to observing $\boldsymbol{D}$, the agent specifies a (subjective) probability distribution for $\boldsymbol{\theta}$, termed the *prior distribution*. Inference on $\boldsymbol{\theta}$ is based on samples from a *posterior distribution*: this distribution is derived by combining the prior distribution and the likelihood via Bayes' rule.

Let $\pi(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{D})$ denote the prior distribution and posterior distribution respectively. By Bayes' rule and the law of total probability,

$$\pi(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{f(\boldsymbol{D};\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{D})} = \frac{f(\boldsymbol{D};\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{D};\boldsymbol{\theta})\pi(\boldsymbol{\theta})}\ ; \tag{2.11}$$

the integration in the denominator is over $\boldsymbol{\theta}$. In higher dimensions, $q(\boldsymbol{D})$ is usually intractable, hence in practice,

$$\pi(\boldsymbol{\theta}|\boldsymbol{D}) \propto f(\boldsymbol{D};\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{2.12}$$

is rather considered. It is often the case that a closed-form analytical solution does not exist for the likelihood and hence direct simulation from the posterior distribution may prove challenging.

2.3.2.1 Markov Chain Monte Carlo methods

Markov chain Monte Carlo (MCMC) methods are a widely used family of computational techniques which enable the estimation of the posterior distribution without requiring the possibility of direct simulation. The goal of MCMC is to obtain samples from the posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{D})$, which may be used to compute a range of summary statistics to describe $\boldsymbol{\theta}$. This goal is achieved by constructing an *ergodic* Markov chain which has $\pi(\boldsymbol{\theta}|\boldsymbol{D})$ as its stationary distribution and running the chain until convergence is reached: samples drawn from the chain after convergence are realisations from $\pi(\boldsymbol{\theta}|\boldsymbol{D})$ (Andersson and Britton, 2000, chap. 11).

MCMC methods have been employed widely for parameter estimation in infectious disease models. Examples include (Andersson and Britton, 2000; Camacho et al., 2015; Finger et al., 2019; Flaxman et al., 2020; Porter and Oleson, 2016; Seymour et al., 2022).

2.3.2.2 Likelihood-free inference

To employ MCMC, one requires access to a likelihood function which can be computed exactly or via an approximation. For complex models with likelihoods that are either computationally expensive or unavailable in closed form, likelihood free-inference methods such as approximate Bayesian computation (ABC) (Beaumont et al., 2002) offer a useful alternative for conducting parameter inference. ABC methods circumvent the evaluation of the likelihood function by employing simulations (Beaumont et al., 2002). The ABC rejection algorithm (Pritchard et al., 1999), the most basic version of ABC, works as follows. First, a value for the parameter $\boldsymbol{\theta}^*$ is sampled from a specified prior distribution $\pi(\boldsymbol{\theta}^*)$. A data set $\boldsymbol{D}^*$ is then simulated from the model of interest given $\boldsymbol{\theta}^*$. With a tolerance level ϵ, a distance measure $d(.)$ and the observed data $\boldsymbol{D}$, $\boldsymbol{\theta}^*$ is accepted if $d(\boldsymbol{D}^*, \boldsymbol{D}) \leq \epsilon$, otherwise $\boldsymbol{\theta}^*$ is rejected. The process is repeated until a desired number of acceptances. The sample of accepted parameter values (so called *particles*)

approximates the desired posterior distribution and may be used in drawing inferences about $\boldsymbol{\theta}$.

Especially in high-dimensional data settings, user-defined summary statistics are mostly employed in the definition of $d(.)$, for the comparison of observations $\boldsymbol{D}$ and simulations $\boldsymbol{D}^*$. A discussion on optimal choice for summary statistics is given by (Nunes and Balding, 2010). The choice of summary statistics as well as the level of tolerance determines the efficiency of an ABC algorithm (Minter and Retkute, 2019).

Generally, the ABC rejection algorithm is computationally inefficient, particularly in high-dimensional parameter settings, since the algorithm requires sampling from the prior distribution at each iteration hence maintaining a wide range of possible values of $d(.)$ that may be obtained. ABC-sequential Monte Carlo (ABC-SMC) (Toni et al., 2009) is an improvement which allows for a gradually decreasing tolerance schedule and multiple intermediate distributions, each distribution corresponding to a smaller tolerance than the previous distribution. In ABC-SMC, a finite set of particles is first drawn from the prior distribution, and particles corresponding to each intermediate distribution are obtained by sampling from the previous distribution based on the weights of particles and perturbing sampled particles via a kernel. Minter and Retkute (2019) demonstrate the superior efficiency of ABC-SMC compared to ABC rejection on a stochastic Susceptible-Exposed-Infected-Recovered model for measles.

2.4 APPROACHES TO STRUCTURAL IDENTIFIABILITY ANALYSIS IN ODE MODELS

Before providing an overview of approaches to structural identifiability analysis in deterministic ODE models, I first introduce the general modelling framework, notations and definitions.

2.4.1 General modelling framework and definitions

Consider a deterministic ODE model $\mathcal{M}$, a generalization of equations (2.1)-(2.3), of the form:

$$\mathcal{M} := \begin{cases} \dot{\boldsymbol{X}}(t) &= f\big(\boldsymbol{X}(t), \boldsymbol{p}, \boldsymbol{u}(t)\big) \\ \boldsymbol{y}\,(t) &= g(\boldsymbol{X}(t), \boldsymbol{p}) \\ \boldsymbol{X}_{t_0} &= \boldsymbol{X}(t_0) \end{cases}, \tag{2.13}$$

with observations on the interval $t_0 \leq t \leq T$, where $\dot{\boldsymbol{X}}(\boldsymbol{t})$ is a system of non-linear ODEs, $\boldsymbol{X}(\boldsymbol{t}) \in \mathbb{R}^{n_X}$ is a vector of time-varying disease states and the unique solution to the system $\dot{\boldsymbol{X}}(\boldsymbol{t})$, $\boldsymbol{p} \in \mathbb{R}^{n_p}$ is a vector of constant unknown model parameters, $\boldsymbol{y}(t) \in \mathbb{R}^{n_y}$ is a vector of time-dependent model outputs corresponding to a specific data type (for example, case incidence rates), $\boldsymbol{u}(t) \in \mathbb{R}^{n_u}$ is a time-dependent input vector, g is the measurement equation which defines the relationship between $\boldsymbol{X}(\boldsymbol{t})$, $\boldsymbol{p}$ and $\boldsymbol{y}$, and $\boldsymbol{X}_{t_0} \subset \mathbb{R}^{n_X}$ is a vector of ICs for the states.

We adopt the definitions for structural identifiability for a model and its parameters, as in Cobelli and Distefano III (1980) and Ljung and Glad (1994). As there may be degenerate parameter values or ICs which render an otherwise identifiable model non-identifiable, the following definitions apply to almost all parameter values and ICs; that is, excluding these special cases (Eisenberg and Hayashi, 2014; Saccomani et al., 2003).

Definition 1 [Parameter structural identifiability] (Cobelli and Distefano III, 1980; Ljung and Glad, 1994)

A parameter $p_i \in \boldsymbol{p}$ is structurally globally identifiable (s.g.i.) on $[t_0, T]$ for a given output choice $\boldsymbol{y}$ if a unique solution exists for p_i; that is, if and only if for almost any $\boldsymbol{p}^*$ and almost any IC (i.e., excluding degenerate values), $\boldsymbol{y}(\boldsymbol{X}, \hat{\boldsymbol{p}}) = \boldsymbol{y}(\boldsymbol{X}, \boldsymbol{p}^*)$ implies $\hat{p}_i = p_i^*$. Otherwise, p_i is structurally globally non-identifiable.

A parameter $p_i \in \boldsymbol{p}$ is structurally locally identifiable (s.l.i.) on $[t_0, T]$ for a given an output $\boldsymbol{y}$ if there exists a neighbourhood $V(\boldsymbol{p})$ of the parameter space within which a unique solution exists for p_i. Otherwise, p_i is structurally non-identifiable (s.n.i.).

Definition 2 [Model structural identifiability] (Cobelli and Distefano III, 1980; Ljung and Glad, 1994)

The model $\mathcal{M}$ is s.g.i. for a given output $\boldsymbol{y}$ if every $p_i \in \boldsymbol{p}$ is s.g.i. given $\boldsymbol{y}$.

The model $\mathcal{M}$ is s.l.i. for a given output $\boldsymbol{y}$ if every $p_i \in \boldsymbol{p}$ is s.l.i. given $\boldsymbol{y}$.

The model M is s.n.i. for a given output $\boldsymbol{y}$ if at least one $p_i \in \boldsymbol{p}$ is s.n.i. given $\boldsymbol{y}$.

Remarks on Definition 1 and Definition 2

1. From the definitions, we can deduce the following:
    a. a s.g.i. parameter is, by definition, s.l.i., and
    b. a s.l.i. parameter may not be s.g.i.
2. Parameters may form *identifiable combinations*; i.e., functional forms of parameters which are identifiable although the individual parameters are not (Cobelli and Distefano III, 1980; Eisenberg and Hayashi, 2014; Hong et al., 2019; Massonis et al., 2021b; Meshkat et al., 2014, 2011, 2009; Ovchinnikov et al., 2021). Several methods for determining the identifiable combinations of parameters of $\mathcal{M}$ have been proposed, based on the Fisher Information Matrix and profile likelihood (Eisenberg and Hayashi, 2014), Gröbner bases (Meshkat et al., 2009), characteristic sets (Ovchinnikov et al., 2021), bootstrap (Hengl et al., 2007), Lie Algebra (Massonis et al., 2021b; Massonis and Villaverde, 2020; Merkt et al., 2015) and Taylor series (Evans and Chappell, 2000; Gunn et al., 1997). To facilitate application, some of these methods have been implemented in publicly available software (Hong et al., 2019; Massonis et al., 2021b; Meshkat et al., 2014).

The knowledge of the identifiable combinations in a model is useful for reparameterizing a model to achieve structural identifiability (Massonis et al., 2021b, 2021a; Meshkat et al., 2014, 2009; Ovchinnikov et al., 2021). Such knowledge also helps to determine those parameters for which data should be obtained in order to recover identifiability (Eisenberg and Hayashi, 2014), thus aiding in the optimization of the design of experiments (Raue et al., 2009).

We do not focus on identifiable combinations in this thesis. The interested reader is referred to the above-cited references for further reading on the topic.

2.4.2 Approaches to testing structural global identifiability

Methods for testing structural identifiability of parameters of $\mathcal{M}$ are varied in terms of concept, applicability and scalability. Methods range from a straightforward approach based on directly applying Definition 1 (Denis-Vidal and Joly-Blanchard, 2000) to more sophisticated approaches based on power series expansion (Pohjanpalo, 1978), generating series (Walter and Lecourtier, 1982), differential algebra (Bellu et al., 2007; Ljung and Glad, 1994), similarity transformation (Vajda et al., 1989), profile likelihood (Raue et al., 2009), observability matrix rank test (Karlsson et al., 2012; Sedoglavic, 2002; Villaverde et al., 2016) and scale invariance (Castro and de Boer, 2020). These testing methods may be classed under two broad categories, namely **global approaches** and **local approaches**, corresponding to the two main types of structural identifiability. We will focus on global approaches here. For an overview of local approaches, the reader is referred to Anstett-Collin et al. (2020).

Global approaches yield structural identifiability results which hold almost everywhere in the parameter space. These approaches are generally analytic in nature, treating the structural identifiability of a model as a mathematical property and thus invoking mathematical theory for querying the model structure (Eisenberg and Hayashi, 2014). Due to their consideration of the entire parameter space, there are considerable difficulties associated with the application of global approaches: these arise mainly from

the need to solve, analytically, large systems of complex, nonlinear algebraic equations (Audoly et al., 2001; Bellu et al., 2007; Walter and Lecourtier, 1982) which are challenging to solve by hand and require the use of symbolic computation tools. Relatedly, the problem of testing structural identifiability has often been described as a challenging one (Audoly et al., 2001; Bellu et al., 2007; Chis et al., 2011a; Meshkat et al., 2014; Muñoz-Tamayo et al., 2018; Villaverde et al., 2016).

I now briefly review the main global approaches proposed in the literature: the direct test, similarity transformation, Taylor series, generating series and differential algebra-based approaches.

2.4.2.1 Direct Test

The direct test approach (Denis-Vidal and Joly-Blanchard, 2000) has been considered the most straightforward conceptually (Chis et al., 2011a; Miao et al., 2011), as it involves the direct application of Definition 1. In this approach, one solves

$$y(\boldsymbol{p_1}) = y(\boldsymbol{p_2}) \tag{2.14}$$

seeking to prove that equation (2.14) implies $\boldsymbol{p_1} = \boldsymbol{p_2}$. A noted downside of this framework is the typical complexity of the resulting algebraic expressions, particularly for high-dimensional problems, thus requiring numerical approximations which yield a non-exact solution (Chis et al., 2011a; Miao et al., 2011).

2.4.2.2 Similarity transformation approach

The similarity transformation approach (Vajda et al., 1989) extends to nonlinear ODE models a method proposed by Walter and Lecourtier (1981) for structural identifiability of linear ODE models. Essentially, the approach establishes global identifiability by

applying the *local state isomorphism theorem* (see Vajda et al. (1989) and references therein) to possible transformations of the model states**.** The existence of transformations which do not change the model structure or the mapping between the input $\boldsymbol{u}$ and output $\boldsymbol{y}$ proves that the model is s.g.i.

A key assumption underlying this approach is that the model must be *observable* (i.e., possible to determine its current state $X(t)$ only through its output y for finite t) and *controllable* (i.e., for any initial state X_i and any final state X_f, there exists a finite input sequence $\boldsymbol{u}(k)$, $k = 0, 1, 2, \ldots f$, capable of transferring the system from X_i to X_f (Hermann and Krener, 1977; Kalman, 1960)). (See Evans et al. (2002) for an extension to uncontrolled rational models.)

Applying the local state isomorphism theorem to $\mathcal{M}$ typically yields a system of partial differential equations whose solution is complex (Chis et al., 2011a), hence the method is not suitable for large-dimensional systems (Audoly et al., 2001). In addition, the observability and controllability conditions limit the set of models to which this approach may be applied, and testing for these conditions may present challenges (Chis et al., 2011a; Miao et al., 2011). These are the major disadvantages of the similarity transformation approach.

2.4.2.3 Taylor series approach

The Taylor series approach (Pohjanpalo, 1978) involves expressing the outputs (y_i's) as a Taylor series about t_0, the derivatives (with respect to time) of which are in terms of the parameters only and hence can be solved to provide information on the structural identifiability of the parameters (by applying Definition 1).

Assume that $\boldsymbol{X}$, $\boldsymbol{y}$, $\boldsymbol{u}$, f, and g have infinitely many derivatives with respect to time. Let $y_i(t, \boldsymbol{p})$ represent the Taylor series expansion of y_i, the ith entry of the output vector $\boldsymbol{y}$ and let $d_i^j(\boldsymbol{p})$ be the jth derivative of $y_i(t, \boldsymbol{p})$ with respect to time, evaluated in the

neighbourhood of t_0. The uniqueness of the solution of $d_i^j(\boldsymbol{p})$ for $j = 0, 1, 2, \ldots, j_m$, and $i = 1, \ldots, n_y$, where j_m is the smallest positive integer for which a solution is yielded by the computations, is a sufficient condition for structural global identifiability of $\mathcal{M}$ (Chis et al., 2011a; Pohjanpalo, 1978).

The value of j_m required to establish structural identifiability of $\mathcal{M}$ cannot be known *a priori* (Chis et al., 2011a). In addition, the system formed from the Taylor expansion of the y_i's is usually complex and can be challenging to manipulate (Vajda et al., 1989) even with symbolic computation software, particularly when j_m is large (Chis et al., 2011a). In such cases, only local identifiability results may be obtained (Chappell et al., 1990). These represent major downsides of the Taylor series approach.

2.4.2.4 Generating series approach

Generating series approaches (Balsa-Canto et al., 2010; Chis et al., 2011b; Walter and Lecourtier, 1982) are based on expressing the observations as a series with respect to time, such that the series coefficients, evaluated at the ICs (i.e., at time t_0), are in terms of the output function g and successive Lie derivatives of g with respect to f. The Lie derivative of g with respect to f is given by:

$$L_f g(\boldsymbol{X}) = \frac{\partial g(\boldsymbol{X})}{\partial \boldsymbol{X}} f(\boldsymbol{X}, u), \tag{2.15}$$

where $L_f g(\boldsymbol{X})$ is a $n_y \times 1$ column vector (as $\partial g(\boldsymbol{X})/\partial \boldsymbol{X}$ is a $n_y \times n_X$ matrix and $f(\boldsymbol{X}, u)$ is a $n_X \times 1$ column vector (Villaverde et al., 2016)). The vector of series coefficients thus obtained is the *exhaustive summary* $s(\boldsymbol{p})$*;* i.e., a vector containing all, but only, the information about $\boldsymbol{p}$ available from $\boldsymbol{u}$ and $\boldsymbol{y}$ for any $u \in U \subset \mathbb{R}$ and for any $t \geq 0$ (Walter and Lecourtier, 1982). The existence of a unique solution for $\boldsymbol{p}$ given $s(\boldsymbol{p})$ is a sufficient condition for $\mathcal{M}$ to be s.g.i.

As the set of series coefficients is infinite, $s(\boldsymbol{p})$ needs to be truncated, in practice, before solving: the resulting solution therefore only offers sufficient conditions for identifiability (Walter and Lecourtier, 1982). Relatedly, a downside of the generating series approach is the lack of a bound on the minimum number of Lie derivatives required to fully establish structural identifiability (Balsa-Canto et al., 2010; Chis et al., 2011a).

Further, although the exhaustive summary thus obtained is simpler compared to the expressions obtained via the Taylor series approach, they can still be quite challenging to solve, especially when the ratio of the number of parameters to the number of outputs is large (Balsa-Canto et al., 2010; Chis et al., 2011a). To facilitate an easier detection of structural identifiability issues in this context via visualization and the systematic solving of $s(\boldsymbol{p})$, Balsa-Canto et al. (2010) propose the use of *identifiability tableaus.* A tableau consists of a table of the non-zero elements of the Jacobian formed from the series coefficients with respect to $\boldsymbol{p}$, and with q rows and n_p columns, where q is the number of non-zero series coefficients (infinite, in principle). Within this framework, M is s.l.i. if the rank of the Jacobian equals n_p.

2.4.2.5 Differential algebra-based approach

Differential algebra-based approaches (Audoly et al., 2001; Bellu et al., 2007; Dong et al., 2021; Hong et al., 2020; Ljung and Glad, 1994; Meshkat et al., 2012, 2011, 2009) are suited to general non-linear models with rational or polynomial f and g, and require the model to be observable from its ICs. The basic idea is to derive, algebraically, from equation (2.13) a set of monic polynomial differential equations (i.e., polynomials with leading coefficients equal to one) constructed to be devoid of the state variables, and only in terms of $\boldsymbol{p}$, $\boldsymbol{u}$ and $\boldsymbol{y}$ (and their derivatives) – the so-called *input-output* equations.

Each model output will have a distinct input-output equation (Kao and Eisenberg, 2018). When normalized, the coefficients of the input-output equations make up the exhaustive summary of the model, which can be solved to obtain solutions for $\boldsymbol{p}$. If these solutions

are unique, $\boldsymbol{p}$ is s.g.i. Due to the typical complexity of the exhaustive summary, computer algebra methods such as the Buchberger (1970) algorithm are usually used to derive solutions (Audoly et al., 2001). For details on the differential algebra-based approach, including steps on constructing the input-output equations, see Audoly et al. (2001) and Meshkat et al. (2012).

2.4.3 Software toolboxes

Towards facilitating easier application, some global approaches have been implemented in publicly available software toolboxes. These toolboxes enable the full characterization of the structural identifiability properties of each unknown model parameter p_i. That is, generally, they provide one of the following results for p_i: s.g.i., s.l.i. or s.n.i. In some cases however, these toolboxes are unable to provide results even after long runtimes, due to model complexity (Hong et al., 2019; Ligon et al., 2018), or may provide only local results (here, we assume $\mathcal{M}$ has been correctly specified so that any difficulty encountered by the software is not due to an inaccuracy in the specification of $\mathcal{M}$). For example, a toolbox may indicate that p_i is s.l.i. but may provide no results on which parameters are s.g.i. In that case, the s.l.i. parameters may be interpreted as being "at least s.l.i.", due to the possibility of p_i being s.g.i. (see Chapter 5).

The following publicly available software toolboxes are built on global approaches: COMBOS (Meshkat et al., 2014), based on the differential algebra approach with methods described by Meshkat et al. (2012, 2011, 2009); DAISY (Bellu et al., 2007), based on the differential algebra approach with methods described by Audoly et al. (2001) and Saccomani et al. (2003); GENSSI 2.0 (Ligon et al., 2018), an improved version of GENSSI 1.0 (Chis et al., 2011b), based on the generating series approach coupled with identifiability tableaus (Balsa-Canto et al., 2010) and SIAN (Hong et al., 2019), based on a combination of the differential algebra and Taylor series approaches (Hong et al., 2020).

Chapter 3 Estimating vaccination threshold and impact in the 2017–2019 hepatitis A virus outbreak among persons experiencing homelessness or who use drugs in Louisville, Kentucky, United States

This chapter presents the following published article:

> **Dankwa, E.A.**, Donnelly, C.A., Brouwer, A.F., Zhao, R., Montgomery, M.P., Weng, M.K. and Martin, N.K., 2021. Estimating vaccination threshold and impact in the 2017–2019 hepatitis A virus outbreak among persons experiencing homelessness or who use drugs in Louisville, Kentucky, United States. *Vaccine 39*(49), 7182-7190. doi: https://doi.org/10.1016/j.vaccine.2021.10.001

A statement of specific contributions by the author of this thesis is provided at the end of the chapter.

ABSTRACT

Background: Between September 2017 and June 2019, an outbreak of hepatitis A virus (HAV) occurred in Louisville, Kentucky, resulting in 501 cases and 6 deaths, predominantly among persons who experience homelessness or who use drugs (PEH/PWUD). The critical vaccination threshold (V_c) required to achieve herd immunity in this population is unknown. We investigated V_c and vaccination impact using epidemic modelling.

Methods: To determine which population subgroups had high infection risks, we employed a technique based on comparing the proportion of cases arising before and after the epidemic peak. We developed a dynamic deterministic model of HAV transmission among PEH/PWUD to estimate the basic reproduction number (R_0), herd immunity threshold, V_c and the effect of timing of the vaccination intervention on epidemic and economic outcomes.

Results: Of the 501 confirmed or probable cases, 385 (76.8%) were among PEH/PWUD. Among PEH/PWUD and within the general population, homelessness was a significant risk factor for infection in the initial stages of the outbreak (odds ratios for homeless versus not homeless: 2.62; 95% confidence interval (CI): 1.62–4.25 for PEH/PWUD and 2.39; 95% CI: 1.51–3.78 for all detected cases). Our estimate for R_0 ranges between 2.85 and 3.54, corresponding to an estimate of 69% (95% CI: 65–72) for herd immunity threshold and 76% (95% CI: 72%–80%) for V_c, assuming a vaccine with 90% efficacy. The observed vaccination program was estimated to have averted 30 hospitalizations (95% CI: 19–43), associated with over US$490,000 (95% CI: $310,000–700,000) in hospitalization cost. Greater impact was observed with earlier and faster vaccination implementation.

Conclusions: Vaccination coverage of at least 77% is likely required to prevent outbreaks of HAV among PEH/PWUD in Louisville, assuming a 90% vaccine efficacy.

Proactive hepatitis A vaccination programs among PEH/PWUD will maximize health and economic benefits of these programs and reduce the likelihood of another outbreak.

Keywords: hepatitis A, critical vaccination coverage, persons who experience homelessness or who use drugs, herd immunity, dynamic modelling

3.1 INTRODUCTION

An estimated 1.5 million cases of HAV infection occur worldwide annually (WHO, 2015). Hepatitis A is a liver disease, typically characterized by fatigue, nausea, jaundice, stomach pain and appetite loss (Nelson et al., 2020). HAV is transmitted via the faecal-oral route, through close personal contact with an infected person or by ingesting contaminated food or water (Lemon et al., 2018; Nelson et al., 2020). Consequently, individuals living in poor sanitation conditions as well as men who have sex with men (MSM) are at increased infection risk (Centers for Disease Control and Prevention, 2020; WHO, 2020). Hepatitis A vaccines are highly effective, offering up to 95% protection (Demicheli and Tiberti, 2003; WHO, 2015).

The US introduced hepatitis A vaccination into its routine vaccination program in 1996 for children ≥24 months of age in high-burden communities and for adults with increased risk for HAV infection or severe disease from HAV. In 2006, recommendations expanded to include vaccination of all children between 12-24 months of age, regardless of risk category or location. Therefore, although vaccination coverage among adolescents (aged 13-17 years in 2019) is moderate (two-dose coverage: 77.1%) (Elam-Evans et al., 2020), it is substantially lower among adults (aged ≥19 years in 2018) (two-dose coverage: 11.9%) (Lu et al., 2021). Additionally, despite being recommended for adults at increased risk such as PWUD, and (since 2018) PEH, vaccination coverage among these groups, measured by antibodies to HAV (anti-HAV) seroprevalence, remains low with estimates from 33% to 52% (Collier et al., 2015; Hennessey et al., 2009; Koekpe et al., 2019).

Since 2016, widespread hepatitis A outbreaks have been reported across various states in the US (Centers for Disease Control and Prevention, 2020). Many of these outbreaks have affected MSM populations and PEH/PWUD (Centers for Disease Control and Prevention, 2020; Foster et al., 2018; Wooten, 2019). As of September 24, 2021, 42,223 cases, 25,666 hospitalizations and 385 deaths had been recorded across 36 states due to the outbreaks (Centers for Disease Control and Prevention, 2020). In response, state

health departments initiated public education and vaccination programs, with the latter mostly targeted at groups at high risk (Foster et al., 2018). However, few data are available on the vaccination coverage required to achieve herd immunity among populations at high risk in the US.

To provide evidence-based recommendations for public health outbreak response, we analyze the 2017-2019 HAV outbreak in Louisville, Kentucky. We assess the risks of different population subgroups to HAV infection during the initial outbreak. Further, using a dynamic model of HAV transmission among PEH/PWUD in Louisville, we estimate the basic reproduction number, the critical vaccination threshold and impact of vaccination strategies within this population.

3.2 METHODS

3.2.1 Data

3.2.1.1 Surveillance

During the outbreak, surveillance was conducted by the Louisville Metro Department of Public Health and Wellness (LMPHW). Cases were categorized based on the 2012 US Council of State and Territorial Epidemiologists (CSTE) clinical description for hepatitis A (Centers for Disease Control and Prevention, 2012). Three case categories were considered: confirmed, probable and suspected. Cases that satisfied the CSTE clinical description and had laboratory confirmation of immunoglobulin M (IgM) anti-HAV were classified as confirmed while cases that satisfied the CSTE clinical description and had an epidemiologic linkage to a person who had laboratory-confirmed hepatitis A were classified as probable. All other cases were classified as suspected.

For each reported case, data were collected on a range of time, demographic and epidemiologic variables. Our study employed a subset of these variables namely, reporting date (year and week), age, sex, housing status, illicit drug-use status, hospitalization status and mortality. Illicit drug use was defined as stipulated by the US

Substance Abuse and Mental Health Services Administration (US Department of Health and Human Services, 2017). Persons who reported homelessness, unstable housing or illicit drug use (intravenous or non-intravenous) were classified as PEH/PWUD.

Although data exist on all reported cases (n = 659), our analysis considers only confirmed or probable cases (n = 501) as evidence for HAV infection among suspected cases is weak. Hereafter, confirmed and probable cases are referred to as *detected cases* or *observed cases*. We use *incident cases* to refer to all cases possessing characteristics of a confirmed or probable case, irrespective of reporting status.

This study used deidentified surveillance data and was determined by the Centers for Disease Control and Prevention to constitute research that does not require review by an Institutional Review Board.

3.2.1.2 Vaccination

In response to the high proportion of cases observed among PEH/PWUD, the LMPHW implemented vaccination programs mainly targeted at PEH/PWUD as well as health and social workers who tend to have regular contact with PEH/PWUD. Vaccines were administered through street outreach and at drug rehabilitation centers, tuberculosis clinics, homeless shelters and correction centers. For each individual vaccinated, data were collected on the date of vaccine administration, age, sex, race, MSM status, housing status and illicit drug use status.

3.2.2 Analysis of risks among sub-populations

To determine HAV infection risk associated with PEH/PWUD and non-PEH/PWUD, we utilized a statistical technique proposed by Worby et al. (2015), based on relative risks. The method involves comparing the pre- and post-peak incidence for population subgroups of interest to ascertain which group(s) stood a higher risk of infection during the initial outbreak stages.

We defined the (epidemic) peak as the period within which the maximum number of cases is observed (weeks 32-34 after the index case (April 9-April 29, 2018), similar for PEH/PWUD and non-PEH/PWUD, Figure 3-2A). We estimated the relative risk (RR), defined for each population subgroup as the ratio of the proportion of detected cases in that subgroup during the pre-peak period versus the corresponding ratio during the post-peak period. Case-detection rate within each subgroup was assumed to be constant over the duration of the outbreak (Worby et al., 2015), enabling the estimation of the odds ratio (OR) for the incidence of cases in subgroup i versus subgroup j for the pre-peak relative to the post-peak:

$$\mathrm{OR_{i,j}} = \frac{\mathrm{RR_i}}{\mathrm{RR_j}}. \tag{3.1}$$

An OR value differing significantly from 1 is indicative of a difference between pre-peak and post-peak incidence rates between the subgroups. Specifically, $OR_{i,j} > 1$ suggests infection risk for subgroup i during the initial epidemic is higher than for subgroup j. We note that the principles of RR and OR used here are the same as the traditional concepts (Sedgwick, 2014).

Among confirmed or probable cases, we investigated risk by PEH/PWUD status, age group and sex. Focusing exclusively on PEH/PWUD cases, we conducted further risk analysis after classifying cases by housing status, illicit drug use, age and sex.

3.2.3 Epidemic modelling

3.2.3.1 Model description

To study HAV outbreak dynamics among PEH/PWUD specifically, we developed a deterministic, compartmental mathematical model for HAV transmission among this population. The population, assumed to be closed and of a fixed size N for the duration of the outbreak, was disaggregated into five mutually exclusive compartments based on their infection status: susceptible (S), latent (L), infectious (I), temporary remission (R) or immune (Z).

The model dynamics are as follows: Susceptible individuals contact infected individuals at an effective rate β, and move to a state of latency for $1/\alpha$ days on average, after which they become infectious. At a rate of $\theta(t)$, individuals are vaccinated. Susceptible individuals enter the immune state at rate $\omega\tau\theta(t)S(t)/N$, the effective number of vaccinations at a given time t. This is the product of the total number of vaccinations $\theta(t)$ at t, the fraction of susceptible individuals $S(t)/N$ at t, the first-dose vaccine efficacy τ, and the fraction ω of vaccine doses given to individuals at risk. A proportion, $1-\eta$, of infectious individuals recover temporarily at an average rate of γ, entering the R state. Individuals in the R state experience a relapse after a period of $1/\sigma$ days on average, becoming infectious. Relapse of symptoms may occur in about 10–15% of cases (Centers for Disease Control and Prevention, 2021; Schiff, 1992) lasting 4–8 weeks, and usually tending to be milder than the initial phase (Glikson et al., 1992). A schematic showing movements between compartments is presented in Figure 3-1.

By limiting the model to PEH/PWUD, we assumed that all infector-infectee pairs were contained within the risk population. We also assumed homogenous mixing. The model does not account for hepatitis A-related deaths due to the low hepatitis A mortality rate observed among the risk group (0.8 deaths per 100 cases). Also, the model does not include background mortality due to the relatively short epidemic duration (<2 years) compared to the average lifespan of PEH/PWUD. Lastly, births were not accounted for as no cases were reported among very young children (age range of detected cases was 10–83 years).

3.2.3.2 Model identifiability

Prior to parameter estimation, we checked for *structural identifiability* of all model parameters; that is, whether the parameters could be uniquely estimated given the model structure. Identifiability is a necessary condition for accurate parameter estimation in dynamic models (Cobelli and Distefano III, 1980). For fixed N, all parameters except the

effective contact rate β, the case detection rate κ, the vaccine efficacy τ, and the fraction of vaccines given to individuals at risk ω were identifiable. We reparameterized the model equations (see Supplementary material), so that all parameters were locally structurally identifiable. Due to the reparameterization, the resulting set of model equations (Supplementary equations (S3) and (S4)) represents the dynamics for observed (detected) cases, rather than true (incident) cases. Subscript 1 denotes variables in the reparameterized model. The identifiability analyses were performed in Mathematica and Maple using differential algebra-based methods (Hong et al., 2019; Meshkat et al., 2009).

Although $\omega \times \tau$, the effective vaccination coverage, was structurally identifiable, it was not *practically identifiable*: that is, the observed data did not hold adequate information to estimate it. This situation can arise, for example, if the impact of vaccination was only in the exponential decay portion of the epidemic and was thus harder to detect. Based on vaccination reports, we assumed $\omega = 1$, that is, all people who received vaccines were at risk of hepatitis A infection, and fixed τ, the first-dose efficacy, at a reasonable value as described below.

3.2.3.3 Model parameterization

Parameters of the reparameterized model are described in Table 3-1. We fixed all natural history parameters ($1/\alpha$, $1/\gamma$, $1/\sigma$ and $1/\eta$) at values informed by the literature (Centers for Disease Control and Prevention, 2021; Glikson et al., 1992). As no Louisville-specific estimates were available for the proportion of PEH/PWUD immune to hepatitis A at the start of the epidemic (ε), its value was fixed at the midpoint of the range of estimates of anti-HAV seroprevalence within similar populations in other US cities: PEH in San Francisco (Hennessey et al., 2009) and PWUD in Wisconsin (Koekpe et al., 2019) and San Diego (Collier et al., 2015).

Bounds for the population size of PEH/PWUD, N, were estimated by aggregating information from a range of sources (Buchino et al., 2019; Coalition for the Homeless,

2017; US Census Bureau, Population Division, 2019; US Department of Health and Human Services, 2019); see Supplementary material.

Due to higher rates of comorbidities among PEH/PWUD (Fazel et al., 2014; Walker et al., 2017), we hypothesized that vaccine protection levels would be lower within this group, since comorbidities tend to decrease an individual's immune response (Zimmermann and Curtis, 2019). We therefore fixed the first dose vaccine efficacy parameter, τ, at 90%, approximately 5% less than that expected in the general population (Centers for Disease Control and Prevention, 2021).

A single value for the effective contact rate β_1 yielded a poor fit to the model (Supplementary Figure S4); thus, we implemented a time-varying β_1, which provided a better fit. This was of the form of a sigmoidal function:

$$\beta_1(t) = \beta_s + \frac{\beta_l - \beta_s}{1+e^{-c(t-t^*)}}, \tag{3.2}$$

where β_s is the value of β_1 at the start of the outbreak, β_l is the value of β_1 later in the outbreak, t^* is the transition midpoint time between β_s and β_l, and c is a rate parameter that controls the speed of the transition ($c > 0$). We note that the structural identifiability analyses conducted here was performed for a constant β_1, instead of the time-varying $\beta_1(t)$ defined in equation (3.2). We however confirmed, using profile likelihood curves ((Raue et al., 2009); curves not shown here), that the parameters upon which $\beta_1(t)$ depends are structurally identifiable.

3.2.3.4 Parameter estimation

Using maximum likelihood estimation and assuming the case counts were Poisson-distributed, we estimated $I_1(0)$, the initial observed number of infected individuals, and the four parameters associated with β_1, namely β_s, β_l, c, and t^*. To account for overdispersion, we initially assumed a negative binomial distribution (parameterized with a mean and dispersion parameter k) for the observed case counts; however, the

estimate of the dispersion parameter was large ($1/k \approx 0.01$) suggesting the data did not significantly differ from a Poisson distribution (Lloyd-Smith, 2007). Simultaneous 95% profile-likelihood-based confidence intervals (CIs) (with bounds at the 2.5th and 97.5th percentiles) were obtained for all estimated parameters (Raue et al., 2009). In the maximum likelihood estimation procedure, N was fixed at its mean value, but in estimating the CI for the observed infection trajectory $I_1(t)$, uncertainty in N was incorporated. Uncertainty estimates for $I_1(t)$ were calculated using a Latin hypercube approximation (McKay et al., 1979) of the parameter space confidence region.

3.2.3.5 The reproduction number and critical vaccination threshold

To quantify the epidemic potential, we calculated the basic reproduction number, R_0, defined as the expected number of secondary infections caused by a single infected individual in a wholly susceptible population. Using the next generation matrix method (Heffernan et al., 2005), we derived an expression for R_0 (see Supplementary material):

$$\mathrm{R}_0 = \frac{\beta_s}{\eta\gamma} \ . \tag{3.3}$$

The herd immunity threshold (ψ), defined as the population proportion that should be immune to ensure decreasing or stable incidence, is given by $\psi = 1 - 1/R_0$ (Dietz, 1975). The critical vaccination threshold (V_c) required to achieve herd immunity is given by (Fine et al., 2011)

$$V_c = \psi/\tau. \tag{3.4}$$

We compute MLEs and 95% CIs for R_0, ψ and V_c.

3.2.3.6 Impact of vaccination program and timing

We assessed the impact of the observed vaccination program conducted among PEH/PWUD by comparing simulations with vaccination during the outbreak to a scenario

with no vaccination. In particular, we assessed impact on the number of detected cases prevented, hospitalizations prevented, and the amount (in US$) saved in hospitalization costs. To estimate the number of hospitalizations based on the number of detected cases, we assumed the observed hospitalization rate among detected PEH/PWUD cases in Louisville. We also assumed hospitalization cost per hepatitis A case was US$16,232, based on a national estimate from the 2017 Healthcare Cost and Utilization Project National Inpatient Sample (Hofmeister et al., 2020), the latest cost estimate available at the time of our analysis.

To assess the potential benefits had vaccination efforts been accelerated, we analysed two vaccination scenarios with earlier program initiation week (3 and 11 weeks earlier than the observed vaccination start date, corresponding to weeks 10 and 18 of the outbreak), with total vaccination counts constant across each scenario. Lastly, we implemented a faster roll-out scenario by modifying the earliest initiation scenario (week 10) such that the number of vaccinations in any given week was double the number of vaccinations observed for that week, up to the total number of observed vaccinations. The potential impact of these programs was compared to the observed scenario in which vaccinations began in week 21 of the outbreak.

3.2.3.7 Sensitivity analysis

We performed a variance-based global sensitivity analysis (Saltelli et al., 2008; Sobol', 1993) to assess the sensitivity of the model's output (quantified by the total number of detected cases) to changes in model parameters. The analysis considered all natural history parameters, ε, N and all estimated parameters and involved varying parameters by $\pm 20\%$ of their respective values as given in Table 3-1. Details are in the Supplementary Material.

3.2.3.8 Reproducibility

All analyses were performed in R version 4.0.5 (R Core Team, 2021). Code for reproducing the results in this paper is freely available on GitHub: emmanuelle-dankwa/HAV-outbreak-Louisville (github.com).

3.3 RESULTS

Between September 2017 and June 2019, there were 501 detected HAV cases in Louisville, among whom 385 (76.8%) were PEH/PWUD. The weekly counts of detected cases by risk group are shown in Figure 3-2A. Counts of detected cases, hospitalizations and deaths by year of detection, sex, age group and risk factor are presented in Table 3-2. The observed hospitalization rate among PEH/PWUD in Louisville (66.2%) was slightly higher than the national rate of 59%, calculated from nationally reported outbreak data up to June 28, 2019 (Centers for Disease Control and Prevention, 2020).

3.3.1 Analysis of risks among subpopulations

Among all cases, there was a significant difference between pre-peak versus post-peak incidence rates for PEH versus non-PEH (OR: 2.39; 95% CI: 1.51–3.78; Supplementary Table S1). No significant difference was found between pre-peak versus post-peak incidence rates for PEH/PWUD versus non-PEH/PWUD (OR: 1.01; 95% CI: 0.64–1.59), or by illicit drug use, sex and age groups (Supplementary Tables S1 and S2).

Among PEH/PWUD cases, the estimated RR was significant for individuals who experience homelessness (RR: 1.92; 95% CI: 1.37–2.69). The OR for the homeless group versus the housed group for the pre-peak versus post-peak period was 2.62 (95% CI: 1.62–4.25). No significant differences were observed by sex, age group and illicit drug use among PEH/PWUD (Table 3-3).

3.3.2 Transmission model results

The model fit well to observed data (Figure 3-3A), with Table 3-1 presenting parameter estimates. The R_0 estimate was 3.20 (95% CI: 2.85–3.54), corresponding to estimates of 69% (95% CI: 65–72) for herd immunity threshold and 76% (95% CI: 72–80) for critical vaccination threshold, assuming a vaccine with 90% efficacy. The results also suggest that the transmission began to decrease rapidly and substantially from mid-April 2018 (Supplementary Figure S3), driving the end of the outbreak.

The weekly vaccinations provided to PEH/PWUD are shown in Figure 3-2B. The vaccination program commenced 21 weeks after the index case and reached 9,999 PEH/PWUD over a 22-week period. We estimated this vaccination program averted 30 hospitalizations (95% CI: 19-43) and US$490,000 (95% CI: US$310,000–US$700,000) in hospitalization costs. The impacts across all measures increase with earlier and faster vaccination interventions, with 99 more detected cases and 66 more hospitalizations averted if vaccinations had been initiated in week 10 of the outbreak, at double the observed coverage, compared to week 21 (Table 3-4). The differences in the epidemic trajectories under the different vaccination scenarios become prominent only after week 20 and are most evident at the peak (Figure 3-3B).

The sensitivity analysis showed the initial effective contact rate (β_s), the fraction of immune individuals (ε) and the duration of infectiousness ($1/\gamma$) as the most influential parameters in driving the variation in the total number of detected cases y_1 (Supplementary Figure S4). A scatterplot of y_1 against these three parameters (Supplementary Figure S6) revealed the following trends: larger mean values of y_1 correspond to small values for ε and large values for β_s and $1/\gamma$, while smaller mean values of y_1 correspond to large values for ε and small values for β_s and $1/\gamma$.

3.4 DISCUSSION

This study analysed a recent HAV outbreak in Louisville, Kentucky, with a focus on outbreak dynamics among PEH/PWUD, who constituted the majority (76.8%) of all detected cases.

Within the general population and among PEH/PWUD, we found homelessness to be an influential risk factor for HAV infection during the initial outbreak (Table 3-3, Supplementary Table S1), underscoring the increased risk of hepatitis A transmission among PEH. Our results point to the importance of proactive targeting of PEH in vaccination campaigns, and to the importance of housing and sanitation programs for those living in unstable housing situations in mitigating HAV infection risk. While drug use has been suggested as a risk factor for hepatitis A transmission (Nelson et al., 2020), it was not significant in our analysis, either in the overall population or among PEH/PWUD (Table 3-3, Supplementary Table S1); however, there was a high rate of non-response for drug use (33.1%, Table 3-2), which may result in an underestimate of the true difference in infection risk between PWUD versus non-PWUD in the initial outbreak. Further work is needed to understand whether drug use on its own is an indicator of high risk of hepatitis A infection.

Our results suggest no significant difference in infection risk by sex and age group among PEH/PWUD (Table 3-3) and in the general population (Supplementary Table S1).

We found a high herd immunity threshold (~69%; 95% CI: 65–72) likely required to prevent HAV outbreaks among PEH/PWUD in Louisville, comparable to corresponding estimates from an MSM population in Australia (~65%) (Zhang and Charlett, 2019). Assuming a vaccine with 90% efficacy, our model yielded a critical vaccine threshold of 77% (95% CI: 73–80). Our 95% CI for R_0 (2.85–3.58) is higher than an estimated range for R_0 of HAV derived for the general US population in the pre-vaccine era (R_0: 1.11–1.55 (Van Effelterre et al., 2006)). We expect this to be the case, as the risk profile of PEH/PWUD is generally higher than that of the general population. In particular, PEH are likely to have more effective contacts providing opportunities for transmission due to

reduced access to and use of sanitation and hygiene facilities. This elevated risk of transmission within this population results in a higher R_0.

We found the implemented vaccination program prevented many cases, but even more could have been prevented if initiated earlier and implemented at a faster rate (Table 3-4, Figure 3-3B), corroborating published evidence on vaccination as a key intervention strategy to mitigate the spread of HAV both in the US and elsewhere (Brouwer et al., 2020; Chodick et al., 2008; Fiore et al., 2006; Lemon et al., 2018; McMahon et al., 1996; Van Effelterre et al., 2006). It is possible that the vaccinations in Louisville also mitigated the spread of HAV to neighbouring counties, as there are indications of a ring effect in spread, with Louisville as the center ((Kentucky Department for Public Health, 2019), Figure 8).

It can be challenging to initiate vaccination programs early on in an outbreak, given the considerable amount of time required to obtain the relevant information (e.g., pathogen specimens, risk factors) that is needed to arrange logistics and funding. To navigate these time constraints, practitioners may benefit from using information from earlier outbreaks. In the current context, this could have been, for instance, risk factor information from surveillance data on earlier HAV outbreaks in other states.

Although data collected at vaccination clinics suggest that all the vaccinations included here were given to PEH/PWUD, making our assumption $\omega = 1$ reasonable, it is not clear that all PEH/PWUD were necessarily at risk, due to the absence of definite risk factor data such as anti-HAV seroprevalence. If ω were substantially below 1, then our estimates of the impact of the alternate vaccination scenarios may be overestimates. More work is needed to understand how risk indicators such as homelessness and drug use translate to actual risk.

We found that beginning mid-April 2018, the effective contact rate β_1 rapidly and substantially decreased to an average of about 20% of its initial value at the start of the outbreak (Table 3-1, Supplementary Figure S3). Dynamically, this change was distinct from the susceptible burnout that generally drives the end of epidemics and distinct from

the impact of vaccination. This time-varying effective contact rate might be driven by seasonality, which would be consistent with the findings of Brouwer et al. (2020) where a model with seasonal transmission yielded a better fit to the data than a model with no seasonal pattern. In general, the evidence for the role of seasonality in HAV transmission is mixed (e.g., compare Centers for Disease Control and Prevention (2021) and Fares (2015)), although it may be stronger for certain groups, such as PEH. Aside from seasonality, changes in behaviour, possibly influenced by education and media reporting, may have resulted in the observed change in the effective transmission rate. The infinite upper bound in the 95% CI for c, the transition midpoint time between the initial and latter values of β_1 (Table 3-1), suggests the data could be modelled with the effective contact rate as a piecewise function; that is, allowing an abrupt change in rates (e.g., (Zhang and Charlett, 2019)) over a one-week period, although this may well be less realistic.

The sensitivity analysis showed that the initial effective contact rate, the fraction of initially immune individuals and the duration of the infectious period, in that order, contributed the most to the variation in the number of detected cases (Supplementary Figure S5). As expected, we found that smaller effective contact rates, shorter durations of infectiousness and larger fractions of initially immune individuals yielded fewer detected cases, given our model assumptions (Supplementary Figure S5). Two recommendations for control are in order: 1) improving early case detection through increased surveillance to decrease the period of exposure of an infectious individual hence decreasing transmission risk, and 2) increasing the rate of vaccination of individuals at risk to ensure a larger fraction of immune individuals and consequently, a smaller chance of take-off in the event of a future outbreak.

Given the substantial overlap between the populations of PEH and PWUD in Louisville (about 73.5% of PEH in Louisville are also PWUD) (Buchino et al., 2019) and the lack of data on mixing between these groups, it was not possible to model transmission within the PEH and PWUD populations separately. The dynamics of HAV transmission within

and between these groups is a worthwhile subject for future investigations, potentially providing useful insights for outbreak control within each subpopulation.

Like all modelling studies, ours contains limitations, mainly due to parameter uncertainty. First, there was appreciable uncertainty associated with N, the population size estimate of PEH/PWUD at risk for HAV. It was particularly challenging to estimate the total population size at risk – in particular, those who are at risk among PWUD. Our variance-based sensitivity analysis indicated that our findings were not overly sensitive to uncertainty in population size, but we acknowledge that these estimates are prone to bias. Our analysis assumed all PWUD in Louisville were at risk, but this number may have overestimated the number truly at risk if not all PWUD are at high risk for HAV infection. On the other hand, drug use is highly stigmatized and is often under-reported and therefore this estimate could possibly have underestimated the number at risk. As such it is difficult to assess implications for our analysis, and better at-risk population size estimates would help inform the public health response and related modelling. Second, no local data were available on baseline immunity against HAV among PEH/PWUD, so studies among similar populations in other locations were used. Additional data collection would improve understanding of existing immunity among local PEH/PWUD populations. Third, we neglect MSM status or risk in our analysis due to substantial missing data in relation to this risk factor. If a proportion of male PEH/PWUD were at an additional risk through sex with men, this risk was missing from our analysis. Studies examining populations with overlapping multiple risks (such as MSM who are homeless and/or who use drugs) are warranted. Fourth, it is probable the assumption of a constant rate of case-detection throughout the outbreak, does not hold completely, likely due to under-reporting among PEH/PWUD. Cases among PEH in particular may go unreported due to multiple factors including inability to afford healthcare costs and fear of hostility by service providers (Institute of Medicine (US) Committee on Health Care for Homeless, 1988). Reporting rates are likely to have been lower during the pre-peak period, compared to the post-peak period, due to the educational campaign introduced later in

the outbreak. Thus, the OR for PEH/PWUD versus non-PEH/PWUD may be an underestimation of the true corresponding values.

In conclusion, we find that hepatitis A vaccination programs will need to achieve vaccination coverage of at least 77% among PEH/PWUD in Louisville, based on a vaccine efficacy of 90%, in order to prevent HAV outbreaks among this population. Proactive hepatitis A vaccination for PEH/PWUD can maximize health and economic benefits.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the authors' affiliated institutions.

Acknowledgments. The authors thank the staff of the Louisville Metro Department of Public Health and Wellness for the collection and organization of the study data. E.A.D is grateful to Ben Lambert for useful discussions on optimization. The authors are also grateful to two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

Funding: This work was supported by the US Centers for Disease Control and Prevention National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (Epidemiologic and Economic Modeling Agreement number NU38PS004650). NKM was additionally supported by the National Institutes of Health NIAID and NIDA (grant number R01AI147490), and the San Diego Center for AIDS Research (SD CFAR), an NIH-funded program (P30 AI036214), which is supported by the following NIH Institutes and Centers: NIAID, NCI, NHLBI, NIA, NICHD, NIDA, NIDCR, NIDDK, NIGMS, NIMH, NIMHD, FIC, and OAR. CAD was supported by the MRC (Centre funding) and NIHR (funded by Vaccine Efficacy Evaluation for Priority Emerging Diseases: PR-OD-1017-20007 and HPRU in Emerging and Zoonotic Infections: NIHR20090). AFB was supported by the National Science Foundation (grant DMS1853032) and the National Institutes of Health (grant U01GM110712).

TABLES

Table 3-1: Descriptions, values and references for model parameters. For estimated parameters, values presented are MLEs with profile-likelihood based 95% CIs in parentheses.

Description	Symbol	Value	Reference
Effective contact rate at the start of the outbreak	β_s	0.61 (0.54-0.67)	Estimated
Effective contact rate later in the outbreak	β_l	0.12 (0.05-0.18)	Estimated
Transition midpoint time from β_s to β_l (in weeks)	t^*	36 (33-39)	Estimated
Speed of transition from β_s to β_l	c	0.75 (0.24-∞)	Estimated
Number of infectious individuals in week 0	$I_1(0)$	0.92 (0.63-2.01)	Estimated
Duration of latent period (in weeks)	1/α	1.57	(Centers for Disease Control and Prevention, 2021)
Duration of infectious period (in weeks)	1/γ	4.64	(Centers for Disease Control and Prevention, 2021)
Duration of remission period (in weeks)	1/σ	4.30	(Glikson et al., 1992)
Probability of experiencing a relapse	1 −η	0.11	(Glikson et al., 1992)
Proportion of initially immune individuals	ε	0.43	(Collier et al., 2015; Hennessey et al., 2009; Koekpe et al., 2019)
First dose vaccine efficacy (%)	τ	90	See Section 3.2.3.3
Fraction of vaccine doses given to at-risk individuals	ω	1	Assumed
PEH/PWUD population size	*N*	69,862	(Buchino et al., 2019; Coalition for the Homeless, 2017; US Census Bureau, Population Division, 2019; US Department of Health and Human Services, 2019)

Table 3-2: Summary of detected cases by year, sex, age group, housing status and illicit drug use. Note that "Unknown" is due to non-response.

Variable	Number of cases (%)	Number of hospitalizations (%)	Number of deaths (%)
Year			
2017	42 (8.4)	32 (9.7)	0 (0)
2018	458 (91.4)	299 (90.3)	6 (100)
2019	1 (0.2)	0 (0)	0 (0)
Sex			
Male	332 (66.3)	223 (67.4)	4 (66.7)
Female	169 (33.7)	108 (32.6)	2 (33.3)
Age (in years)			
10-19	5 (1.0)	0 (0)	0 (0)
20-29	95 (19.0)	59 (17.8)	0 (0)
30-39	185 (36.9)	119 (36.0)	0 (0)
40-49	120 (24.0)	85 (25.7)	2 (33.3)
50-59	65 (13.0)	44 (13.3)	3 (50.0)
60-69	23 (4.5)	18 (5.4)	0 (0)
70+	8 (1.6)	6 (1.8)	1 (16.7)
Housing status			
Homeless	128 (25.5)	80 (24.2)	1 (16.7)
Not homeless	358 (71.5)	244 (73.7)	5 (83.3)
Unknown	15 (3.0)	7 (2.1)	0 (0)
Illicit drug use, intravenous			
Yes	276 (55.1)	192 (58.0)	3 (50.0)
No	134 (26.7)	89 (26.9)	2 (33.3)
Unknown	91 (18.2)	50 (15.1)	1 (16.7)
Illicit drug use, non-intravenous			
Yes	189 (37.7)	129 (39.0)	2 (33.3)
No	146 (29.2)	107 (32.3)	2 (33.3)
Unknown	166 (33.1)	95 (28.7)	2 (33.3)
PEH/PWUD [a]			
Yes	385 (76.8)	255 (77.0)	3 (50.0)
No	116 (23.2)	76 (23.0)	3 (50.0)
Total	501 (100)	331 (100)	6 (100)

[a] PEH/PWUD: Persons experiencing homelessness or who use drugs.

Table 3-3: RR and OR estimates for detected cases among PEH/PWUD (n=385) by risk status (homelessness and illicit drug use), sex and age group (in years). For each variable level, total case counts as well as case counts by period (pre-peak, peak and post-peak) are presented. RR and OR used here are as defined in Worby et al. (2015).

Variable	Number of cases				RR[a] (95% CI)	OR[b] (95% CI)
	Pre-peak	Peak	Post-peak	Total		
Homelessness						
Yes	70	22	36	128	1.92 (1.37–2.69)	2.62 (1.62–4.25)
No	92	41	124	257	0.73 (0.63–0.86)	
Illicit drug use						
Yes	155	56	155	366	0.99 (0.95–1.03)	0.71 (0.22–2.30)
No	7	7	5	19	1.38 (0.45–4.27)	
Sex						
Male	102	43	112	257	0.90 (0.77–1.05)	0.73 (0.46–1.16)
Female	60	20	48	128	1.23 (0.91–1.68)	
Age group (years)						
10-19	2	0	0	2	-	
20-29	40	9	32	81	1.23 (0.82–1.86)	Refer to Table
30-39	72	25	63	160	1.13 (0.87–1.46)	S3[c]
40-49	33	16	42	91	0.78 (0.52–1.16)	
50-59	14	10	16	40	0.86 (0.44–1.71)	
60-69	1	2	6	9	-	
70+	0	1	1	2	-	
Period totals	162	63	160	385		

[a] Estimates are not computed for groups with fewer than ten cases in total. Reference for RR is the pre-peak period.

[b] First rows for all variables (except age group) used as reference for OR.

[c] See Supplementary Material

Table 3-4: Comparison of vaccination scenarios by week of initiation and administration rate, using key impact measures. Main values presented are model estimates computed based on MLEs of estimated parameters, while 95% CIs (in parentheses) are computed based on the corresponding profile-likelihood based 95% CIs of estimated parameters.

	Vaccination scenarios by week of initiation			
Measure	**Week 21**	**Week 18**	**Week 10**	**Week 10, rate doubled**
Number of detected cases prevented	46 (29–65)	65 (43–90)	121 (86–160)	145 (104–188)
Number of hospitalizations prevented	30 (19–43)	43 (28–59)	80 (57–106)	96 (69–124)
Amount (in millions of US$) saved in hospitalization cost	0.49 (0.31–0.70)	0.69 (0.46–0.96)	1.30 (0.92–1.72)	1.55 (1.12–2.02)

FIGURES

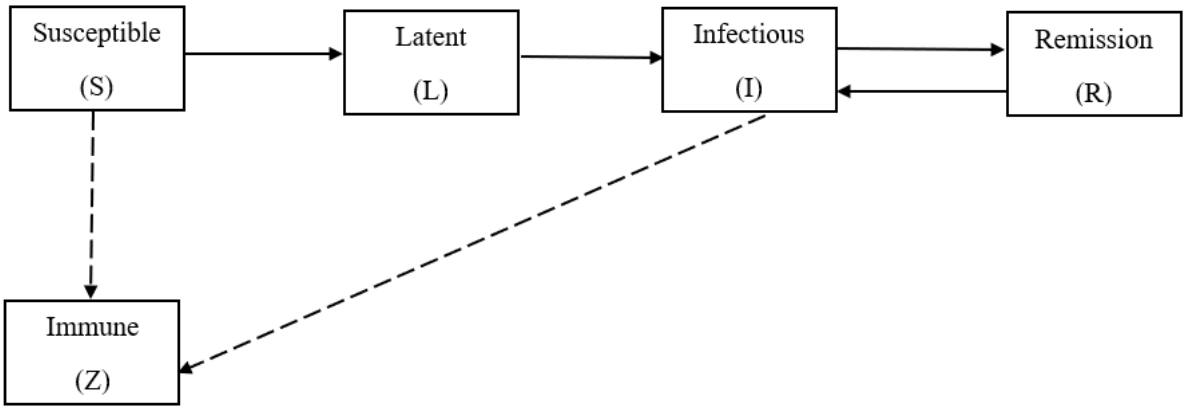


Figure 3-1: Schematic diagram illustrating the transitions between states for the HAV transmission model among PEH/PWUD in Louisville, Kentucky. Transition rates between compartments are detailed in the main text. Solid arrows indicate movements for infected individuals while dashed arrows indicate movements for immune individuals.

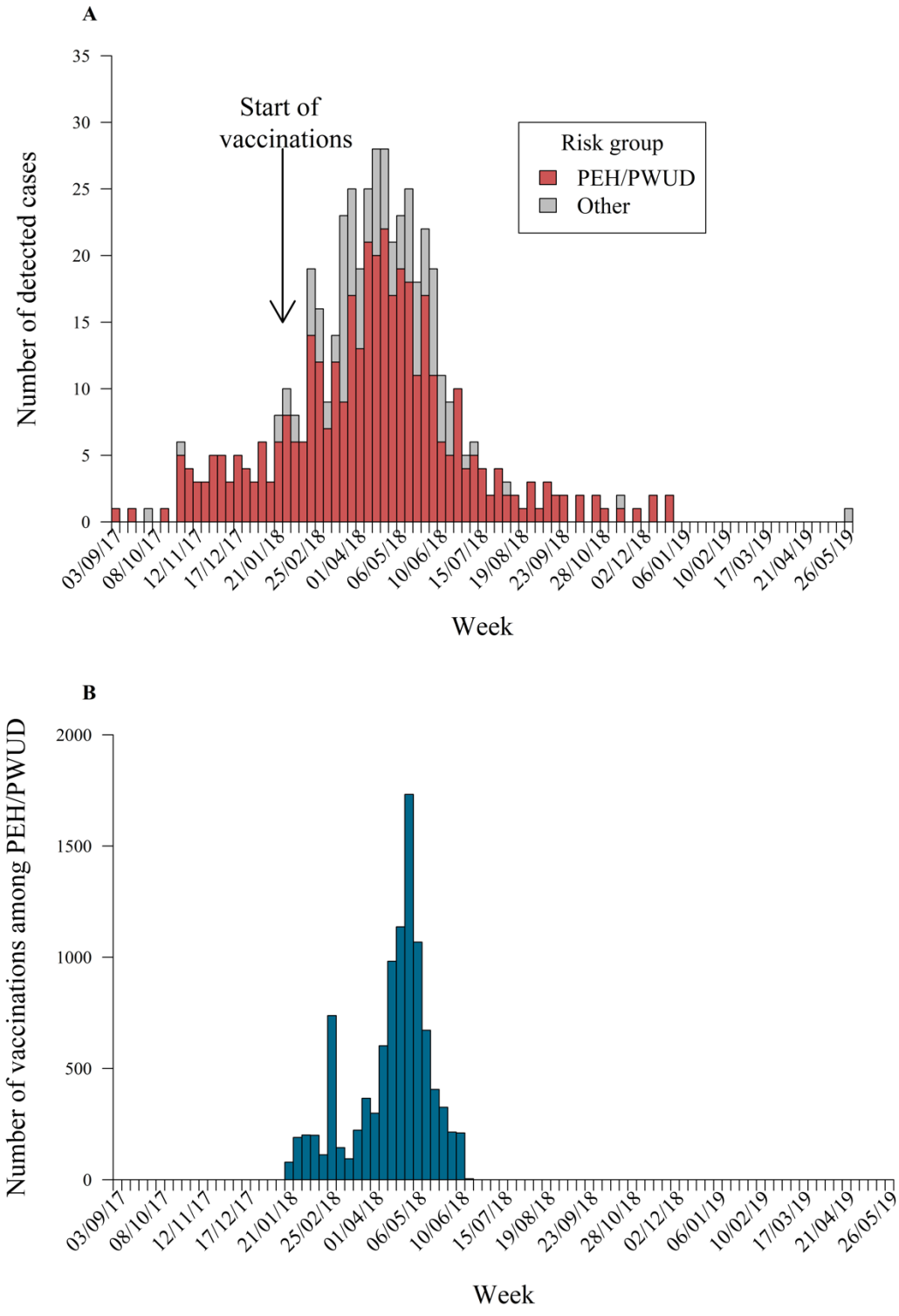


Figure 3-2: **A**: Distribution of all detected cases plotted by risk group; colored bars represent cases among PEH/PWUD. Week labels correspond to the first day of the week. **B**: Weekly distribution of vaccines administered to PEH/PWUD in Louisville during the outbreak. Week labels correspond to the first day of the week.

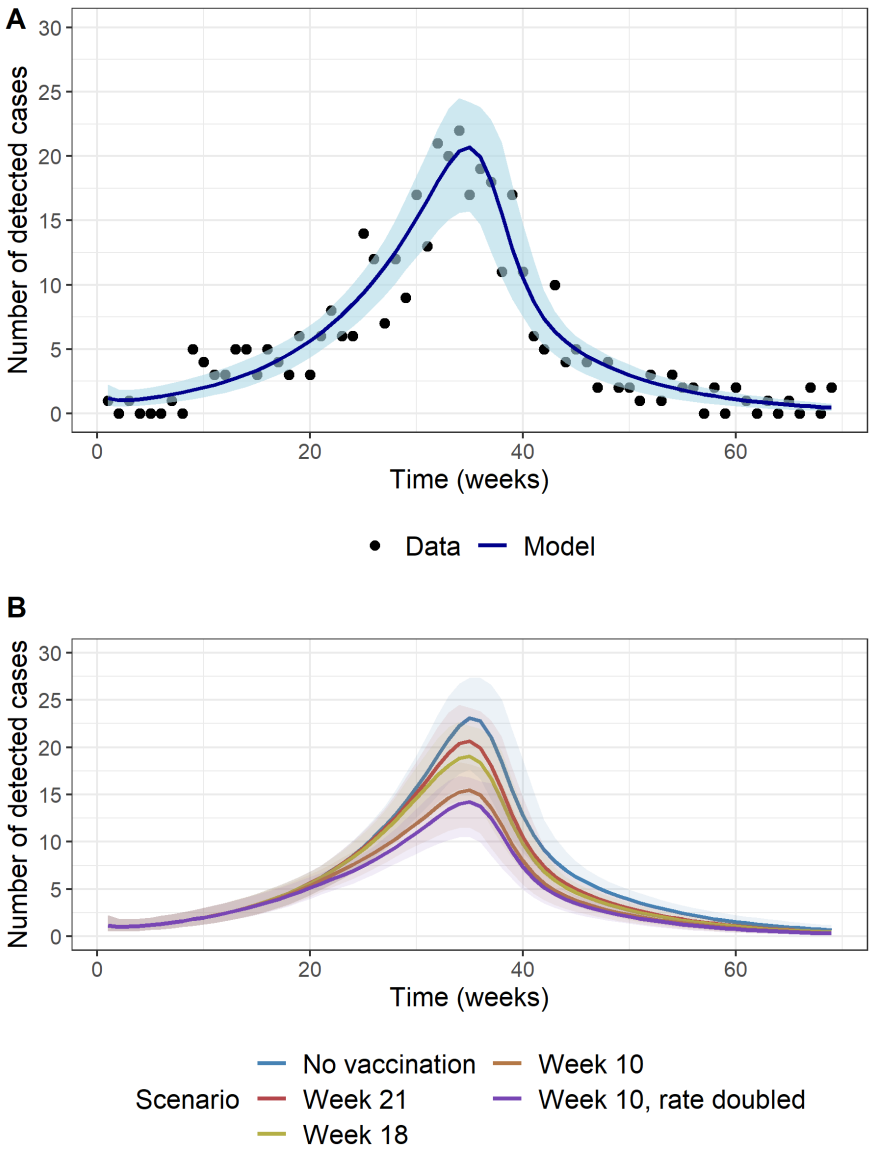


Figure 3-3: **A:** Model fit (blue line) to weekly case counts of detected cases of hepatitis A among PEH/PWUD in Louisville (black dots). Shaded area is the 95% confidence region for the model estimates. **B:** Model estimates for the weekly number of detected cases under the various vaccination scenarios: no vaccination (blue line), vaccination initiation in week 21 as observed (reddish brown line), week 18 (green line), and week 10 with the observed coverage (brown line) and double the observed coverage (magenta line). 95% confidence regions for all scenarios are shaded in corresponding colours.

SUPPLEMENTARY MATERIAL

S1: Relative risk analysis

Table S1 presents RR and OR estimates by age, sex and risk group — PEH/PWUD versus others — along with 95% CIs. Table S2 presents OR estimates with 95% CIs for the various age groups and Table S3 presents corresponding estimates among PEH/PWUD cases. None of the estimates presented in Tables S2 and S3 are significant at the 5% level. Figure S1 displays epidemic curves by risk group, housing status (among PEH/PWUD) and illicit drug use (among PEH/PWUD). The terms RR and OR used here are as defined in Worby et al. (2015).

Table S1: RR and OR estimates for detected cases by risk group, sex and age groups (in years). For each estimate, the 95% CI is in parentheses. For each variable level, total case counts as well as case counts by period (pre-peak, peak and post-peak) are presented.

Variable	Number of cases				RR[a] (95% CI)	OR[b] (95% CI)
	Pre-peak	Peak	Post-peak	Total		
PEH/PWUD						
Yes	162	63	160	385	1.00 (0.90–1.11)	1.01 (0.64–1.59)
No	49	18	49	116	0.99 (0.70–1.40)	
Homelessness						
Yes	70	22	36	128	1.92 (1.37-2.69)	2.39 (1.51-3.78)
No	141	59	173	373	0.80 (0.77-0.83)	
Illicit drug use						
Yes	155	56	155	366	0.99 (0.95-1.03)	0.96 (0.62-1.49)
No	56	25	54	135	1.02 (0.76-1.39)	
Sex						
Male	134	55	143	332	0.93 (0.81–1.07)	0.80 (0.54–1.20)
Female	77	26	66	169	1.16 (0.88–1.51)	
Age group						
10-19	2	0	3	5	-	
20-29	48	12	35	95	1.36 (0.92–2.01)	Refer to Table
30-39	82	28	75	185	1.08 (0.85–1.39)	S2
40-49	46	21	53	120	0.86 (0.61–1.21)	
50-59	25	15	25	65	0.99 (0.59–1.67)	
60-69	7	3	13	23	0.53 (0.22–1.31)	
70+	1	2	5	8	-	
Period totals	211	81	209	501		

[a] Estimates are omitted for groups with fewer than ten cases in total. Reference for RR is the pre-peak period.
[b] First rows for all variables (except age group) used as reference in computation of the OR.

Table S2: Pairwise ORs (row versus column) for age groups (in years) for the pre- vs post-peak periods. 95% CIs for estimates are in parentheses. Estimates are not displayed for age groups with fewer than ten cases.

Age group				
30-39	0.80 (0.47–1.36)			
40-49	0.63 (0.35–1.14)	0.79 (0.48–1.31)		
50-59	0.73 (0.36–1.48)	0.91 (0.48–1.73)	1.15 (0.58–2.28)	
60-69	0.39 (0.14–1.09)	0.49 (0.19–1.30)	0.62 (0.23–1.67)	0.54 (0.18–1.58)
Age group	**20-29**	**30-39**	**40-49**	**50-59**

Table S3: Pairwise ORs (row versus column) for age groups (in years) among PEH/PWUD for the pre- vs post-peak periods. 95% CIs for the estimates are in parentheses. Estimates are not displayed for age groups with fewer than ten cases.

Age group			
30-39	0.91 (0.51–1.62)		
40-49	0.63 (0.33–1.21)	0.69 (0.39–1.21)	
50-59	0.70 (0.30–1.66)	0.77 (0.35–1.69)	1.11 (0.48–2.61)
Age group	**20-29**	**30-39**	**40-49**

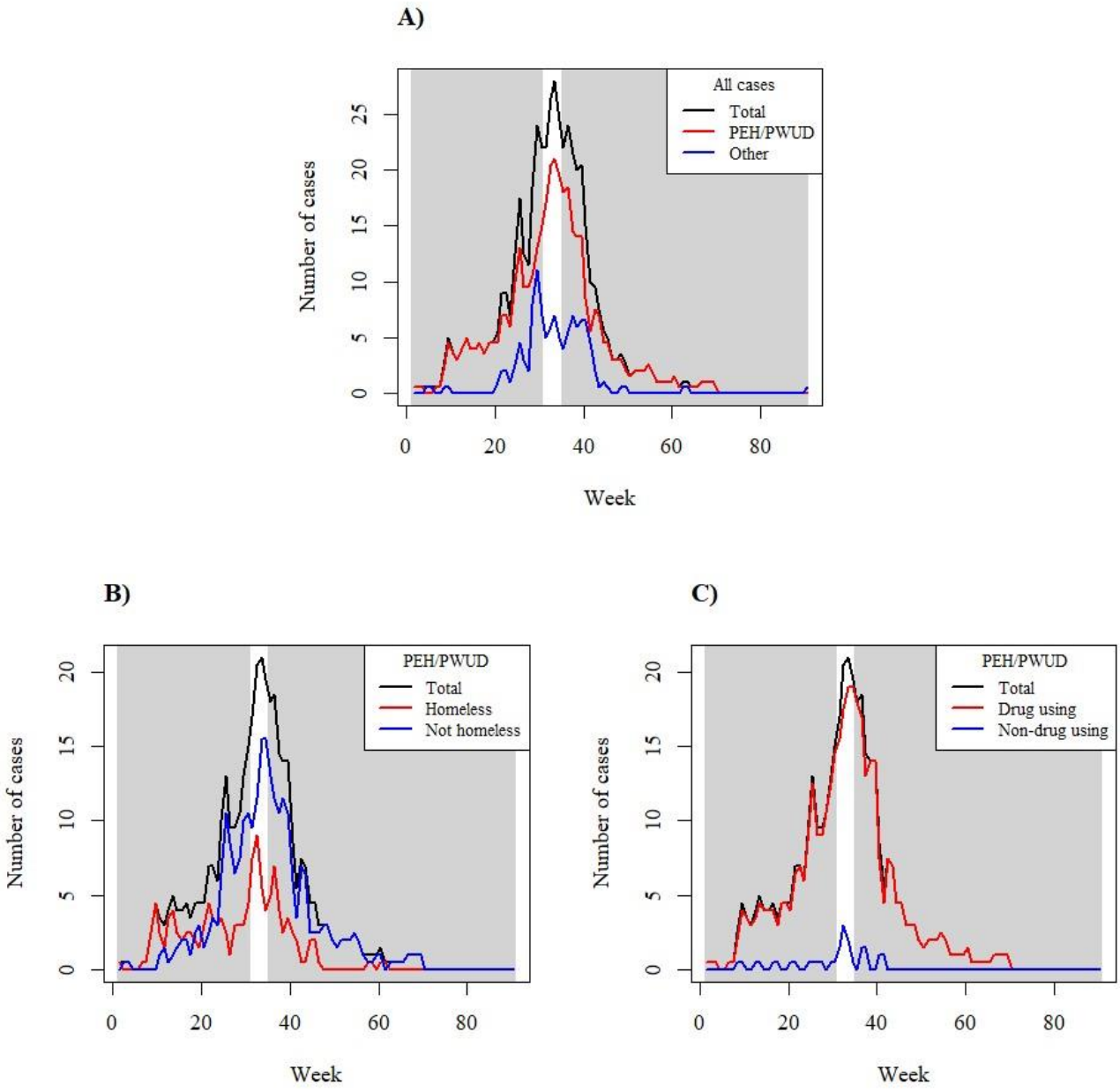


Figure S1: Epidemic curves by risk group, housing status (among PEH/PWUD) and illicit drug use (among PEH/PWUD). A) Epidemic curves for detected cases during the HAV outbreak in Louisville**.** Curves are plotted for all such cases (black curve), cases among PEH/PWUD (red curve) and others (blue curve). **B)** Epidemic curves by housing status among PEH/PWUD. Curves are plotted separately for all PEH/PWUD (black curve), PEH/PWUD cases reporting homelessness (red curve) and PEH/PWUD not reporting homelessness (blue curve). **C)** Epidemic curves by illicit drug use status among PEH/PWUD**.** Curves are plotted separately for all PEH/PWUD (black curve), PEH/PWUD cases reporting illicit drug use (red curve) and PEH/PWUD cases not reporting illicit drug use (blue curve). All: The grey-shaded portions indicate the pre-peak (left) and post-peak (right) periods. Curves are plotted using a 2-week moving average.

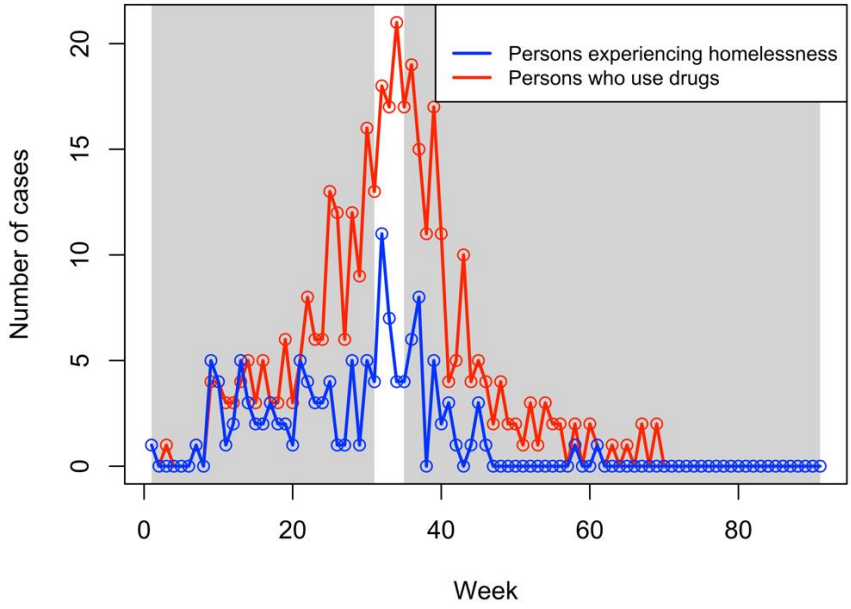


Figure S2: Epidemic curves for persons experiencing homelessness (PEH, red line) and persons who use drugs (PWUD, blue line). Dots indicate the number of cases in each category in any given week. Note that the two groups are not mutually exclusive (see Table 3-2) exists between the two groups. The grey-shaded portions indicate the pre-peak (left) and post-peak (right) periods.

S2: The model

Model identifiability and reparameterization

To study the dynamics of the 2017-2019 HAV outbreak among PEH/PWUD in Louisville, we developed a deterministic compartmental model which classifies individuals in the population according to their infection status. At any time, t, an individual may be in one of the following states: susceptible (S), latent (L), infectious (I), temporary remission (R) or immune (Z). Individuals in the S state are susceptible to the disease, individuals in the L state have the disease but are not yet infectious, individuals in the I state are infectious, individuals in the Z state are immune to the disease as a result of vaccination or past infection and individuals in the R state will experience a relapse of symptoms at some point in the future. In what follows, Greek letters represent model parameters as in Table 3-1 (main text).

Based on the model dynamics as described in the "Model description" section of the main text, the model equations are as follows.

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta SI}{N} - \omega\tau\theta(t)\mathrm{S(t)}/N \\
\frac{dL}{dt} &= \frac{\beta SI}{N} - \alpha L \\
\frac{dI}{dt} &= \alpha L + \sigma R - \gamma I \\
\frac{dZ}{dt} &= \omega\tau\theta(t)\mathrm{S(t)}/N + \eta\gamma I \\
\frac{dR}{dt} &= (1-\eta)\gamma I - \sigma R
\end{aligned} \qquad \text{(S1)}$$

where the parameters τ, $1/\alpha$, $1/\gamma$, $1/\sigma$, $1-\eta$, β and ω represent vaccine efficacy, duration of latency, duration of infectiousness, duration of the remission phase, proportion of relapsing individuals, effective contact rate and fraction of vaccines given to individuals truly at risk, respectively.

We computed the number of detected HAV cases at day t, $\mathrm{y}(t)$, as

$$y(t) = \int_t^{t+1} \kappa\,(\,\alpha L(s) + \sigma R(s))\,ds, \qquad \text{(S2)}$$

where κ is the rate of case detection.

The model as expressed in equations (S1) and (S2) is unidentifiable. In particular, for a fixed N, the parameters β, κ, τ, and ω are not separately identifiable but form identifiable combinations: β/κ and $\tau\omega\kappa$. To ensure identifiability of the model, we reparameterized the model equations.

Consider $S_1 = S \times \kappa$, $L_1 = L \times \kappa$, $I_1 = I \times \kappa$, $R_1 = R \times \kappa$, $Z_1 = Z \times \kappa$, $\beta_1 = \beta/\kappa$, and $\omega_1 = \tau\omega$.

Substituting these expressions into equations (S1) and (S2), we have

$$\begin{aligned}
\frac{dS_1}{dt} &= -\frac{\beta_1 S_1 I_1}{N} - \omega_1 \theta(t) \mathrm{S}_1(\mathrm{t})/N \\
\frac{dL_1}{dt} &= \frac{\beta_1 S_1 I_1}{N} - \alpha L_1 \\
\frac{dI_1}{dt} &= \alpha L_1 + \sigma R_1 - \gamma I_1 \\
\frac{dZ_1}{dt} &= \omega_1 \theta(t) \mathrm{S}_1(\mathrm{t})/\mathrm{N} + \eta\gamma I_1 \\
\frac{dR_1}{dt} &= (1-\eta)\gamma I_1 - \sigma R_1
\end{aligned} \tag{S3}$$

and

$$y_1(t) = \int_t^{t+1} \alpha L_1(s) + \sigma R_1(s))\, ds. \tag{S4}$$

With this reparameterization, the model now represents the dynamics for observed cases, rather than true cases, and all parameters are locally structurally identifiable.

Parameter estimation

Likelihood function

Assuming a Poisson likelihood for the observed case counts $\boldsymbol{x} = x_1, \ldots x_{69}$ (outbreak among PEH/PWUD lasted 69 weeks), we estimated the parameters of the reparameterized model (Table 3-1, main text) via maximum likelihood estimation. Let $\boldsymbol{\theta}$ represent the set of parameters to be estimated. The likelihood function given $\boldsymbol{x}$ is

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{69} \frac{y_t^{x_t} e^{-y_t}}{x_t!},$$

where $y_t > 0$ is the model outcome at day t.

Maximum likelihood estimation enables the estimation of the set of parameter values $\widehat{\boldsymbol{\theta}}$ which maximizes $L(\boldsymbol{\theta})$ and thus best describes the observed epidemic. Uncertainty quantification for the estimated parameters is described in the main text.

Initial conditions

Following Brouwer et al. (2020), we chose the IC for L_1, $L_1(0)$, to be in equilibrium with $I_1(0)$. The ICs for all states are as follows: $I_1(0) = I_0, L_1(0) = I_0\alpha/\gamma,\ Z_1(0) = \varepsilon N,\ R_1(0) = 0$ and $S_1(0) = N - L_1(0) - I_1(0) - Z_1(0) - R_1(0)$.

Evolution of effective contact rate

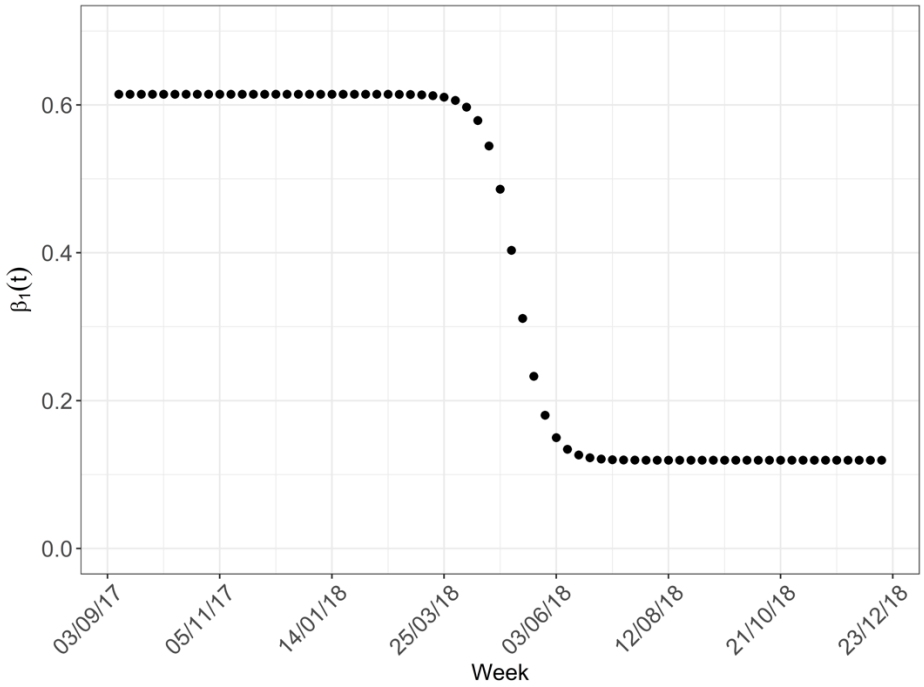


Figure S3: Evolution of the effective contact rate $\beta_1(t)$ in the HAV transmission model. The rate drops sharply and steadily between the second week of April 2018 and the third week of June 2018 but remains constant before and after this period.

Model fit for a single effective contact rate

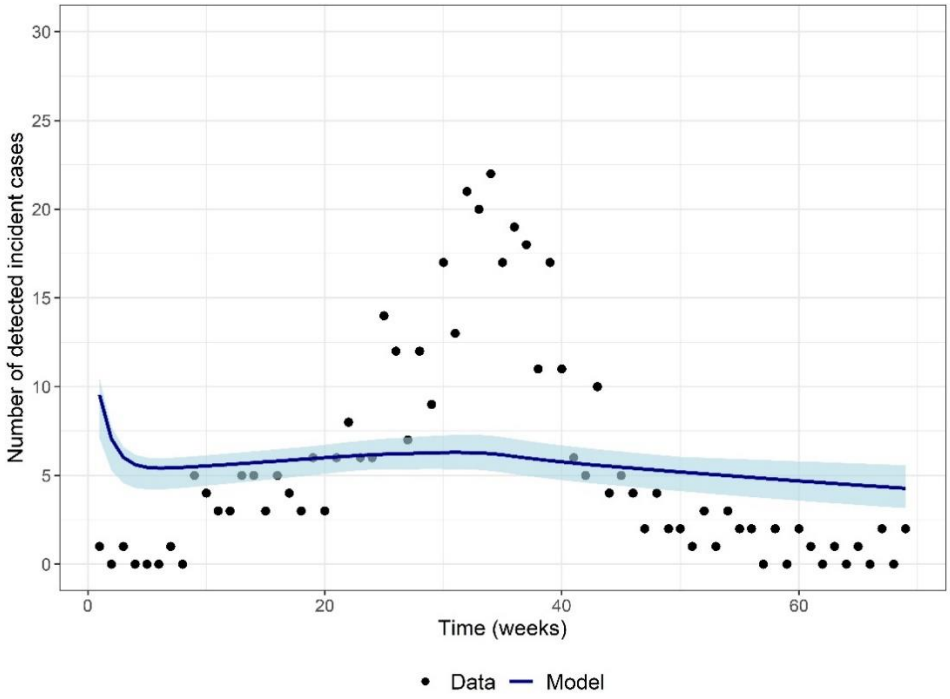


Figure S4: Model fit with one value for the effective contact rate β_1 in the HAV transmission model. Two parameters (β_1 and I_0) are estimated in this model while all other parameters are held fixed.

S3: Estimation of population size of PEH/PWUD in Louisville

In estimating the population size of PEH/PWUD in Louisville, we utilized publicly available data from the US Census Bureau (US Census Bureau, Population Division, 2019), the Coalition of Homeless Persons in Louisville (Coalition for the Homeless, 2017) and the US National Survey on Drug Use and Health (NSDUH) (US Department of Health and Human Services, 2019). The NSDUH provides estimates on various measures of substance use and mental illness among individuals aged 12 and older at national, state and substate levels. Due to the absence of recent substate estimates for illicit drug use (the latest data set at the time writing is dated 2012), we used the Kentucky (i.e., state-level) estimate for the percentage of citizens (aged 12 or older) who are drug users and multiplied this by the population size of Louisville to obtain an estimate for the population size of illicit drug users in Louisville.

Let *A*, *B*, *C* and *D* represent the estimates for the population size of 10-84-year-olds in Louisville (as of July 1, 2017), the population size of homeless individuals in Louisville, the population size of illicit drug users in Louisville and the population size of homeless individuals who use drugs in Louisville, respectively. Also, let *C* (upper) and *C* (lower) represent the upper and lower bounds for *C*. Upper and lower bounds for the size of the PEH/PWUD population were calculated by summing the population sizes of homeless and illicit drug-users in Louisville and subtracting the number of homeless drug-users from the result, as follows:

- Upper bound: *B* + *C* (upper) - *D* and
- Lower bound: *B* + *C* (lower) - *D*.

Thus, we obtained 59,222 and 80,501 as lower and upper bounds respectively for the population size of PEH/PWUD in Louisville. Table S4 contains descriptions and references for all estimates.

Table S4: Estimating the population size of PEH/PWUD in Louisville, Kentucky.

Code	Description	Estimate	Reference	Note
A	Population size of Louisville, 10-84 years (as of July 1, 2017)	658,811	(US Census Bureau, Population Division, 2019)	The age range was chosen to match that of observed cases (10-83 years).
B	Population size of homeless individuals in Louisville (2017 estimate)	6,695	(Coalition for the Homeless, 2017)	This includes both sheltered and unsheltered homeless persons.
C	Population size of illicit drug users in Louisville	67,330 (95% Bayesian credible interval (CrI): 57,448–78,727)	(US Department of Health and Human Services, 2019)	The NSDUH estimates are given in percentages of population size: 10.22% (95% CrI: 8.72%–11.95%).
D	Population size of homeless individuals in Louisville who use drugs	4,921 (73.5% of homeless individuals)	(Buchino et al., 2019)	

S4: Derivation of R_0

The next generation matrix method (Heffernan et al., 2005) was employed in deriving an expression for R_0 given our defined model. The method considers the model dynamics of infected compartments only and focuses on the flow of individuals between these compartments. In the derivation below, we follow the notation by Heffernan et al. (2005).

Let x_i denote the proportion of individuals in compartment i and consider the matrices F and V:

$$F = \left[\frac{\partial F_i(x_0)}{\partial x_j}\right];\ V = \left[\frac{\partial V_i(x_0)}{\partial x_j}\right],\quad \text{(S5)}$$

where $F_i(x_i)$ represents the rate at which *new* infections occur in compartment i, $V_i(x_i)$ represents the net transfer of individuals with respect to compartment i and x_0 represents the disease-free equilibrium (DFE). $V_i(x_i)$is given by $V_i(x_i)= V_i^-(x_i) - V_i^+(x_i)$, where $V_i^-(x_i)$ and $V_i^+(x_i)$ are the outward and inward rates of movement of individuals with respect to compartment i. Then, R_0 is equal to the dominant eigenvalue (or spectral radius) of the matrix FV^{-1}. Note that only infected compartments are considered in this analysis.

For our dynamic model of HAV transmission among PEH/PWUD (equation (S3)), the equations describing the dynamics of infected states are given below.

$$\frac{dL_1}{dt} = \beta_1 S_1 I_1/N\ - \alpha L_1 \quad \text{(S6)}$$

$$\frac{dI_1}{dt} = \alpha L_1 + \sigma R_1 - \gamma I_1 \quad \text{(S7)}$$

$$\frac{dR_1}{dt} = (1-\eta)\gamma I_1 - \sigma R_1 \quad \text{(S8)}$$

At the DFE ($S_1 = N$), we have

$$F = \begin{pmatrix} 0 & \beta_1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } V = \begin{pmatrix} \alpha & 0 & 0 \\ -\alpha & \gamma & -\sigma \\ 0 & \gamma(\eta-1) & \sigma \end{pmatrix}, \quad \text{(S9)}$$

giving

$$R_0 = \frac{\beta_1}{\eta\gamma}\ . \quad \text{(S10)}$$

S5: Sensitivity analysis

To determine the sensitivity of the model to changes in model parameters, we employed a global variance-based sensitivity index initially proposed by Sobol' (Sobol', 1993). Variance-based sensitivity indices quantify the influence of a model's input (a parameter or group of parameters) on its output based on changes in the variance of the output when the input is fixed (Saltelli et al., 2008).

Let $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots \theta_k\}$ denote the set of all model parameters and y the model output. For a parameter θ_i, the first-order Sobol' sensitivity index S_i quantifies the reduction in the variance of y when θ_i is fixed and the other parameters allowed to vary, while the total-order Sobol' sensitivity index T_i quantifies the reduction in the variance of y when the higher order

interactions involving θ_i are fixed and other parameters allowed to vary (Saltelli et al., 2008; Sobol', 1993). Thus, the larger the values of S_i and T_i, the greater the contribution of θ_i to the variance in y and hence the more important θ_i for explaining model dynamics. For all θ_i, $S_i \leq T_i \leq 1$ and $T_i \gg S_i$ means higher order interactions involving θ_i contribute more to the variance in y than θ_i does as a stand-alone (Saltelli et al., 2008).

We computed first- and total-order Sobol' sensitivity indices for the following parameters: all natural history parameters, all estimated parameters, N and ε (see Table 3-1, main text). Parameter sets were initially obtained via a quasi-random number generator (Sobol' sequence (Sobol', 1967)) and then for each parameter, the sampled values were transformed into uniform distributions with bounds at $m_i \pm 0.2m_i$, where m_i is the mean value of parameter θ_i as indicated in Table 3-1. These parameter sets were the model input while the total number of detected cases by the end of the outbreak y_1 was the output. Uncertainty estimates for the sensitivity indices were computed by bootstrap. To ascertain the level of numerical approximation error, we additionally computed sensitivity indices for a dummy parameter with no relationship with the model output and considered the upper (resp. lower) bound of the uncertainty in its sensitivity estimate as the threshold beyond (resp. below) which an estimate is significantly distinguishable from approximation error (Puy et al., 2021). All analyses and plots in this section were produced using the R package **sensobol** (Puy et al., 2021). The package's default methods for first- and total-order sensitivity estimators were followed, in keeping with best practices in sensitivity analysis (see (Puy et al., 2021) for references).

The results (Figure S5) show the initial effective contact rate (β_s), the fraction of immune individuals (ε) and the duration of infectiousness ($1/\gamma$) as the most influential in driving the variation in y_1. The first-order sensitivity indices of all other parameters except t^* and $1/\alpha$ could not be distinguished from approximation error. Also, β_s is shown as the only parameter with a total order sensitivity index significantly higher than its first-order index estimate and the approximation error, suggesting that higher-order interactions involving β_s contribute substantially to variation in y_1. The form of the relationship between y_1 and the three most influential parameters were visualized with the aid of scatterplots (Figure S6). Parameter values and corresponding outputs plotted in Figure S6 are exactly those used in the sensitivity analysis.

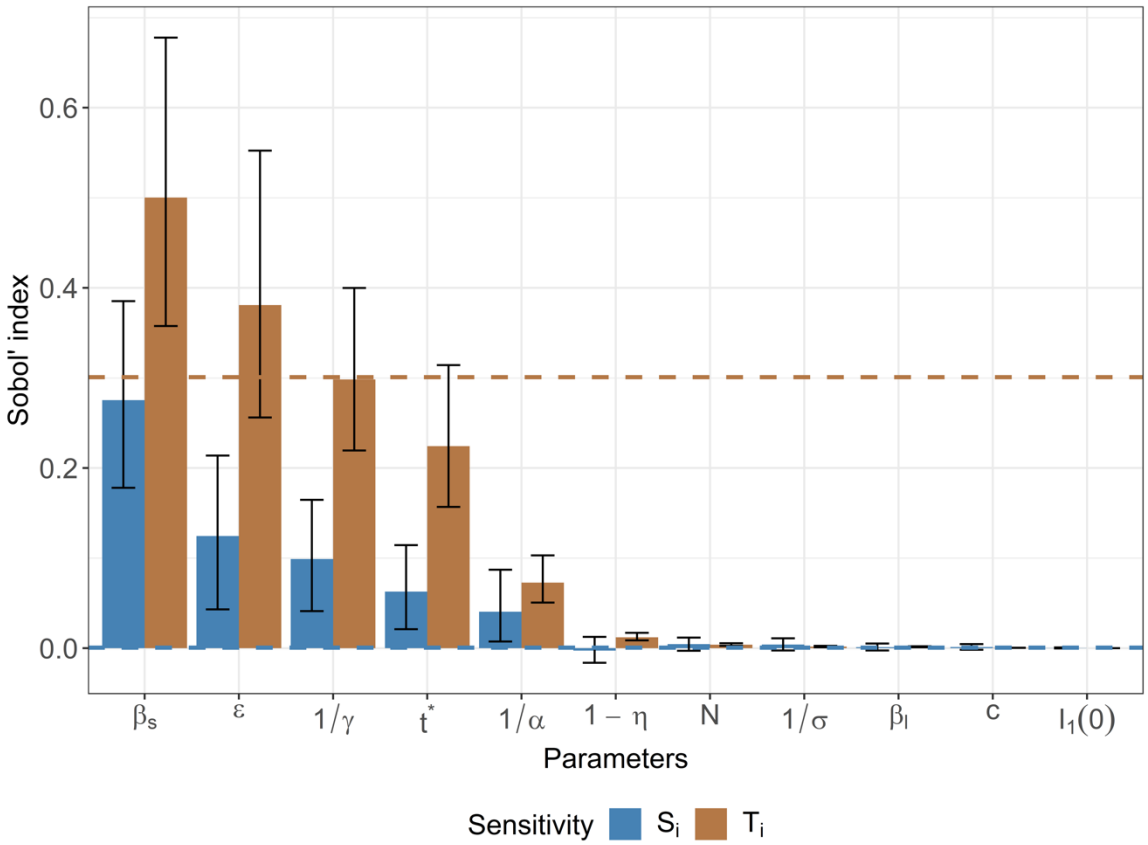


Figure S5: First-order (S_i, blue) and total-order (T_i, brown) Sobol' sensitivity indices for selected parameters of the HAV transmission model. The lengths of the error bars (black) indicate the widths of bootstrap 95% CIs for the sensitivity estimates. Blue dashed line indicates the lower limit of the 95% CI of the first-order sensitivity index estimate for the dummy parameter. Brown dashed line indicates the upper limit of the 95% CI of the total-order sensitivity index for the dummy parameter.

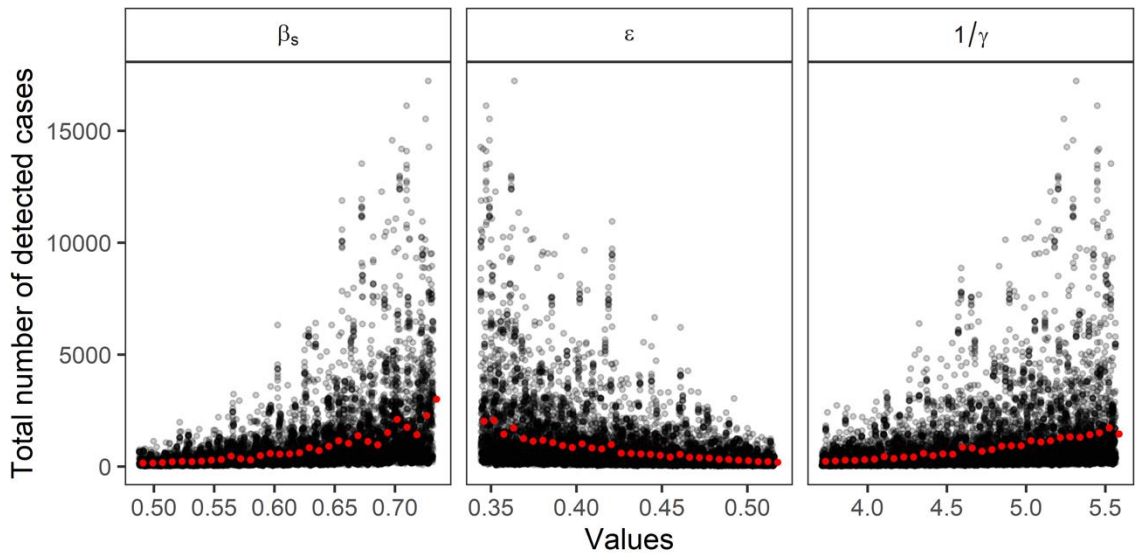


Figure S6: Scatterplots of the number of detected cases by the end of the outbreak against the three most influential parameters per the sensitivity analysis: the initial effective contact rate (β_s), the fraction of immune individuals (ε) and the duration of infectiousness ($1/\gamma$). Red dots indicate the mean values of the response in each bin (bin size = 30).

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

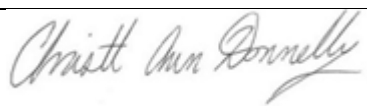
Title of Paper	Estimating vaccination threshold and impact in the 2017–2019 hepatitis A virus outbreak among persons experiencing homelessness or who use drugs in Louisville, Kentucky, United States
Publication Status	☑Published □ Accepted for Publication □Submitted for Publication □Unpublished and unsubmitted work written in a manuscript style
Publication Details	**Dankwa, E.A.**, Donnelly, C.A., Brouwer, A.F., Zhao, R., Montgomery, M.P., Weng, M.K. and Martin, N.K., 2021. Estimating vaccination threshold and impact in the 2017–2019 hepatitis A virus outbreak among persons experiencing homelessness or who use drugs in Louisville, Kentucky, United States. *Vaccine 39*(49), 7182-7190. doi: https://doi.org/10.1016/j.vaccine.2021.10.001

Student Confirmation

Student Name:	Emmanuelle A. Dankwa		
Contribution to the Paper	adapted the model, organized data, performed all analysis, interpreted results, performed all visualization, wrote initial version of manuscript, co-revised manuscript, documented code, created repository for code and data		
Signature 		Date	15 March 2022

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Christl A. Donnelly		
Supervisor comments Emmanuelle A. Dankwa was the leading author on this manuscript and made the substantial contributions detailed above		
Signature 	Date	16 March 2022

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 4 Stochastic modelling of African swine fever in wild boar and domestic pigs: epidemic forecasting and comparison of disease management strategies

This chapter presents the following manuscript which is undergoing peer review:

> **Dankwa, E.A**, Lambert, S., Hayes, S., Thompson, R.N., Donnelly, C.A. Stochastic modelling of African swine fever in wild boar and domestic pigs: epidemic forecasting and comparison of disease management strategies.

A statement of specific contributions by the author of this thesis is included in an authorship statement within the manuscript.

ABSTRACT

African swine fever (ASF), caused by the ASF virus, is highly virulent in domestic pigs and wild boar (*Sus scrofa*), causing up to 100% mortality. The recent epidemic of ASF in Europe has a serious economic impact and poses a threat to global food security. Unfortunately, there is no effective treatment or vaccine against ASFV, limiting the available disease management strategies. Mathematical models allow us to further our understanding of infectious disease dynamics and evaluate the efficacy of disease management strategies. The ASF Challenge, organised by the French National Research Institute for Agriculture, Food, and the Environment, aimed to expand the development of ASF transmission models to inform policy makers in a timely manner. Here, we present the model and associated projections produced by our team during the challenge. We developed a stochastic model combining transmission between wild boar and domestic pigs, which was calibrated to synthetic data corresponding to different phases describing the epidemic progression. The model was then used to produce forward projections describing the likely evolution of the epidemic under various disease management scenarios. Our model projected that incidence in pig herds was low and mostly driven by transmission from wild boar. Despite the interventions implemented, long-term projections forecasted persistence of ASFV in wild boar, and hence repeated outbreaks in domestic pigs. A key finding was that it is important to consider the timescale over which different measures are evaluated: interventions that have only limited effectiveness in the short term may yield significant long-term benefits. Our model faces several limitations, partly because it was developed in real-time. Nonetheless, the model informed understanding of the likely development of the epidemic and the efficacy of disease management strategies, proving that our framework could be used for a real epidemic, should the virus continue its spread in Europe.

Keywords: mathematical modelling; African swine fever virus; wildlife-livestock interface; spatial model; real-time analysis

4.1 INTRODUCTION

4.1.1 Background to ASF

ASF is a highly contagious viral disease capable of infecting all swine. It is caused by ASFV, a double-stranded DNA virus that is the sole member of the Asfarviridae family. ASFV is endemic across much of sub-Saharan Africa (Penrith et al., 2019). An ancient sylvatic cycle involving warthogs (*Phacochoerus africanus*) and soft ticks of the species *Ornithodoros* exists in eastern and southern Africa (Chenais et al., 2018; Costard et al., 2013; Dixon et al., 2019; Penrith et al., 2019). Juvenile warthogs are infected with the virus within the first few weeks of their lives when they are bitten by ticks living within their burrows. They develop a transient viraemia and remain infected for life but do not show any clinical signs of disease (Jori et al., 2013). The situation is very different in domestic pigs and wild boar (*Sus scrofa*) in which ASFV causes a range of clinical signs from sudden death to inapparent infection. Mortality rates range between 0-100% depending on the strain of the virus, the host, the viral dose, and the route of exposure (Blome et al., 2013, 2012; Costard et al., 2013). The existence of a carrier state following recovery from lower virulence strains has been proposed (Dixon et al., 2019).

Transmission routes for ASFV include direct contact between swine, contact with infected carcasses, meat products, fomites, the environment, and tick vectors (Costard et al., 2013; Guinat et al., 2016a; Pepin et al., 2020). A transmission cycle involving haematophagous flies has been suggested to occur in Europe, but its significance is still uncertain (Mellor et al., 1987; Olesen et al., 2018; Vergne et al., 2021). Transmission between wild boar and domestic pigs has been demonstrated (Dixon et al., 2019; Guinat et al., 2016a) and is thought to play an important role in the spread of ASFV. In high-biosecurity commercial pig farms where contact with wild boar has been excluded as a means of transmission, indirect transmission mediated by humans via fomites, for example, is usually considered the most likely route of introduction (Guinat et al., 2016a; Olesen et al., 2018). Infected animal carcasses have also been identified as a potential route of transmission and under certain conditions could pose a risk of infection for several months (Fischer et al., 2020). Wild boar have been shown to have

frequent contact with conspecific carcasses (Probst et al., 2017) and carcass-based transmission may be especially important in locations with low host density (Pepin et al., 2020).

The strains currently circulating in Europe have shown high virulence during experimental infection of domestic pigs and wild boar (Blome et al., 2020). Typically, death occurs within 7-10 days post-infection, but survival up to 36 days post-infection has been reported (Blome et al., 2020; Pietschmann et al., 2015). There is currently no approved treatment or vaccine against ASFV. Instead, disease management measures include culling on infected pig farms, disinfecting farm equipment, imposing restrictions on pork trade, conducting epidemiological surveillance of domestic pig and wild boar populations and managing wild boar populations (OIE, 2021).

4.1.2 Current situation in Europe

Although ASFV was eradicated from most of Europe in the 1990s, it was reintroduced to the continent via Georgia in 2007, most likely by importation of infected pork products (Beltrán-Alcrudo et al., 2008; Rowlands et al., 2008). Following its introduction, ASFV became established in the local wild boar population leading to further outbreaks in domestic pigs, with direct contact between infected wild boar and domestic pigs thought to play an important role in the spread of the disease (Dixon et al., 2019; Gogin et al., 2013; Oganesyan et al., 2013). In 2014, the first cases were reported in the EU (European Food Safety Authority, 2015). Since then, a series of outbreaks have been recorded resulting in major economic losses for the European pig industry (Danzetta et al., 2020; Guinat et al., 2016a). EU countries that have been affected by ASFV include Belgium, Bulgaria, Czech Republic, Estonia, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Poland, Romania, Serbia, and Slovakia (Blome et al., 2020). The first ASFV cases in Germany were reported in wild boar in September 2020 (Sauter-Louis et al., 2021b). Then, in July 2021, ASFV was confirmed in two domestic pig herds in the Brandenburg region bordering Poland (International Society for Infectious

Diseases, 2021). Most recently, in January 2022, ASFV was confirmed in wild boar in Italy (International Society for Infectious Diseases, 2022). Belgium and the Czech Republic are the only EU countries that have successfully eradicated ASFV following its introduction during the current epidemic. Outbreaks in these two countries were geographically localised and confined to the wild boar population. Disease management measures implemented included fencing off high-risk areas to limit movement of wild boar, active search and removal of wild boar carcasses and alterations to hunting patterns (Dellicour et al., 2020; Marcon et al., 2020). Whilst ASFV was eradicated from most of Europe in the 1990s, it has remained endemic in Sardinia since its introduction in 1978. The epidemiology of ASFV in Sardinia is complicated by the presence of free-roaming domestic pigs (FRPs) which have hindered previous eradication efforts. However, recent evidence highlighting the central role that FRPs play in maintaining ASFV and the implementation of a new eradication programme have led to marked reductions in the levels of ASFV in Sardinia and eradication appears achievable if the controls are maintained (Viltrop et al., 2021).

Whilst we focus on the situation in Europe here, it is worth noting that ASFV continues to circulate across sub-Saharan Africa (as of 4 March 2022 (OIE, 2022)) and, since its introduction to Asia in 2018, ASFV has spread to many Asian countries causing substantial economic impacts and posing a threat to vulnerable and endangered wild pig species (Luskin et al., 2020; Mighell and Ward, 2021; Tian and von Cramon-Taubadel, 2020).

4.1.3 Mathematical Models

The first mathematical models of ASFV were published in 2011 following introduction of the virus to Europe (Hayes et al., 2021). Since then, mathematical models have been widely utilised to further our understanding of the transmission dynamics of ASFV (see Hayes et al. (2021) for a recent review). Modelling studies have involved estimation of epidemiological parameters (Barongo et al., 2015; de Carvalho Ferreira et al., 2013; Guinat et al., 2018, 2016b; Gulenkin et al., 2011; Hu et al., 2017; Lange and Thulke, 2017; Loi et al., 2020; Nielsen et al.,

2017; Pietschmann et al., 2015; Shi et al., 2020), investigation of transmission dynamics in particular species (Halasa et al., 2016a, 2019; Mur et al., 2018; O'Neill et al., 2020; Pepin et al., 2020, 2021; Taylor et al., 2021), exploration of the role of vectors in transmission (Vergne et al., 2021) and assessment of the potential impacts of interventions (Barongo et al., 2016; Croft et al., 2020; Gervasi et al., 2020; Halasa et al., 2016b; Lange, 2015; Lange et al., 2018; Lange and Thulke, 2015; Lee et al., 2021; Taylor et al., 2021; Thulke and Lange, 2017).

Single-species models are used most frequently, despite the important role of between-species transmission of ASFV (Taylor et al., 2021). Domestic pig models may incorporate within-herd and/or between-herd transmission whilst wild boar models are frequently individual-based spatially structured models that incorporate existing knowledge of wild boar demography. Many models of ASFV transmission use the classic Susceptible-Exposed-Infectious-Recovered structure (Guinat et al., 2018; Halasa et al., 2019), with some including an extra class for infectious/non-infectious carcasses (Pepin et al., 2021). Other variations include using a Susceptible-Latent-Subclinical-Clinical-Removed structure (Halasa et al., 2016a), in which the infectious stage is split into sub-clinical and clinical stages, and a Susceptible-Infectious-Carrier-Removed structure (O'Neill et al., 2020) which incorporates the possibility that pigs that have recovered from lower virulence strains continue to carry the virus.

4.1.4 Challenge overview and objectives

Motivated by the ongoing global spread of ASFV, the French National Research Institute for Agriculture, Food and the Environment (INRAE) organised the ASF Challenge to expand the development and application of mathematical methods for ASF epidemic forecasting and to better understand the strengths and limitations of different modelling approaches (Picault et al., 2021). An additional goal was to improve the readiness of modelling teams and hence their ability to advise policy makers in a timely manner when faced with emerging epidemic threats such as ASF (https://www6.inrae.fr/asfchallenge/). Similar events in the past have led

to significant statistical and computational innovations for epidemic forecasting and have fostered fruitful collaborations between research teams and policy makers (Johansson et al., 2019; McGowan et al., 2019; Viboud et al., 2018).

The challenge took place from August 2020 to January 2021. It was comprised of three phases, describing different stages of a simulated ASF epidemic on a fictional island (Merry Island). In each phase, simulated data regarding the number and location of reported infected pig herds and wild boar – as well as simulated data describing movements of pigs exchanged or traded between herds – were provided. Modelling teams were then asked to provide projections, informed by their own analyses, of the course of the epidemic over a specified future time period (generally 30 days) incorporating specified disease management measures where indicated. Day 0 represented the date of the first reported case of ASF on Merry Island. In phase 1, simulated data from days 0 to 50 (period 1) were provided, with projections requested for days 51 to 80 (period 2). In phase 2, simulated data from periods 1 and 2 were provided to underpin projections for days 81 to 110 (period 3). Finally, in phase 3, simulated data were provided from days 0 to 110 with two sets of projections requested: one short-term set for days 111 to 140 and one longer-term set for days 111 to 230.

Here we provide details of the model and the associated projections that were produced by our team in each phase of the challenge. To facilitate reproducibility, we have made all code and relevant data files used for this analysis freely available on GitHub: https://github.com/emmanuelle-dankwa/ASF_model.

4.2 MATERIALS AND METHODS

4.2.1 Synthetic data provided by the challenge coordinators

4.2.1.1 Demographic data

In the challenge, a simulated ASF epidemic on Merry Island – a fictional island located in the North Atlantic Ocean with area 144,208 km^2 divided into 25 administrative units – was considered. There were 4,532 registered pig farms on the island. Farms were classified as either backyard or commercial, and either outdoor or indoor. For each farm, additional information provided included the size of the pig herd, its geographical coordinates, its production type (farrow-to-wean, farrow-to-finish or finishing), and whether it belonged to the same producer as other farms (i.e., "multisite farms" comprising several geographically distinct farms). Farms belonging to the same producer were expected to be more epidemiologically connected to each other than to any other farm. In addition to these registered farms, there was an unknown number of small, unregistered farms. Eight unregistered farms were identified in total from period 1 to 3 and were added to the 4,532 registered farms: two farms were identified because they became infected, four farms were identified because they became part of a surveillance zone, and two others were identified because they were culled preventively. Registered movements of pigs in the trade network were also provided for the two months before the first detected infected pig herd (suspected on 8th July 2020, referred to as "day 0" during the challenge, and confirmed three days later), with the day at which each movement occurred, the source and destination herds, and the number of pigs traded.

Data on the hunting bag (number of boar hunted during a hunting season) in each administrative unit in 2019 (the calendar year before the detection of ASFV) were also provided (260,675 hunted wild boar in total). It was estimated by the ASF Challenge coordinators that around 50% of wild boar are shot during a hunting season, giving a rough wild boar population size estimate of 521,350 for Merry Island.

4.2.1.2 Epidemiological data

In each of the three phases of the challenge, incidence data for both pigs and wild boar were provided. This synthetic epidemiological data originated from an epidemiological model developed by the challenge coordinators, which remained unknown to the participating teams at the time of the challenge. Briefly, the model used to produce synthetic epidemiological data was a discrete-time, stochastic, spatially explicit and agent-based model. Agents were pig herds and individual wild boar, each with their specific location. Transmission pathways included transmission between wild boar, transmission between pig herds (via introduction of infected pigs through trade movements or indirect contact with infectious farms), and transmission from wild boar to pig herds and vice versa. Trade movements were determined based on a temporal directed graph between farms. All other transmission pathways were modelled based on an exponential transmission kernel, assuming that the probability of transmission between two farms/boar decreased with increasing distance between them. The synthetic data corresponded to one stochastic simulation of the model. Further details on the original model and data generation can be found in the first article in this special issue (Picault et al., 2021).

For pig herds, the data provided included the identity of each herd in which infection was detected, with the mode of detection as well as the dates of suspicion, confirmation, and culling. The first reported infected herd was herd 3594, and two other infected herds were detected during period 1 (days 0 to 50). Nine new infected herds were confirmed during period 2 (days 51 to 80), followed by 14 others during period 3 (days 81 to 110), resulting in a total of 26 detected infected pig herds from period 1 to period 3 (days 0 to 110). For wild boar, the locations of tested wild boar found each day through passive surveillance, active surveillance and hunting were provided, as well as the date of confirmation and the test results (positive or negative). A total of 2,984 detected infected wild boar were reported from period 1 to period 3 (days 0 to 110). Although the epidemiological data provided were synthetic, we later refer to

them as “observed data” to clearly distinguish the results of our model from the data to which they were compared.

Contextual information on disease management measures in both pig herds and wild boar was also provided by the challenge coordinators in each phase (see details in Model section below).

4.2.2 Model

We modelled ASFV transmission on Merry Island using a model combining a stochastic, spatial Susceptible-Infectious-Post-infectious (SIP) model for transmission across the island via wild boar and a stochastic, metapopulation Susceptible-Latent-Infectious-Recovered-Deceased (SLIRD) model for pig herds in keeping with the known epidemiology of ASF. The model included transmission from wild boar to pig herds but did not account for transmission from pig herds to wild boar since we saw no evidence, based on the data provided, of such transmission. All simulations and analyses were performed with R version 4.1.2 (R Core Team, 2022).

4.2.2.1 Transmission

4.2.2.1.1 Wild boar

Transmission via wild boar was modelled using a stochastic SIP model with an exponential spatial dispersal kernel. Due to the large number of wild boar on the island and for computational efficiency, only wild boar within a defined area around the initial detected cases in wild boar and pig herds were considered (see Appendix A for details). Wild boar live in matrilineal groups, with reported mean group sizes of 4-8 individuals (Maselli et al., 2014; Pepin et al., 2020; Podgórski et al., 2014). These groups are typically composed of adult and subadult females and their offspring (Pepin et al., 2020). Reports of home range sizes vary

but are typically estimated as between 2–10 km^2 (Janeau et al., 1995; Keuling et al., 2008; Podgórski et al., 2013). To represent this social structure, we divided the area into 2500 rectangular patches and assumed that the infectious pressure on every susceptible wild boar within a specific patch was identical. Thus, patches were treated as model units. The area per patch was 16 km^2 in phase 1, and 7.5 km^2 in phases 2 and 3. Patch dimensions were reduced in the latter phases to enable a finer-grained resolution for more accurate results.

In addition to tracking the health states of patches, we also tracked the health state of individual wild boar in a patch to enable the assessment of individual-focused control measures. Each day, a patch was either susceptible *S*, infectious *I*, or post-infectious *P*: a patch was considered infectious if at least one boar within the patch was infectious; a patch was considered susceptible if there were no infectious boar within the patch; and a patch was considered post-infectious if all boar in the patch were carcasses that were no longer infectious (see Table A4-6, Appendix A). We assumed that infected wild boar were infectious immediately after infection. As we assumed 100% mortality in wild boar for this strain of ASFV (Blome et al., 2013, 2012), the infectious wild boar transitioned to become a carcass after a fixed period of 14 days after infection (ASF Challenge coordinators, 2020; Blome et al., 2013; Pietschmann et al., 2015) if it had not been removed (by hunting) before then. The period from infection to death was set to 14 days based on the upper limits of estimates for infection to onset of infectiousness (4-5 days) and the duration of infectiousness in live boar (2-9 days) (ASF Challenge coordinators, 2020; Hayes et al., 2021; Pietschmann et al., 2015). Carcasses could not be re-infected but remained infectious for 90 days after death (Fischer et al., 2020) if not removed by surveillance before then. Thus, a carcass could either be in the I or P state, depending on whether the boar died less than/more than 90 days ago.

For each patch, the number of new infections at each daily time step was chosen stochastically assuming a Poisson distribution with rate equal to the total “infectious pressure” being exerted on the patch. The infectious pressure on a patch j (accounting for the number of susceptible

boar in patch j that are available for infection) exerted by an infected patch i at day t was given by

$$\beta_{ij}(t) = \beta \exp\left(\frac{-d_{ij}}{\alpha}\right) I_i(t) S_j(t), \quad (4.1)$$

where $\beta > 0$ determines the overall infection rate, $\alpha > 0$ is the scale parameter of the exponential dispersal kernel, $\exp(.)$, d_{ij} is the Euclidean distance between the midpoints of patches i and j (measured in km), $I_i(t)$is the number of infected wild boar in patch i on day t and $S_j(t)$ is the number of susceptible wild boar in patch j on day t. The total infectious pressure on patch j at day t, $\omega_j(t)$, was computed as the sum of the infectious pressures exerted on j:

$$\omega_j(t) = \sum_i \beta_{ij}(t) \quad (4.2)$$

The number of new infections in boar in each patch was determined by treating the infectious pressures (unit: day^{-1}; see Table 3-1) as Poisson rates:

$$n_j \sim \text{Pois}(\omega_j) \quad (4.3)$$

where n_j is the number of new infections in patch j. The specific wild boar infected in each patch on any day were randomly selected from the remaining susceptible boar in the patch. If n_j was greater than the number of susceptible wild boar remaining in the patch, all susceptible boar in the patch became infected.

4.2.2.1.2 Pig herds

The transmission model for pig herds considered individual pigs by describing the numbers of animals in each compartment within a herd, and the flows between the different compartments. Each herd was considered as a homogeneous, random-mixing population, not accounting for any within-herd structure. Health statuses were susceptible S, exposed and

pre-infectious E, infectious I, and immune (recovered) R. In addition, we distinguished infectious animals into subclinical (Isc) and clinical stages (Ic).

The force of infection λ_i (equation (4.4)) exerted on susceptible pigs in herd i at time t was calculated based on: (1) the proportion of infectious pigs and of infectious residues from dead pigs within herd i (Fischer et al., 2020; Halasa et al., 2016a); (2) the local spread due, for example, to shared material and fomites from neighbouring infected herds within a 2 km radius (Andraud et al., 2019; Halasa et al., 2016c); and (3) for outdoor herds, the number of infectious wild boar (alive and carcasses) in each infected patch and the distance between the herd and each infected patch using an exponential kernel:

$$\lambda_i(t) = 1 - \exp\left(-\beta^{PH} * \frac{I_i(t) + D_i(t)}{N_i(t)} - \sum_j \left(\frac{\rho}{d_{ij}} * \frac{I_j(t) + D_j(t)}{N_j(t)} \right) - \Omega_i(t) \right) \quad (4.4)$$

where $\beta^{PH} > 0$ is the transmission rate within pig herds (Table 4-1), $I_i(t)$ is the total number of infectious pigs in herd i (subclinical and clinical cases), $D_i(t)$ is the contribution of residues from dead pigs in herd i to transmission, $N_i(t)$ is the total number of live pigs in herd i, $\rho > 0$ is the transmission rate by local spread (Table 4-1), d_{ij} is the distance between herds i and j, and $\Omega_i \geq 0$ is the infectious pressure exerted by wild boar on outdoor herd i ($\Omega_i = 0$ for indoor herds).

Susceptible pigs S that acquired infection moved to the exposed pre-infectious compartment E, where they stayed during the pre-infectious period (with average duration δ), and then moved into the infectious compartment I, where they stayed during the infectious period (with average duration γ). Infectious pigs were first subclinical (Isc) during the subclinical period (average duration ϕ), and then became clinical (Ic) for the rest of the duration of the infectious period. Infectious pigs either survived and became immune and moved into the recovered compartment R, or died with probability μ (Halasa et al., 2016a). Although dead pig carcasses were assumed to be removed, dead pigs entered the compartment D to represent residues from dead animals contributing to transmission. These residues stayed in the environment

during the mean lifetime of the virus in residues (with average duration τ). We assumed lifelong immunity in the R compartment. Parameter values are given in Table 4-1.

In addition to transmission by local spread, between-herd transmission was explicitly driven by the modelling of animal movements in the trade network (Brooks-Pollock et al., 2014), where animals in each compartment could enter or leave a herd, representing opportunities of contacts and transmission between individuals from different herds (see "Movements" (Section 4.2.2.3) and Appendix A for details).

For outdoor herds, the total infectious pressure exerted by wild boar on herd i at time t was given by:

$$\Omega_i(t) = \sum_k \beta \exp\left(-\frac{d_{ik}}{\alpha}\right) I_k(t) \quad (4.5)$$

where $\beta > 0$ and $\alpha > 0$ are the same parameters as in the wild boar model (equation (4.1) and Table 4-1), d_{ik} is the Euclidean distance between herd i and the centre of infected wild boar patch k, and I_k is the number of infectious boar (alive and carcasses) in patch k.

4.2.2.2 Population dynamics

4.2.2.2.1 Wild boar

We assumed a constant population size among wild boar in the absence of hunting, carcass removal and ASFV-related mortality. Across Europe, wild boar breeding is typically seasonal, commencing in late autumn/early winter with peaks in November/December. Following a gestation period of 115 days, peak birthing of piglets occurs between February and April (Alves da Silva et al., 2004; Podgórski and Śmietanka, 2018; Rosell et al., 2012; Sabrina et al., 2009). Therefore, we did not account for births since the period over which projections were required (27th August 2020 until 23rd February 2021, corresponding to days 51-230) was not within the known peak birthing period for wild boar. We also did not account for natural mortality due

to the short duration of the projection period relative to the average lifespan of wild boar (Herrero et al., 2008; Jezierski, 1977) and given that the predominant causes of mortality over the projection period were likely to be hunting and ASFV, both of which our model accounted for. Thus, the population of wild boar, both dead and alive, remaining in the landscape at any time could only be decreased through removal by hunting (all hunted boar were removed from the landscape) or via surveillance. For wild boar, two main types of surveillance were carried out in the island: (1) passive surveillance, which involved the removal and reporting of found wild boar carcasses, and (2) active surveillance, which involved active search for wild boar carcasses around already detected infected carcasses.

4.2.2.2.2 Pig herds

We assumed a constant population size in each herd in the absence of ASFV-related mortality, with two population dynamics processes depending on the production type of the herds: birth of susceptible pigs in farrow-to-wean and farrow-to-finish herds, and animals sent to the abattoir in finishing and farrow-to-finish herds. Natural mortality was not accounted for.

Thus, in farrow-to-wean and farrow-to-finish herds, the number of pigs leaving the herd (outgoing movements) was compensated by the entry of the same number of pigs (susceptible only). On the other hand, in finishing and farrow-to-finish herds, the number of pigs entering the herd (ingoing movements) was compensated by the same number of pigs leaving to the abattoir. Such processes were considered an acceptable approximation of the population dynamics of the pig herds given the batch system used in swine production and the timescale of the simulations.

4.2.2.3 Movements

4.2.2.3.1 Wild boar

In phase 3, to make the model more representative of wild boar movement dynamics, we implemented a threshold – the maximum infection range (MIR). This was chosen to be 8 km to reflect reports of the maximum distance travelled and the estimated home range of wild boar (Janeau et al., 1995; Podgórski et al., 2013). Consequently, an infectious wild boar in a patch A could infect other susceptible wild boar in the same patch or in another patch B if the centre of B was situated less than 8 km from the centre of A (see Figure A4-7, Appendix A). Similarly, a pig herd i could only be infected by infectious wild boar in a patch k whose centre was located less than 8 km from herd i (i.e., k such that $d_{ik} \leq M\mathrm{IR}$ in equation (4.5)). In phases 1 and 2, no threshold was set for this maximum distance (Table 4-2).

4.2.2.3.2 Pig herds

Data on pig movements up to day 50, 80 and 110 (for phases 1, 2 and 3, respectively) were provided by the challenge coordinators and therefore these pig shipments between herds were considered as deterministic (day of the shipment, source and destination herds, number of pigs shipped). Pig movements from day 51, 81 or 111 onwards (for phases 1, 2 and 3, respectively) were projected using exponential random graph models (ERGMs) to determine a pair between a source herd and a destination herd (Relun et al., 2017), and using generalized linear models (GLMs) with zero-truncated negative binomial distribution to determine the number of pigs exchanged (more details are provided in Appendix A).

4.2.2.4 Initial conditions

4.2.2.4.1 Wild boar

The initial size and spatial distribution of the wild boar population in each patch was estimated using the hunting bag data and hunting rate estimates for 2019. At each phase, the model was seeded with some observed ASFV infections among wild boar, as explained in Appendix A.

4.2.2.4.2 Pig herds

Pig herds were distributed according to the coordinates provided. We considered all known herds in the island. The number of known pig herds changed at each phase as the number of unregistered farms identified increased (see details in Data section above): 4533 for phase 1, 4537 for phase 2 and 4540 for phase 3. Each herd was initialized with susceptible pigs based on its size provided by the challenge coordinators. In all analyses, ASFV was introduced in pig herd 3594 (the first detected infected pig herd) by replacing a susceptible pig (*S*) by an exposed pre-infectious pig (*E*) at day -31 (i.e., 31 days before day 0), giving a median suspicion date in the model at day 0 and a median detection date at day 3, as observed in the data.

4.2.2.5 Disease management measures and model implementation

4.2.2.5.1 Wild boar

4.2.2.5.1.1 Fence

As part of the measures to curb the spread of the virus out of the forest area, a 300 km rectangular fence was set up around the area where ASFV had been initially detected in wild boar (ASF Challenge coordinators). The fence was operational from day 60 and was assumed to have no impact on transmission before this date. In the models we presented for phases 1

and 2, the fence was assumed to be 100% effective from day 60. However, by phase 3, the observed locations of some newly detected infections suggested that the fence was not fully effective. Thus, we allowed for a “leaky” fence in all directions, such that ASFV could be transmitted between two patches on opposite sides of the fence if the distance between their centres was less than or equal to half the MIR (Table 4-2 and Appendix A). This also applied to transmission from wild boar to pig herds: in phase 3, only wild boar patches situated on the same side of the fence as outdoor herd i and satisfying $d_{ik} \leq M\mathrm{IR}$ and wild boar patches situated on the other side of the fence and satisfying $d_{ik} \leq \mathrm{MIR}/2$ were considered in the computation of the infectious pressure (equation (4.5)). In phases 1 and 2, only wild boar patches situated on the same side of the fence as herd i were considered.

4.2.2.5.1.2 Normal hunting pressure

The “normal hunting pressure” scenario involved hunting according to the usual hunting rates for a typical hunting season, which corresponds to a hunting rate of 50% of the wild boar population from day -36 (i.e., 36 days before day 0) to day 204 (an 8-month period) and at a uniform rate over the period. This measure was maintained outside the fence and buffer zone for the entire duration of the projection period. Under this scenario, both surveillance types were possible and 20% of all hunted wild boar were tested.

4.2.2.5.1.3 Increased hunting pressure

Within the fenced area and a 15 km-buffer zone around the fence, an “increased hunting pressure” management strategy was implemented, beginning at day 60. This involved applying a hunting rate of 90% of the wild boar population (much higher than that observed in a typical hunting season) from day 60 to day 120 (2 months), at a uniform rate, to decrease wild boar density and thus slow the spread of ASFV (ASF Challenge coordinators). Under this

measure, active surveillance ceased within the fence and the buffer zone, given “the potential risks posed by hunts” (ASF Challenge coordinators). However, passive surveillance was still possible. Moreover, 100% of all hunted wild boar were tested.

4.2.2.5.2 Pig herds

4.2.2.5.2.1 Baseline regulatory interventions

According to the challenge coordinators, disease management measures defined by European regulations were immediately implemented in Merry Island in response to the epidemic, when the first detected infected pig herd was confirmed (day 3). These regulations were originally established by the European Union (European Commission, 2002) and are now described in the new “Animal Health Law” (European Commission, 2016) and its supplement as regards rules for the prevention and management of diseases such as ASF (European Commission, 2020a).

Based on the description of the disease management measures provided by the challenge coordinators, the following measures were implemented in our model: (1) suspected pig herds were confirmed infected three days after suspicion, assuming perfect ASFV detection tests; (2) all herds with confirmed infection were culled the day after confirmation (four days after suspicion), implemented in our model by setting all compartments to zero (including residues from dead pigs, i.e., assuming cleaning and disinfection were effective immediately); (3) after a herd was confirmed infected, a protection zone (3 km radius for 40 days) and a surveillance zone (10 km radius for 30 days) were defined, and at-risk herds that traded pigs with infected herds (ingoing or outgoing movements) within the previous three weeks were traced; (4) all movements of pigs (ingoing or outgoing) were banned for 40 days in protection zones and at-risk herds, and for 30 days in surveillance zones; (5) awareness of farmers about ASF was improved in surveillance and protection zones, as well as for at-risk herds; (6) repopulation of

a culled herd was allowed 50 days after culling (except if the herd was still in a protection or surveillance zone), assuming that all pigs used for repopulation were susceptible.

In our model, suspicion of a herd was assumed to occur when two conditions were met: (1) when the mortality rate caused by ASFV during the previous 14 days in the herd was more than 6% (Andraud et al., 2019); and (2) the number of clinical or dead animals in the herd reached a minimum value of five during the previous 14 days (Halasa et al., 2016a). The minimum number of clinical or dead animals was introduced to represent more accurately the probability of detecting abnormal events, especially in small herds, where only one death could make the mortality rate exceed the threshold (Halasa et al., 2016a). Increased awareness of farmers in protection and surveillance zones and in at-risk herds was represented in our model by reducing the minimum number of clinical or dead animals required for detection to one (Halasa et al., 2016a).

4.2.2.5.2.2 Additional interventions (phases 2 and 3)

During phase 2, additional disease management strategies in pig herds were incorporated into the model as asked by the challenge coordinators (see Appendix A for more details): (1) preventive culling of all herds in a protection zone ("cullPZ"); (2) increasing the size of the surveillance zone from 10 km (the standard radius used) to 15 km ("incrSZ"); (3) preventive culling of all pig herds located at less than 3 km from positive wild boar carcasses ("cullWB"); and (4) preventive culling of all at-risk herds ("cullTR"). Those additional interventions were implemented in forward projections during phase 2, i.e., from day 81 to day 110. During phase 2, pig herds preventively culled before detection in scenarios cullPZ, cullWB and cullTR were not tested (Table 4-2). Culling was assumed to take place 24 hours after the event triggering the intervention, as for confirmed herds in baseline interventions.

During phase 3, cullWB was implemented starting day 90 according to the challenge coordinators. Preventive culling happened 5-7 days after a wild boar case was confirmed, and tests were performed rapidly in all culled herds, providing results the day after (Table 4-2).

4.2.3 Analysis considerations

4.2.3.1 Comparison of scenarios

Using model simulations, we compared epidemic outcomes (number and locations of cases) under the range of scenarios discussed, to determine the effectiveness of each at limiting the epidemic.

4.2.3.2 Probability of epidemic fade out by day 230

A key question of interest posed by the challenge coordinators in phase 3 was how likely the epidemic was to fade out by day 230 given the following conditions: (1) a cessation in increased hunting pressure at day 120 (due to a reduction in reported incidence), (2) end of the hunting at day 204 (usual last day of hunting on the island), and (3) possibility of passive discovery of wild boar carcasses beyond day 204. To estimate the probability of fade-out, we simulated from our model under these conditions and computed the proportion of simulations having at least one case by day 230. This was done for both wild boar and pig herd populations.

4.2.4 Parameter estimation

The wild boar model was calibrated using ABC (Beaumont et al., 2002). In phase 1, the type of ABC algorithm employed was ABC-SMC (Minter and Retkute, 2019; Toni et al., 2009) while in phases 2 and 3, the ABC rejection algorithm (see Toni et al. (2009)) was employed.

In all phases, the transmission parameter β was estimated. In phases 2 and 3, to improve the runtime of the estimation algorithm, the scale parameter α of the dispersal kernel was fixed based on its estimated value at Phase 1 and results from some trial runs of the model.

The wild boar summary statistics used in the ABC estimation were: (1) the daily number of detected infected wild boar, and (2) the area of the minimum convex polygon enclosing the locations of infected patches. By choosing these summary statistics, we sought to make our model fit reflect well both the size of the observed epidemic in wild boar and the spatial extent of the epidemic. The parameter values producing simulated summary statistics closest to the observed summary statistics were retained for model predictions. The tolerances used in the ABC were chosen based on an iterative sequence of trial runs which compared simulated model outputs to the observed data.

Parameters exclusively associated with the pig herd model were derived from published estimates (Table 4-1). After a graphical comparison between simulated and observed daily and cumulative numbers of detected infected pig herds over time at each phase, the same transmission parameter values for transmission from wild boar to pigs in outdoor herds were used as those calibrated for wild boar-to-wild boar transmission (α and β).

4.2.5 Simulations and outputs

During the challenge, the number of stochastic repetitions decreased from 500 for phase 1, to 72 for phase 2 and 32 for phase 3 because of runtime constraints imposed by real-time analysis. However, the results presented in this paper were expanded to include 500 stochastic repetitions for each phase.

For the wild boar model, the simulated period was from day 1 in phases 1 and 2, but from day 60 in phase 3, due to computational constraints. For the pig herd model, the simulated period was from day –59 (when data on pig movements started) in all phases. Each repetition corresponded to a given set of parameter values retained by ABC (α and β in phase 1, only β in phases 2 and 3). In addition, model stochasticity was driven by drawing events randomly from probability distributions.

For wild boar, model outputs across all phases were the daily number of detected/infected wild boar and the locations of infected wild boar patches. In phases 2 and 3, additional outputs were infected wild boar locations. For pig herds, model outputs were the daily number of suspected/confirmed/infected herds, and the probability of suspicion/detection/infection for each herd (expressed as the proportion of simulations where a given herd was suspected/detected/infected). Model outputs were expressed as the median of the simulations and the associated 95% equal-tailed credible interval (CrI), using the 2.5% and the 97.5% percentiles of the simulations as lower and upper bounds of the 95% CrI, respectively.

4.2.6 Sensitivity analysis

To assess the sensitivity of our model to changes in parameter values and assumptions, we conducted sensitivity analyses.

First, we assessed the influence of the MIR, the scale parameter of the dispersal kernel α, the duration of infectiousness in wild boar carcasses, and the duration of infectiousness in live boar on the daily number of infections and detected cases in wild boar and pig herds. For this analysis, we focused on phase 3, from day 60 to 110, corresponding to the period over which the phase 3 model was fitted. We chose to use the phase 3 model since it is the only model which incorporates the MIR. We considered values ranging from 2 km to 20 km for the MIR; from 0.6 km to 1.2 km for α; from 10 days to 130 days for the duration of infectiousness of wild boar carcasses (Fischer et al., 2020; Mazur-Panasiuk and Woźniakowski, 2020); and from 5 days to 14 days for the infectious period in live boar (Gervasi et al., 2020; Gervasi and Guberti, 2021; Halasa et al., 2019; Hayes et al., 2021; O'Neill et al., 2020; Pepin et al., 2020). As intervention measures in pig herds were tested during phase 2, we employed the phase 2 model to test the sensitivity of the epidemic trajectory among pig herds to changes in values of parameters governing these interventions; namely, the delay (in days) for the implementation of culling, and the minimum number of clinical or dead pigs required for detection when awareness is increased. For both parameters, we considered three values: 1

(baseline), 4 and 7, for the delay before the implementation of culling; and 1 (baseline), 3 and 5, for the minimum number of clinical or dead pigs required for detection when awareness is increased. These values were chosen to reflect imperfect surveillance and delays in the implementation of culling pig herds, which are likely to occur, especially in the early phase of an ASF outbreak. The outcomes measured in this analysis were the final number of infected pig herds for each disease management strategy. As we assumed no transmission from pig herds to wild boar, we did not assess the influence of these parameters on outputs for wild boar.

For each parameter X assessed, we compared model outcomes under different values of X, including the baseline value employed in our model. For each value of X studied, the sensitivity analysis involved running 100 stochastic repetitions of the model. In these model simulations, all other parameters and all model assumptions, including control measures, were as in the baseline model. We then computed the median and 95% CrIs for each outcome across the 100 stochastic repetitions.

4.3 RESULTS

4.3.1 Parameter estimation and model fit

Parameter estimates at each phase are provided in Table 4-1. Our model fitted well to the observed temporal and spatial dynamics of the epidemic (Figure 4-1 and Table 4-3; Figure B4-10, Appendix B). To evaluate the ability of our model projections to capture the observed dynamics of the epidemic, we also compared model projections for the detected number of cases in wild boar and pig herds during phase 1 (up to day 78) and phase 2 (up to day 110) to the observed data provided by the challenge coordinators after these two phases were completed. We were not provided with observed data corresponding to the projection period for phase 3 (beyond day 110), thus precluding comparison of our projections in that phase with observed data. The 95% CrIs for the number of detected infected pig herds and wild boar captured the observations in phase 1 (Table 4-4). In phase 2, the 95% CrIs for the

number of detected infected pig herds captured the observed number although the corresponding statistic for the number of detected infected wild boar did not; the median estimate for wild boar overestimated the observed number by 7.7% (Table 4-4).

Although ASFV was seeded in both wild boar and pig herds, pig herd incidence was driven by the wild boar epidemic, as illustrated in Figure B4-11, Appendix B. Indeed, in the absence of ASFV transmission from wild boar to pig herds, the cumulative number of detected infected pig herds up to day 230 remained very low (median: 2, 95% CrI: (2–7)).

4.3.2 ASF management strategies

4.3.2.1 Fencing and increased hunting pressure

The challenge coordinators were interested in the difference in effectiveness between the scenario involving the implementation of the fence alone and that involving the implementation of the fence combined with increased hunting pressure within the fence (and from phase 2 also in the buffer zone; see Table 4-2). For all phases, we report model projections of the daily number of detected infected wild boar and the daily number of detected infected pig herds under the increased hunting pressure and normal hunting pressure scenarios (Figure 4-2). In general, our model projections showed a better efficacy of the combination of fence with increased hunting pressure in comparison with fence and normal hunting pressure (Figure 4-2, Table 4-3).

In wild boar, for phase 1, there were 90% more detected cases under increased hunting pressure compared to normal hunting pressure (Table 4-3). However, for phases 2 and 3, there were more cases under normal hunting pressure than under increased hunting pressure: the projected median estimates for normal hunting pressure were 1.8% and 131% higher than corresponding estimates for increased hunting pressure, for phases 2 and 3, respectively (Table 4-3). The projected number of detected infected pig herds was very similar for both

scenarios in phase 1 (Figure 4-2, Table 4-3). In phase 2, increasing the hunting pressure led to 9% fewer detected infected pig herds on average compared to the normal hunting pressure scenario (Figure 4-2, Table 4-3). It was only in phase 3 that increasing hunting pressure had a strong impact, with 46.5% fewer detected infected pig herds on average compared to the normal hunting pressure scenario (Figure 4-2, Table 4-3).

4.3.2.2 Additional interventions in pig herds

The model projections showed that culling all pig herds in protection zones ("cullPZ"), culling all herds that have traded pigs with an infected farm less than three weeks before detection ("cullTR"), or increasing the size of the surveillance zone from 10 km to 15 km ("incrSZ") all had a negligible impact on the number of infected herds compared to the baseline management strategies in pig herds (Figure 4-3). However, culling of all pig herds located less than 3 km from positive wild boar carcasses ("cullWB") led to 4 fewer infected herds on average, compared to the baseline management strategies, a 18.5% reduction over a 30-day period (Figure 4-3). This reduced number of infected herds was obtained by culling 67 more herds on average compared to the baseline management strategies, a 122% increase over a 30-day period.

4.3.3 Probability of epidemic fade-out by day 230

Our model simulations showed the persistence of the virus within the population by day 230 in all projections (Figure 4-4), given the new disease management measures introduced at day 120. The estimated daily numbers of detected cases beyond day 120 were generally lower than had been observed at the start of the increased hunting pressure activities (day 60) and followed a steady trend up to day 204, after which even fewer cases were detected daily, given the end of the hunting season. The probability of fade-out in pig herds depended on the

probability of fade-out in wild boar (Figure B4-11, Appendix B). Indeed, as long as the virus persists within the wild boar population, further infections of pig herds are to be expected.

4.3.4 Sensitivity analysis

There was no substantial difference between the trajectories for the daily number of detected infected boar corresponding to MIR values of 8 km, 14 km and 20 km, although there was a marked difference between these trajectories and that corresponding to a MIR of 2 km (Figure 4-5A). Increasing the MIR from 2 km to 8 km resulted in a 187% (95% CrI: 160%–206%) increase in the number of detected infected boar within the period considered (days 60–110), whereas increasing from 8 km to 14 km resulted in only a 2.02% (95% CrI: -5.4%–12%) increase (Table 4-5). Similarly, the number of detected infected pig herds increased by 100% (95% CrI: 61%–152%) when increasing the MIR from 2 km to 8 km but did not change further for values above 8 km. Similar results as for detected infected boar and pig herds were obtained when considering the number of infections (detected or not: Figure C4-12A, Appendix C).

For the duration of infectiousness in boar carcasses, we observed a similar trend where the results changed only for the smallest parameter value. Indeed, there was no notable difference between the median trajectories corresponding to the 50-day, 90-day and 130-day durations, but the trajectory corresponding to a 10-day duration was slightly lower starting from day 70 (Figure 4-5C). However, CrIs corresponding to estimates for all parameter values were largely overlapping (Figure 4-5C, Table 4-5). Similar results were observed for the number of infections and the number of detected infected pig herds (Table 4-5 and Figure C4-12C, Appendix C).

The trend was however different for the scale parameter of the dispersal kernel, α (Figure 4-5B), and the duration of infectiousness in live boar (Figure 4-5D). For these parameters, larger parameter values resulted in larger values of the daily number of detected

infected wild boar. This was especially true for the scale parameter α, for which the number of detected infected wild boar increased by 190% (95% CrI: 169%–212%) from 0.6 km to 1.2 km, and the number of detected infected pig herds increased by 140% (95% CrI: 82%–204%) from 0.6 km to 1.2 km (Table 4-5). A similar trend was observed for the number of infected wild boar (Figure C4-12B, Appendix C).

Decreasing the infectious period of live boar in the baseline model (14 days) by 4 days, 7 days, and 9 days resulted in a decrease of 16% (95% CrI: 10%–23%), 30% (95% CrI: 24%–37%) and 40% (95% CrI: 35%–45%), respectively, in the number of infected detected boar, compared to baseline (Figure 4-5D, Table 4-5). However, decreasing the infectious period in live boar led to no substantial changes in the number of infected wild boar (Figure C4-12D, Appendix C) nor in the number of detected infected pig herds (Table 4-5).

Finally, in the baseline management scenario, the number of infected pig herds was not sensitive to increasing the delay for the implementation of culling from 1 to 7 days and to increasing the minimal number of clinical or dead pigs required for detection from 1 to 5 (Figure 4-6). The efficacy of the four additional scenarios remained the same when changing the minimal number of clinical or dead pigs required for detection, with negligible impact except for culling pig herds less than 3km away from infected wild boar (Figure 4-6D and E). However, the number of infected pig herds increased with increasing delay before implementing culling for this latter scenario, from 35 (95% CrI: 27–49) for a one-day delay to 39 (95% CrI: 28–48) for a seven-day delay, levelling all scenarios to the same efficacy for a seven-day delay (Figure 4-6B and C).

4.4 DISCUSSION

We have developed a stochastic spatiotemporal model describing the transmission dynamics of ASFV in a multispecies context involving wild boar and domestic pigs. Our model captured the shape of the observed epidemic trajectory as well as its spatial characteristics (Figure 4-1,

Table 4-4; Figure B4-10, Appendix B) and was complex enough to allow for the incorporation of a range of disease management measures and for the estimation of their respective effects on the epidemic trend (Figure 4-3 and Figure 4-2, Table 4-3).

4.4.1 What do our results show and what do they mean?

4.4.1.1 Key point 1: Increased hunting pressure effective, long-term evaluation more beneficial

Toward informing recommendations for ASF management measures in the wild boar population – assumed to be the reservoir for ASFV in the island we considered – we evaluated the effectiveness of an increased hunting pressure scenario and a normal hunting pressure scenario, both including a fenced area to restrict wild boar movement beyond an identified epicentre. Our model results showed a superior efficacy associated with the increased hunting pressure scenario (Figure 4-2, Table 4-3). It is worthy of note that the benefits (reduction in the number of infected boar removed) realized under the increased hunting pressure scenario were more apparent in the longer term, in both the wild boar and pig herd populations (Figure 4-2). For wild boar, the benefit of increased hunting pressure could only be seen in phases 2 and 3 (Figure 4-2B-C), where the number of detected infected boar decreased despite testing more (as 100% of hunted boar within the fence and buffer zones were tested in the increased hunting pressure scenario, compared to only 20% in the normal hunting pressure scenario). For pig herds, increased benefit in the longer term can be visually observed in the cumulative curves in Figure 4-2G-I, where the divergence between the scenario curves is seen to increase as the epidemic progresses. In the face of emerging threats such as ASF, where there is typically a haste to suppress disease spread, mechanisms which do not prove highly effective in the short term might be quickly abandoned or less favoured. These results suggest that the timescale over which different interventions are evaluated may influence the evaluation outcomes: interventions which have only limited effectiveness in the short term may yield significant long-term benefits. Thus, instead of comparing scenarios over a fixed period of time (30 days during the ASF Challenge), it could

be recommended to compare scenarios based on the time needed to reach a pre-defined level of efficacy.

4.4.1.2 Key point 2: Preventive culling around positive wild boar was most effective in pig herds

As done for wild boar, we compared various ASF management measures in pig herds, that could complement the baseline interventions defined by the European Union and that were implemented in Merry Island. These additional measures included increasing the size of the surveillance zone from 10 km (the standard radius used) to 15 km, or the preventive culling of herds either in a protection zone, defined as being at-risk (based on previous trade with infected herds), or located at less than 3 km from positive wild boar. These measures were evaluated and compared during phase 2 of the challenge.

Increasing the size of the surveillance zone by 5 km was not effective in reducing the number of infected pig herds (Figure 4-3). We also found that preventive culling of herds connected to detected infected herds had a negligible impact on the number of infected pig herds (Figure 4-3). Similar to our results, increasing the size of the surveillance zone or pre-emptive culling around infected herds were not predicted to improve the management of a hypothetical ASFV epidemic in Denmark (Halasa et al., 2018, 2016c). In our case, these results can be explained by the fact that incidence in pig herds was largely driven by transmission from wild boar (Figure B4-11, Appendix B). Therefore, these scenarios strictly relating to pig-to-pig transmission had only limited impact. All these scenarios were insensitive to delays in implementing culling and to imperfect detection despite increased awareness.

In contrast, culling pig herds located less than 3 km away from positive wild boar decreased the number of infected pig herds by 18.5% in one month (Figure 4-3). This type of preventive culling was more effective as it prevented boar-to-pig transmission, by depleting pig herds before they were exposed to transmission from wild boar. However, this scenario required the culling of 67 additional herds in a month compared to the baseline scenario. Although the costs of disease management interventions were not directly evaluated in our model, the costs

associated with this scenario would probably be substantial. The cost-benefit ratio of this strategy should therefore be evaluated by comparing the costs of culling additional herds with the benefits of preventing ASF in a few herds.

One key result of our sensitivity analysis was that the efficacy of culling pig herds located less than 3 km away from positive wild boar was the only scenario sensitive to the delay before the implementation of culling: although it showed a better efficacy compared to baseline when the delay was one day, the efficacy was not different from baseline for a delay of seven days (Figure 4-6). Therefore, recommending this strategy to policy makers should include a warning that it would only be effective if this delay can remain very low, which may not be possible in practice depending on the number of herds that need to be culled. As an illustration, during phase 3, the challenge coordinators informed the players that culling pig herds located less than 3 km away from infected boar was implemented from day 90, but with a delay of 5-7 days, to reflect what could be achieved in real-life situations.

One possible refinement of this scenario would be to preventively cull the herds most at-risk of transmission from wild boar, i.e., outdoor herds. Here, all herds (indoor and outdoor) were indiscriminately culled, whereas only outdoor herds were assumed to be exposed to transmission from wild boar. By culling only outdoor herds close to positive wild boar, this scenario would be expected to produce similar benefits while decreasing the number of preventively culled herds, hence reducing the costs and potentially reducing delays required to implement such culling. Double fencing of outdoor pig herds as an alternative to keeping FRPs has been implemented as part of the recent eradication programme in Sardinia (Viltrop et al., 2021). Whilst the epidemiology of ASFV in Sardinia is different to that of Merry Island as FRPs rather than wild boar are considered the main drivers of transmission, double fencing of outdoor pig farms could also be considered as an alternative or addition to culling as a means of reducing transmission between wild boar and outdoor pig herds. Scenarios such as these could have been evaluated using our model; however, given the time restrictions

imposed in the challenge to mimic real-time analysis and decision making, we restricted our analyses to the scenarios asked by the challenge coordinators.

4.4.1.3 Key point 3: ASFV persistence beyond day 230 and what this means for disease control

Concerning the probability of epidemic fade-out, our model estimates suggest the strong likelihood of the persistence of ASFV in the landscape by day 230 (Figure 4-4), translating to a probability of incidence among pig herds, since the epidemic in pig herds is sustained by that in wild boar (Figure B4-11, Appendix B). The fact that our simulations stopped at the beginning of the peak birthing season (February – April (Alves da Silva et al., 2004; Podgórski and Śmietanka, 2018; Rosell et al., 2012; Sabrina et al., 2009)), also suggests a potential for endemicity of the virus in the landscape with seasonal epidemics, since the peak of introduction of new susceptible individuals into the population represents increased opportunities for transmission (Altizer et al., 2006). In addition, boar piglets have been seen to survive for longer periods after infection compared to adult boars (Sánchez-Cordón et al., 2019), meaning an increased potential for effective contacts per infected individual and hence a higher chance of epidemic take-off.

However, the simulated epidemic which was used to provide the data used in the challenge showed a decline in real incidence in wild boar from around day 35, down to almost no new cases by day 230 (ASF Challenge coordinators). This discrepancy between our results and the original model could have originated from the differences in assumptions. For instance, ASFV was introduced into a single wild boar in the original model 112 days before the first detected case (ASF Challenge coordinators). In our model, ASFV infections were seeded in wild boar at day 1 for phases 1 and 2 or at day 60 for phase 3, based on the number of observed infected boar (Appendix A). As the number of seeded infections was assumed based on the number of observed infected boar and not estimated, this number could have been underestimated. This could have resulted in a temporal shift of the epidemic according to our model, whereby day 230 would be in earlier stages of the epidemic than observed in

the “real” trajectory, thus overestimating virus persistence. Estimating the number of infections at the beginning of the simulations or the date of introduction of the virus could represent possible refinements to avoid this issue. Another possible explanation could be the spatial spread of the virus. Spatial diffusion is dependent on the probability of the virus reaching new areas with susceptible individuals. Differences in parameter estimates or in spatial structure (patches in our model versus individual boar in the original model) could explain a faster diffusion in our model compared to the original one, increasing the chances of reaching new areas with susceptible boar and therefore increasing virus persistence.

However, virus persistence, as predicted by our model, is more reflective of the current situation and challenges being faced by many countries within Europe. Since its re-introduction to Europe in 2007, only two countries – the Czech Republic and Belgium – have managed to eradicate ASFV when it has been present in wild boar (European Food Safety Authority et al., 2021, 2020; Sauter-Louis et al., 2021a). In both countries, ASFV was restricted solely to wild boar following a focal introduction. The control measures utilised reflect measures currently recommended by the EU when a focal introduction within wild boar has occurred in a previously disease-free area (European Commission, 2020b; European Food Safety Authority et al., 2020). Three separate zones are demarcated – a core zone, a buffer zone and an intensive hunting zone (European Food Safety Authority et al., 2020, 2018). The core zone is the area within which ASFV-positive boar have been identified. This area is fenced to control the movement of the wild boar with the goal of reducing disturbance and avoiding dispersal of infected animals over a wider area. Mortality associated with ASFV is allowed to occur and carcasses promptly removed. During the period of active ASFV transmission it is recommended that boar are undisturbed within both the core and buffer zones. Once the epidemic starts to decline, active population management is recommended under strict biosecurity. Within the intensive hunting zone (which is the outermost of the three zones), the goal is to reduce the population of wild boar to below a level at which transmission of ASFV cannot be sustained. Following the success of these strategies in Europe, a similar

approach has been adopted in South Korea but has met with variable success, thought in part to be due to failures in biosecurity measures and training (Jo and Gortázar, 2021). The challenge presented by ASF management highlights the importance of developing accurate mathematical models of ASFV transmission in wild boar and domestic pigs, to improve our understanding of ASFV transmission dynamics and to evaluate potential disease management strategies in various situations and locations.

4.4.2 What were the main drivers of model dynamics?

The sensitivity analysis allowed to gain insights on the key parameters influencing infection and detection dynamics.

We found that the detected incidence was largely unaffected by changing MIR values when MIR ≥ 8 km (Figure 4-5A). To understand this, it is helpful to consider the value of the dispersal kernel (where $\alpha = 1$ km as in the baseline model): both the value of the dispersal kernel and, consequently, the infectious pressure exerted on a susceptible patch j by an infectious patch i, decreases with increasing distance between patches (equation (4.1)). Thus, although the MIR increases, resulting in an increase in the number of possible patches j that could be infected by i, the infectious pressure exerted by i on patches located at least 8 km away is negligible (value of dispersal kernel for $d_{ij} =$ 8 km is 3.3×10^{-4}), hence such long-distance infection events are unlikely in the model. Consequently, increasing the MIR beyond 8 km does not contribute substantially to the number of new infections, as observed in Figure 4-5A. Hence, fixing the MIR at 8 km did not artificially restrict the dynamics.

Larger values of α led to larger estimates for the detected incidence (Figure 4-5B and Table 4-5). Indeed, given constant β, d_{ij}, and the prevalence in patches i and j, larger values of α will result in higher infectious pressures on a susceptible patch j (equation (4.1)) and hence more infections (Figure C4-12B, Appendix C), and consequently, detections, than would be realized with a smaller value of α.

When the duration of infectiousness in carcasses was 50 days or more, there was almost no sensitivity of both infection and detection dynamics to changes in the values of this parameter (Figure 4-5C; Figure C4-12C, Appendix C). This is due to the fact that once boar became carcasses, they persisted in the landscape no more than 43 days on average (by day 110) consistent with observed removal dynamics and model assumptions. Thus, values larger than 43 days will be expected to produce similar dynamics. However, values smaller than 43 days will be expected to produce different dynamics; in particular, the number of daily infections and consequently, detections will be generally lower, as infectious carcasses spend less time in the landscape.

Finally, we found that when the length of the infectious period in live boar was assumed to be shorter than 14 days (as in the baseline model), the detected incidence was generally lower than realized with the baseline model (Figure 4-5D). Indeed, the more days infectious boar spend alive, the higher the proportion of infected boar among all hunted boar, given that within the period considered (day 60-day 110), the major mode of detection of infected boar was hunting. The graph of the corresponding dynamics for all infections (i.e., including undetected infected boar; Figure C4-12D, Appendix C) reveals that compared to the detection dynamics, infection dynamics were less sensitive to changing values of this parameter. The number of days an infected boar spends alive is expected to influence detection dynamics more than it does the infection dynamics because: (1) the bulk of detections targeted live boar; hence, more live infected boar means a higher probability of detecting an infected boar; and (2) transitioning to become a carcass – thereby shortening the duration of infectiousness as a live boar – does not immediately affect a boar's infectiousness (in the baseline model, once a carcass, infectiousness lasts 90 days).

Further sensitivity analysis will consider the impact on projections when levels of intervention effectiveness are varied jointly.

4.4.3 What challenges did we face?

The main challenge faced in model implementation concerned computation time. The complexity of the models, coupled with the increasing amounts of data as the modelling challenge progressed, made simulations and parameter estimation slow. This efficiency drawback was even more evident during the early stages of phases 2 and 3 for two reasons: (1) in these phases, we included information on individual boars and locations because this level of granularity was needed for the implementation of pig management strategies, such as the culling of pig herds less than 3 km away from an infected wild boar, and (2) candidate models had to be iteratively tested prior to parameter estimation. To mitigate this issue, we employed three approaches. First, algorithms were parallelized where possible and useful, taking advantage of high-performance computing clusters. Second, cross-language programming was utilised where needed. Although the main programming language was R, some sections of the model were written in the faster C++ language to improve the overall speed. Third, the final model (at phase 3) was fitted to observed data from day 60, rather than from day 1. (This choice likely contributed to the accuracy of the phase 3 projections: in a previous epidemic modelling challenge, it was observed that models fitted to more recent data performed better than those fitted to data over the entire observed period (Viboud et al., 2018)). Still, the computational resources required were substantial (see Table A4-11, Appendix A for algorithm runtimes). It is crucial, particularly in real-time analysis of epidemics, for modellers to have efficient tools in order to provide timely evidence-based recommendations for disease management. Therefore, more work is required on the efficient design of epidemic models to minimize computational burden upon implementation. Also, work to develop highly efficient parameter estimation methods which have the potential to scale with large datasets and complex models will be useful for real-time epidemic response.

4.4.4 How could the modelling approach/choices be improved?

The modelling approach presented here can be improved in a number of ways. First, the component of the model describing transmission dynamics in wild boar could be made more

realistic by including a latent compartment, as in the pig herd model. The sensitivity analysis on the duration of infectiousness in live boar may be considered an approximate test on the length of a latent period on infection dynamics: one may think of the baseline model as allowing for no latent period and of the alternative models as allowing for a latent period of D days, where D is the decrease in infectious period between the baseline and the alternative model. The difference in the number of infected detected boar between the baseline model and the alternative models (decrease in infectious period by 4, 7 and 9 days) was notable – the lower the infectious period, the lower the number of infected detected boar relative to the baseline (Figure 4-5, Table 4-5). The absence of a latent period in the wild boar model may therefore explain the overestimation in the number of infected detected cases in phase 2 (Table 4-4 and Table 4-5).

Second, our model could be fitted to pig herd incidence, to better characterize infection dynamics between herds. We were not able to fit our model to pig herd data because the data were restricted to the number of detected infected pig herds, which was very low especially during the early phases of the challenge. Incorporating the analysis of data on pig herds could have allowed the separate estimation of α and β parameters for boar-to-boar and boar-to-pig transmission. However, the close similarity between model projections and observed data for pig herds (Table 4-3; Figure B4-10, Appendix B; Table 4-4) show that the use of common parameters for boar-to-boar and boar-to-pig transmission was sufficient for the purposes of our model. In addition, this avoided the need to perform ABC for both components of the model, which would have increased an already long computation time. Parameters for within- and between-herd transmission were based on experimental infections (Gallardo et al., 2017; Guinat et al., 2016b), previous modelling work (Halasa et al., 2016c, 2016a) or adapted from knowledge from classical swine fever virus. More detailed data, for instance on the number of infected or dead pigs in each herd, could have been useful to estimate within-herd parameters (Guinat et al., 2018). This kind of data could be collected when facing a real ASF epidemic to better inform mathematical models used.

4.4.5 How can the projections be improved?

Our projections could be improved by utilising multi-model ensembles as these have consistently demonstrated superior prediction abilities and lower variance, on average, compared to single models for epidemic forecasts (Johansson et al., 2019; McGowan et al., 2019; Reich et al., 2019; Viboud et al., 2018), deriving advantage from their ability to incorporate various signals from their constituent models, each of which may capture a distinct combination of system characteristics (McGowan et al., 2019). In the context of modelling challenges or real-time analysis of epidemics, the limited time available for analysis may make it challenging to develop multiple, diverse models needed for a good ensemble. For some modelling challenges, an ensemble based on the presented models have been developed (for example, (McGowan et al., 2019; Viboud et al., 2018)) and such ensembles could serve as useful tools for informing disease management in the event of a real epidemic.

4.4.6 Comparison to previous modelling studies

The pig herd model used within this study was broadly based on the models reported by Halasa et al. (2016c, 2016a). For within-herd transmission, the main modification that was implemented within our model related to the duration of the latent and infectious periods. We based these on experimental data reported by Guinat et al. (2016b) rather than on expert knowledge and we incorporated uncertainty in these parameter values as in Vergne et al. (2021).

For between-herd transmission, a number of modifications were implemented. We only considered disease spread via animal movements and via local transmission, as these were the main drivers of between-herd transmission in Halasa et al. (2016c) and Andraud et al. (2019). For spread via animal movements, instead of computing probabilities of virus transmission via movements (Halasa et al., 2016c), we explicitly modelled animal movements as a potential source of introduction (e.g., as in Brooks-Pollock et al. (2014)), using observed movement data and projected movements using ERGMs. For local transmission, we used a

continuous function of distance to represent the decreasing probability of transmission with increasing distance, instead of using discrete values for certain distance ranges as in Halasa et al. (2016c).

The wild boar model was developed independently and was not based on any previously published modelling studies. As noted in Hayes et al. (2021), until 2020 the majority of the published ASFV transmission models for wild boar were based on Lange and Thulke's ASF model (Halasa et al., 2019; Lange, 2015; Lange et al., 2018; Lange and Thulke, 2017, 2015; Thulke and Lange, 2017) or parameterized as per that model (Croft et al., 2020). Our model is similar to that by Lange and Thulke (2017) in that it is a spatially explicit model. However, whilst many of the published wild boar ASF modelling studies include detailed demographic information (age and sex of individual boar, births, sub-adult dispersal, annual reproduction, litter sizes and mortality) (Croft et al., 2020; Gervasi and Guberti, 2021; Halasa et al., 2019; Lange, 2015; Lange et al., 2018; Lange and Thulke, 2017, 2015; O'Neill et al., 2020; Pepin et al., 2020; Thulke and Lange, 2017) we chose to simplify the demographic processes included within our model due to the short time-frame modelled within the ASF Challenge.

The duration of infectiousness of carcasses is variable across studies and varies from 4 weeks (Lange and Thulke, 2017) to 12 weeks (Gervasi and Guberti, 2021). The 90-day period used in our model would thus be at the upper end of this range. More recently, studies have varied the rate of carcass decomposition by season to reflect different seasonal rates of carcass decomposition (Gervasi and Guberti, 2021; Pepin et al., 2020; Thulke and Lange, 2017). Another study has demonstrated the influence of temperature and environmental conditions on ASFV persistence in carcasses (Fischer et al., 2020; Mazur-Panasiuk and Woźniakowski, 2020). The projection periods for the ASF modelling Challenge ran from August to February and thus these seasonal and temperature variations in the duration of infectiousness of carcasses could have been considered in our model.

Prior to 2020 (when the ASF Challenge started), there had been a lack of diversity among ASFV models in both domestic pigs and wild boar although the situation has been improving

(Hayes et al., 2021). Our model, alongside the other models produced in the ASF Challenge, provides a valuable contribution to increasing the diversity in the ASFV modelling literature. The number of studies modelling transmission between wild and domestic hosts remains small (Pietschmann et al., 2015; Pollock et al., 2021; Taylor et al., 2021; Yoo et al., 2021). Given the importance of wild boar in the transmission of ASFV in Europe, the multi-host nature of our model is one of the major strengths of our study.

4.5 CONCLUSIONS

In summary, we have developed a framework for modelling ASFV transmission during outbreaks. The model can be parameterized in real-time during outbreaks and refined as additional outbreak data become available. The model can be used to generate forward projections and to predict the effectiveness of different proposed disease management strategies.

For the simulated epidemic on Merry Island, our model indicated that transmission between wild boar (and from wild boar to pig herds) was the main driver of epidemic dynamics. Effective control measures included the construction of a fence around the main area of the island with high prevalence, following by increased hunting of wild boar both within and near the fenced region. Culling of pig herds was generally not an effective control strategy, except in regions with substantial numbers of infections in wild boar. This is because there was only a low risk of transmission through the pig trade network. Our model predicted that the virus is likely to persist in future on Merry Island, at least in the short to medium term.

An important general finding is that it is important to consider the timescale over which different control strategies are evaluated. Interventions that have only limited effectiveness in the short term may yield significant long-term benefits.

Further refinement of our modelling framework is necessary going forwards. Nonetheless, we have demonstrated the potential for this approach to be used to generate projections and

assess different possible control measures during future ASFV outbreaks. This will help animal health policy makers optimise disease management decisions during future outbreaks.

Acknowledgments

The authors would like to acknowledge the organizers of the ASF modelling challenge for developing this interesting modelling problem and for the opportunity to participate in the challenge.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. E.A.D was funded by a Rhodes Scholarship and a studentship at the Department of Statistics, University of Oxford. S.L. was funded by the EDCTP program supported by the European Union (FibroScHot ref: RIA2017NIM-1842). S.H. was funded by the Engineering and Physical Sciences Research Council. C.A.D. was supported by joint Centre funding from the UK Medical Research Council and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union. C.A.D. was funded on grants from the UK National Institute for Health Research (NIHR) [Vaccine Efficacy Evaluation for Priority Emerging Diseases: PR-OD-1017-20007 and HPRU in Emerging and Zoonotic Infections: NIHR200907]. The views expressed in this publication are those of the authors and not necessarily those of their funding institutions.

Declarations of interest: none

CRediT author statement

Emmanuelle A. Dankwa: Conceptualization, Methodology, Formal analysis, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization. **Sébastien Lambert**: Conceptualization, Methodology, Formal analysis, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization. **Sarah Hayes**: Conceptualization, Methodology, Data Curation, Writing – Original Draft, Writing – Review & Editing. **Robin Thompson**: Conceptualization, Methodology, Writing – Original Draft, Writing – Review & Editing, Supervision. **Christl A. Donnelly**: Conceptualization, Methodology, Writing – Original Draft, Writing – Review & Editing, Supervision.

TABLES

Table 4-1: Epidemiological parameters. For estimated parameters, mean values along with 95% CrIs, in parentheses, are reported.

	Description	Mean value(s) (95% CrI)	Source(s)
Wild boar			
α	Scale parameter of dispersal kernel	Phase 1: 0.8225 km (0.8006-0.8800) Phase 2: 0.87 km Phase 3: 1 km	Estimated at phase 1 Fixed at phases 2 & 3
β	Overall infection rate	Phase 1: 0.0018 day^{-1} (0.0011-0.0028) Phase 2: 0.0077 day^{-1} (0.0069-0.0084) Phase 3: 0.0035 day^{-1} (0.0034-0.0036)	Estimated
	Time from infection to death	14 days	(Blome et al., 2012; Pietschmann et al., 2015)
	Infectious period for carcasses	90 days	(Fischer et al., 2020)
	ASFV-related mortality rate	100%	(Blome et al., 2013, 2012)
Additional parameters for phase 1 model			
r_1	Fraction of positive boar in a patch when day <28, averaged over all patches	0.16 (0.09-0.20)	Estimated*
r_2	Fraction of positive boar in a patch when day ≥38, averaged over all patches	0.25 (0.20-0.38)	Estimated*
d	Detection rate for positive boar	0.10	Assumed
Domestic pig herds			
β^{PH}	Transmission rate	0.60 day^{-1}	(Guinat et al., 2016b)
τ	Mean lifetime of ASFV in residues from dead pigs	$1/\log(2)$ days	Adapted from: (Halasa et al., 2016a)
δ	Average duration of the pre-infectious period	$\mathrm{PERT}(3;4;5)$ days†	(Guinat et al., 2016b; Vergne et al., 2021)
ϕ	Average duration of the subclinical period	2 days	ASF Challenge coordinators
γ	Average duration of the infectious period	$\mathrm{PERT}(3;7;14)$ days†	(Guinat et al., 2016b; Vergne et al., 2021)
μ	Probability of pigs dying following infection	0.95	(Gallardo et al., 2017; Halasa et al., 2016a)
ρ	Transmission rate by local spread	0.005 km.day^{-1}	Adapted from: (Halasa et al., 2016c)

* This parameter was defined to be time-varying to reflect the spread of the infection (in the absence of disease management measures) as time progressed. For day ∈ [28, 37], the fraction of positive boar in a patch was given by: $r_1 + \frac{r_2 - r_1}{38-28} \times (\mathrm{day} - 28)$

† PERT distribution of parameters (minimum; mode; maximum)

Table 4-2: Differences between models across the three phases of the challenge. As the challenge progressed, the models had to be slightly adapted to account for new data and information provided by the challenge coordinators and/or to answer new questions.

	Phase 1	**Phase 2**	**Phase 3**
Explicit modelling of individual infected boar in a patch (and their locations)	No	Yes	Yes
Detection of infected wild boar	Yes (fixed rate)	Yes (through testing of hunted boar and active surveillance of boar carcasses)	Yes (through testing of hunted boar and active surveillance of boar carcasses)
Increased hunting pressure in fence and buffer zone	Yes (no buffer zone)	Yes	Yes
Permeability of the fence	No	No	Yes
MIR	No	No	Yes (8 km)
Test of preventively culled pig herds	Not applicable	No	Yes
Delay before preventive culling	Not applicable	Yes (24 hours)	Yes (5-7 days)

Table 4-3: Model fit and projections for the cumulative number of detected infections under the two main disease management scenarios considered in wild boar: increased hunting pressure and normal hunting pressure. The model fits are median model estimates for the observed period (days 1-50 for phase 1, days 1-80 for phase 2 and days 1-110 for phase 3) while the model projections are median model estimates for the unobserved periods over which projections were computed (days 51-78 for phase 1, days 81-110 for phase 2, days 111-230 for phase 3). Model estimates are medians of 500 simulations along with 95% CrIs in parentheses.

			Wild boar			Pig herd		
Phase	**By day**	**Disease management scenario***	**Observed**	**Model fit (95% CrI)**	**Model projections (95% CrI)**	**Observed**	**Model fit (95% CrI)**	**Model projections (95% CrI)**
	50		397	396 (358-435)		3	4 (2-6)	
1	78	Increased hunting pressure			1770 (1445-2503)			8 (5-14)
		Normal hunting pressure			933 (751-1289)			8 (5-14)
	80		2007	2009 (1912-2102)		12	12 (8-17)	
2	110	Increased hunting pressure			3214 (3112-3378)			28 (22-35)
		Normal hunting pressure			3272 (2973-3868)			30 (23-38)
	110		2984	2994 (2897-3077)		26	25 (21-31)	
3	140	Increased hunting pressure			3442 (3372-3514)			38 (32-48)

		Normal hunting pressure		7954 (6891-8827)	87 (67-100)
	230	Increased hunting pressure		4599 (4480-4711)	113 (99-129)

* For all phases, scenarios are only indicated for projected periods and not for observed periods. For the observed periods, the scenario for phase 1 is normal hunting pressure with no fence whereas the scenario for phases 2 and 3 is increased hunting pressure.

Table 4-4: Comparison of model projections and observed data on the cumulative number of detected infected wild boar and pig herds up to days 78 and 110. For both wild boar and pig herds, model projections shown here assumed disease management measures as implemented during the indicated periods. Model estimates are medians of 500 simulations along with 95% CrIs in parentheses.

Population	**Category**	**Cumulative number of detected infections up to:**	
		Day 78	**Day 110**
Wild boar	Model	1770 (1445-2503)	3214 (3112-3378)
	Observed	1903	2984
Pig herds	Model	8 (5-14)	31 (23-39)[1]
	Observed	10	26

[1] To adequately compare the results of the model with the observed data, the projections for pig herds up to day 110 (phase 2 model) were simulated using disease management measures as implemented during the indicated period, i.e., incorporating preventive culling of all pig herds located at less than 3 km from positive wild boar carcasses from day 90, with a delay of 5-7 days between the confirmation of a wild boar case and pig herd culling, and performing tests in all culled herds which provided results the day after.

Table 4-5: Number of detected infections in wild boar and pig herds by day 110 under alternative parameter values tested in the sensitivity analysis. Estimates presented are medians and 95% CrIs, in parentheses, of 100 stochastic repetitions of the model. Parameter values as used in the baseline model (MIR = 8 km, α = 1 km, duration of infectiousness in carcasses = 90 days; and duration of infectiousness in live boar = 14 days) and corresponding outcomes are in bold.

Parameter	**Values**	**Median number of detected infections by day 110 (95% CrI)**	
		Wild boar	**Pig herds**
MIR (km)	2	1435 (1395-1476)	12 (10-16)
	8	**2991 (2856-3125)**	**24 (20-29)**
	14	3043 (2891-3199)	24 (21-31)
	20	3061 (2924-3189)	24 (20-30)
Scale parameter of dispersal kernel, α (km)	0.6	1804 (1736-1861)	14 (11-19)
	0.8	2282 (2202-2361)	17 (14-21)
	1.0	**2991 (2856-3125)**	**24 (20-29)**
	1.2	4094 (3945-4279)	33 (28-40)
Duration of infectiousness in carcass (in days)	10	2756 (2639-2890)	23 (19-28)
	50	2995 (2868-3106)	24 (20-30)
	90	**2991 (2856-3125)**	**24 (20-29)**
	130	2982 (2835-3103)	24 (20-30)
Duration of infectiousness in live boar (in days)	5	2071 (1986-2138)	24 (20-29)
	7	2308 (2224-2434)	24 (20-29)
	10	2638 (2501-2774)	24 (20-28)
	14	**2991 (2856-3125)**	**24 (20-29)**

FIGURES

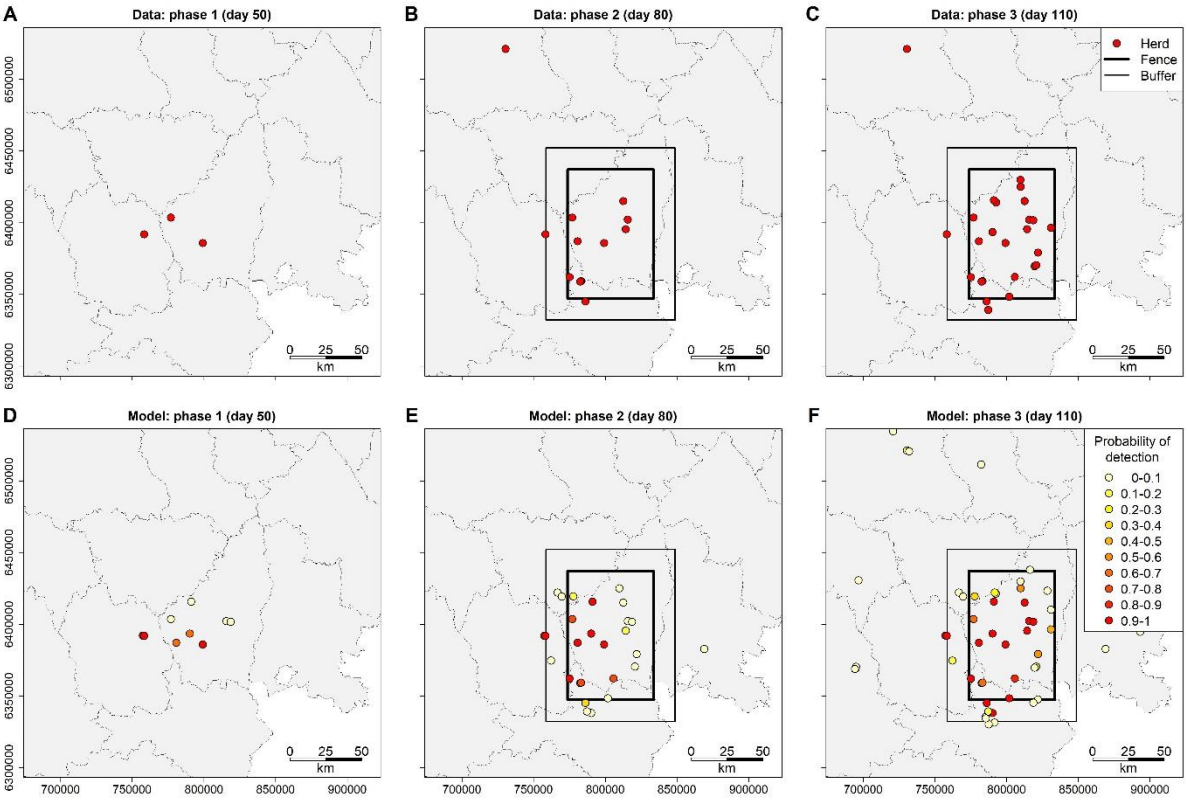


Figure 4-1: Distribution of detected infected pig herds on Merry Island: comparison between data (top panels showing only detected infected pig herds) and model simulations (bottom panels showing all pig herds that were detected as positive in at least one simulation) for each phase. **Top panels (A-C)**: detected herds by (A) day 50, (B) day 80 and (C) day 110 (for phase 1, 2 and 3 respectively) in the data provided by the challenge coordinators. Detected herds are indicated by red dots, while the fence and buffer zones (implemented during phases 2 and 3) are indicated by thick and thin rectangles, respectively. **Bottom panels (D-F)**: detected herds by (D) day 50, (E) day 80 and (F) day 110 in the model simulations run with estimated parameter values. Dots indicate herds that were detected in at least one simulation, with colours indicating the proportion of simulations in which a given herd was detected (among 500 simulations).

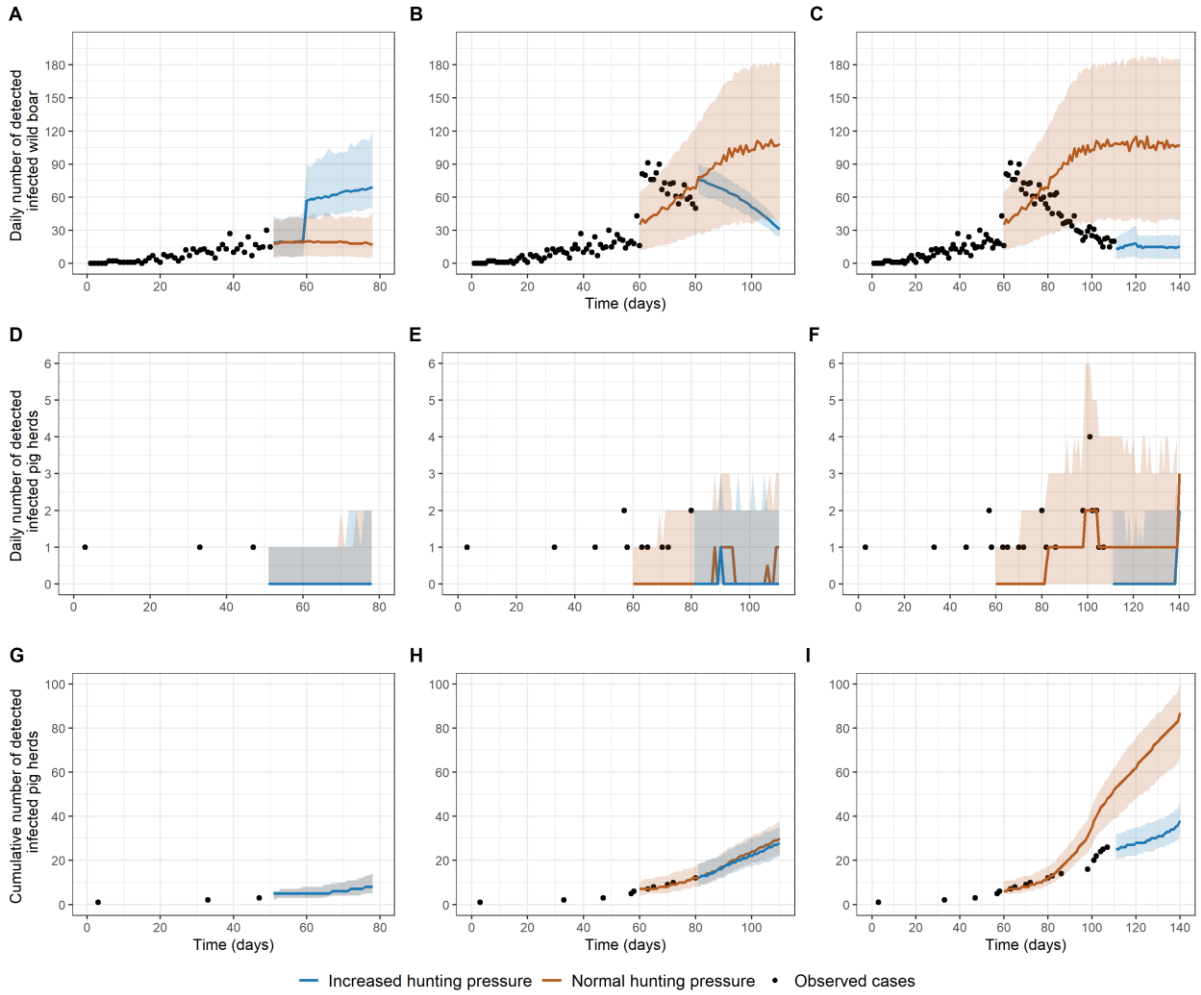


Figure 4-2: Comparison of disease management scenarios for wild boar (increased hunting pressure and normal hunting pressure) across all phases. **Top panels (A-C)**: Observed (black dots) and projected daily number of detected infected wild boar under the increased hunting pressure (blue) and normal hunting pressure (light orange) scenarios. In panel C, the drop at day 120 is due to a cessation in increased hunting pressure activities. **Middle panels (D-F)**: Observed (black dots) and projected daily number of detected infected pig herds under the increased hunting pressure (blue) and normal hunting pressure (light orange) scenarios. **Bottom panels (G-I)**: Observed (black dots) and resulting cumulative numbers of detected infected pig herds from the daily projections (D-F), under the increased hunting pressure (blue) and normal hunting pressure (light orange) scenarios. **All panels**: Median model projections are shown along with 95% CrIs (shaded areas with corresponding colours). Projections were obtained using the model calibrated to data up to days 50, 80 and 110 for the increased hunting pressure scenario in Phase 1 (A, D, G), phase 2 (B, E, H) and phase 3 (C, F, I) respectively. For the normal hunting pressure scenario, the model was calibrated using data up to day 50 for phase 1 and up to day 59 for phases 2 and 3. **[Note: the observed data in all cases arose in the challenge scenario in which hunting pressure increased from day 60, making these data not directly comparable with the normal-hunting-pressure-throughout projections (light orange) in the middle and right-hand columns.]**

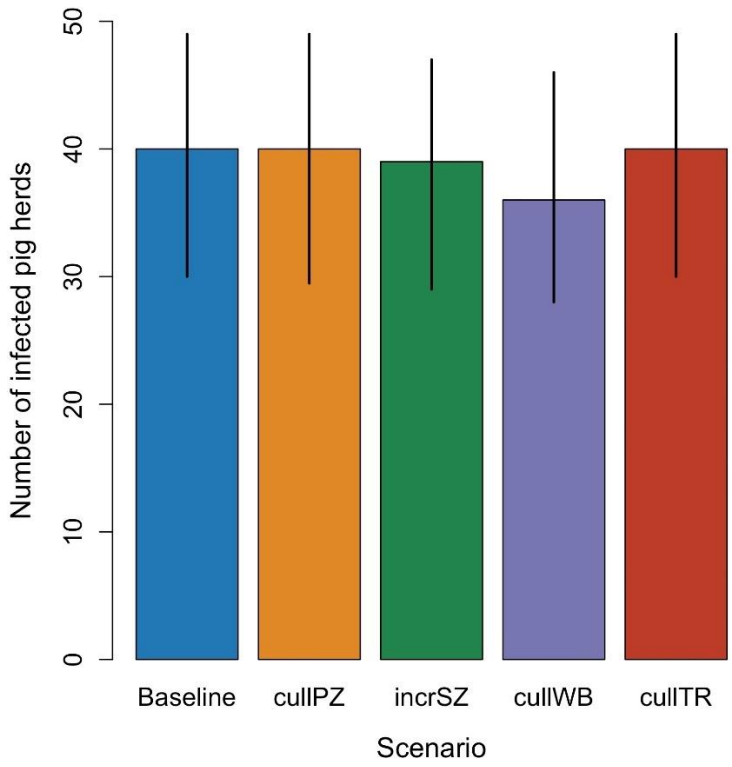


Figure 4-3: Comparison of additional disease management measures for pig herds (phase 2). Barplots represent the median number of infected herds at the end of the projections for phase 2 (day 110), with 95% CrIs indicated by black central bars, for a baseline scenario and four additional disease management measures implemented in pig herds. The baseline scenario ("Baseline") involved regulatory interventions in pig herds and the implementation of fencing and increased hunting pressure in wild boar. The four disease management measures implemented in addition to the baseline scenario are: (1) "cullPZ": culling of all pig herds in protection zones; (2) "incrSZ": increasing the size of the surveillance zone from 10 km (the standard surveillance radius used) to 15 km; (3) "cullWB": culling of all pig herds located at less than 3 km from positive wild boar; (4) "cullTR": culling of all herds that have traded pigs with an infected farm less than three weeks before detection.

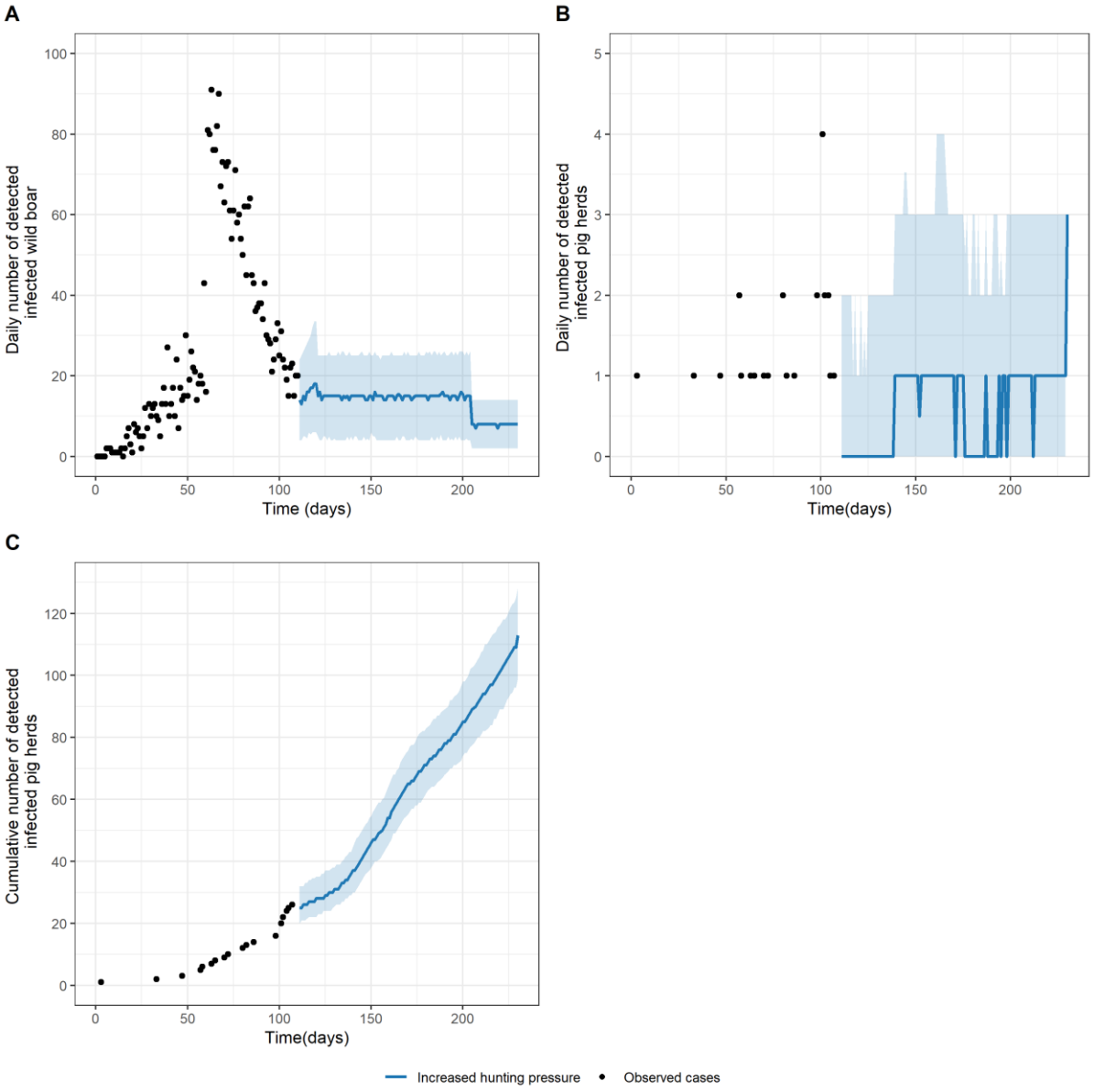


Figure 4-4: Model projections from day 111 to day 230 under the increased hunting pressure scenario. **A.** Observed and projected daily numbers of detected infected wild boar. Projections were obtained using the disease management measures as implemented over the indicated period: (1) increased hunting pressure from day 111 to day 120, (2) normal hunting pressure from day 121 to day 203 and (3) cessation in hunting activities (end of the hunting season) from day 204 onwards but permitting passive discovery of wild boar carcasses. The drop at day 204 is due to the cease in hunting activities. **B.** Observed and projected daily numbers of detected infected pig herds. **C.** Observed and projected cumulative numbers of detected infected pig herds. **All**: Black dots, blue line and shaded areas represent the observed data, median model projections and 95% CrIs, respectively. Projections were obtained using the model fitted to data up to day 110.

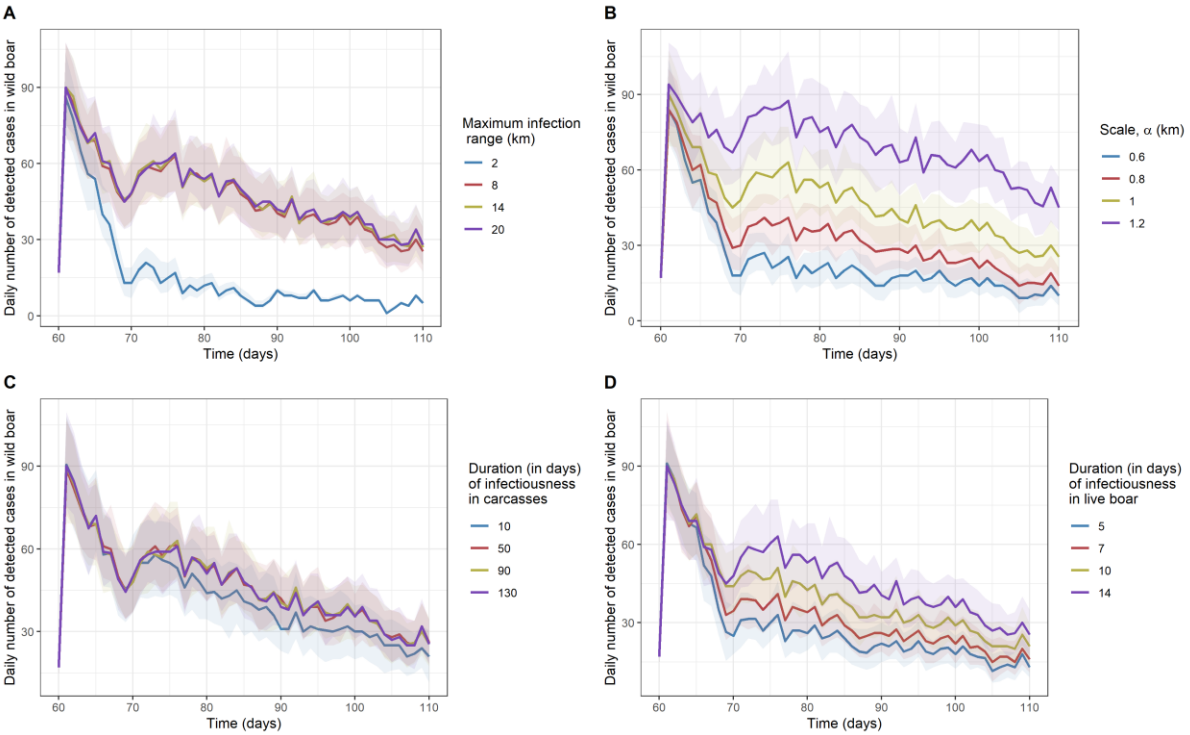


Figure 4-5: Sensitivity of the daily number of detected infected wild boar (from day 60 to day 110) to **A**: the MIR (values: 2 km, 8 km, 14 km, 20 km); **B**: the scale parameter α of the dispersal kernel (values: 0.6 km, 0.8 km, 1 km, 1.2 km); **C**: the duration of infectiousness in carcasses (values: 10 days, 50 days, 90 days, 130 days); and **D**: the duration of infectiousness in live boar (values: 5 days, 7 days, 10 days, 14 days). Trajectories are medians computed from 100 stochastic repetitions of the model. Shaded regions are corresponding 95% CrIs. Simulations corresponding to baseline parameter values (MIR = 8 km, α = 1 km, duration of infectiousness in carcasses = 90 days; and duration of infectiousness in live boar = 14 days) are the same across panels.

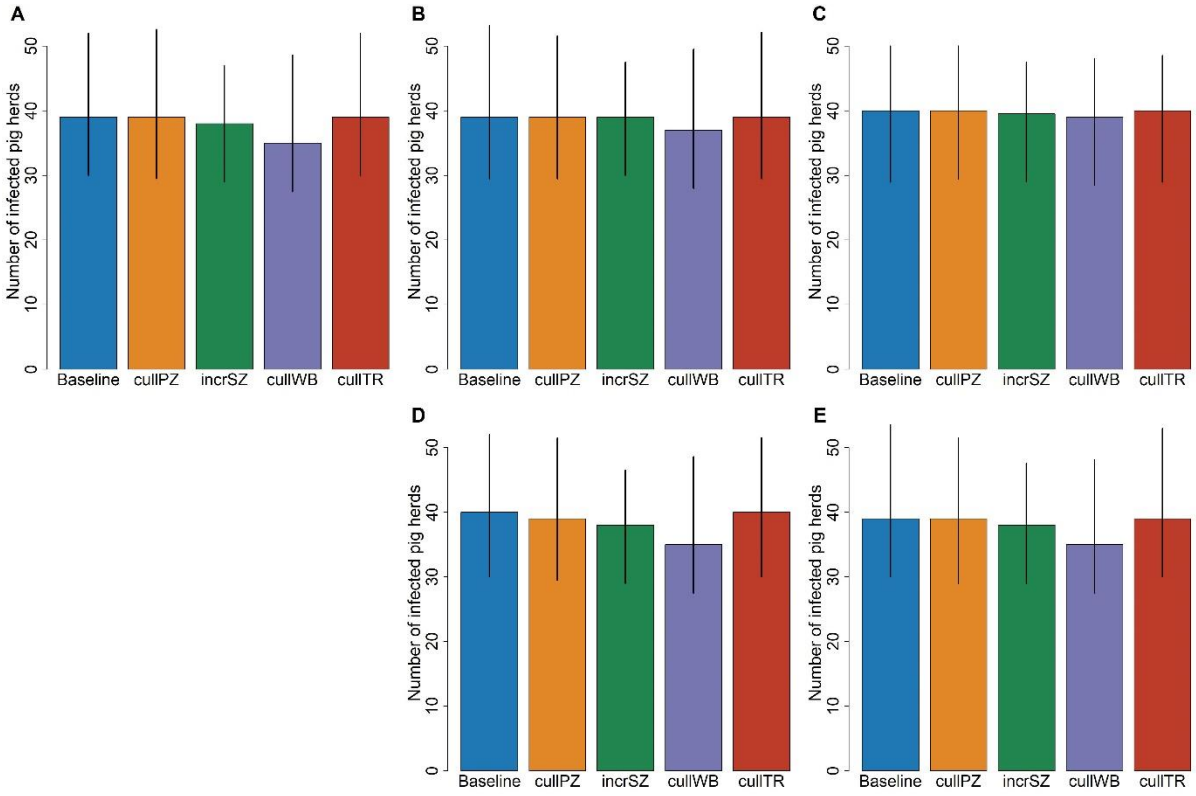


Figure 4-6: Sensitivity of the cumulative number of infected pig herds to the delay between confirmation and culling and to the minimum number of clinical or dead pigs required for detection. **A**: delay of one day, minimum of one clinical or dead pig required for detection (baseline); **B-C**: delay of four and seven days respectively, still with a minimum of one clinical or dead pig required for detection; **D-E**: minimum of three and five clinical or dead pig required for detection, respectively, keeping a delay between confirmation and culling of one day. Barplots represent the median number of infected herds at day 110, with 95% CrIs indicated by black central bars. The baseline scenario ("Baseline") involved regulatory interventions in pig herds and the implementation of fencing and increased hunting pressure in wild boar. The four disease management measures implemented in addition to the baseline scenario are: (1) "cullPZ": culling of all pig herds in protection zones; (2) "incrSZ": increasing the size of the surveillance zone from 10 km (the standard surveillance radius used) to 15 km; (3) "cullWB": culling of all pig herds located at less than 3 km from positive wild boar; (4) "cullTR": culling of all herds that have traded pigs with an infected farm less than three weeks before detection.

4.6 APPENDIX A: DETAILED MODEL DESCRIPTION

4.6.1 Model description

We modelled ASFV transmission on Merry Island using a model combining a stochastic SIP model with an exponential spatial transmission kernel describing viral transmission across the island via wild boar and a stochastic, metapopulation SLIRD model for pig herds. All pig herds within the island were considered, but, for computational efficiency, only wild boar within a defined area around the detected cases in wild boar and pig herds were considered (see Section 4.6.1.4 for details).

All simulations were performed with R version 4.1.2 (R Core Team, 2022), with a discrete time step of one day.

4.6.1.1 State variables

4.6.1.1.1 Wild boar

Epidemiological units / entities modelled: scale

For the wild boar model, patches were considered as model units; thus, transmission was modelled directly among patches rather than among individual boar. However, we still tracked the *health states* of boar within each patch.

Health states

Health states for patches and boar were susceptible *S*, infectious *I*, or post-infectious P. Latent periods were not considered (i.e., there was no exposed class for boar).

A patch was considered infected if at least one boar within the patch was infected. A patch was considered susceptible if there were live boar but no infectious boar within the patch. A patch was considered post-infectious if all boar in the patch were carcasses that were no longer infectious. Wild boar health states and corresponding patch states are summarized in Table A4-6.

Spatial distribution

In addition to reported boar locations, we simulated locations for boar which had either been removed and had not been tested (hence not reported) and boar that had not yet been removed (alive or dead in the landscape). For each administrative region, the total number of boar locations simulated was equal to the difference between the expected population size in that region in the previous year and the number of boar for which (location and infection) data had been reported.

For the sake of computational efficiency, we considered only the part of the island where infected wild boar had been found. There were 2500 rectangular patches considered in all: patches were chosen to reflect the social structure of wild boar. For the final model (phase 3), these patches each had a 7.5 km^2 area (2.5 km in the x-direction and 3 km in the y-direction)

with population sizes ranging from 0 to 110. Section 4.6.1.4 contains details on how patch populations were determined.

Table A4-6: Health states of wild boar with corresponding patch states. S, I, and P denote the susceptible, infectious, and post-infectious states, respectively.

Health state of wild boar in patch	Health state of patch
All *S*	*S*
All *P*	*P**
At least one *I*	*I*
Mixture of *S* and *P*	*S*

*Cannot be re-infected.

Transmission

Transmission between boar in any two patches was assumed to depend on the number of susceptible and infectious boar in both patches, the distance between their patch centres, and whether the patches were situated on the same side of the fence (By day 60, a fence had been erected as a disease management strategy to limit the spread of infection to other parts of the island). Section 4.2.2.1 in the main text contains details on transmission mechanisms.

4.6.1.1.2 Pig herds

Epidemiological units / entities modelled: scale

The pig herd model considers individual pigs by describing the numbers of pigs in each compartment within a herd, and the flows between the different compartments. Each herd was considered as a homogeneous, random-mixing population, not accounting for any within-herd structure.

Health states

Health states were susceptible S, exposed and pre-infectious E, infectious and subclinical (i.e., not showing clinical signs of ASF) I_{sc}, infectious and clinical (i.e., showing clinical signs of ASF) I_c, and immune (recovered) R.

Herd characteristics

Transmission from wild boar to pig herds was only possible for outdoor herds. Other herd characteristics were accounted for to predict pig movements based on the data provided by the challenge coordinators: the size of the herd, its production type, whether it belonged to the same producer as other farms (i.e., belonging to a group of “multisite farms” comprising several geographically distinct farms), whether it was commercial, and whether it was outdoor (see Section 4.6.1.2.2). The production type (farrow-to-wean, farrow-to-finish, finishing) was also accounted for in pig herd population dynamics (see Section 4.6.1.2.1).

Spatial distribution

Herds were distributed according to the coordinates provided by the challenge organizers: these coordinates were used to calculate Euclidean distances between herds, between herds and wild boar patches (for the computation of the force of infection exerted on susceptible herds), and between herds and individual infected boar (for evaluating the disease management measure consisting in culling herds within a 3 km radius of a positive boar).

4.6.1.2 Processes and parameter values

4.6.1.2.1 Population dynamics

4.6.1.2.1.1 Wild boar

We assumed a constant population size among wild boar in the absence of hunting and ASFV-related mortality. We did not account for birth or natural mortality in the timescale modelled. This was because the reproductive cycle of wild boar is reported to be seasonal with peak birthing occurring between February and April (Alves da Silva et al., 2004; Podgórski and Śmietanka, 2018; Rosell et al., 2012; Sabrina et al., 2009), hence the introduction of additional susceptible animals via births was considered unlikely to contribute significantly to transmission throughout the period over which predictions were required (27th August 2020 until 23rd February 2021, corresponding to days 51-230 of the epidemic).

The population of wild boar – dead and alive – remaining in the landscape (i.e., the portion of the island we considered) at any time could only be decreased through removal by hunting, active surveillance, or passive surveillance.

For model fitting, the number of boar hunted daily was estimated based on the data provided on the number of hunted boar tested daily. The proportion of hunted boar tested was 20% under normal hunting pressure and 100% under increased hunting pressure (see Main text for hunting pressure definitions). Thus, under normal hunting pressure, the number of boar hunted daily was equal to five times the number of tested boar, whereas under increased hunting pressure, the number of boar hunted daily was equal to the number of tested boar. The number of boar carcasses found daily (by passive or active surveillance) was solely determined based on the data provided, since all found boar carcasses were tested and hence reported. For model projections, the daily number of hunted boar and found carcasses was computed based on the observed fractions of removed boar and found carcasses. For both model fitting and model projections, after the number of boar to be removed had been determined, the particular boar to be removed was randomly sampled from the remaining boar within the area considered.

From day 60, a 300 km fence was erected around the initially infected area to limit disease spread (see Figure 4-1 in the main text). In addition, an area within 15 km of the fence was demarcated as a buffer zone. Within the fence and buffer zone, increased hunting pressure was implemented for 60 days (between day 60 and day 120). In phases 2 and 3, within the fence and buffer zone, the probability of removal by hunting was dependent on the infection status of the boar. We assigned a higher removal probability to infected boar than to

susceptible boar, such that infected boar were more likely to be hunted or found as carcasses. This was based on our assumption that infected boar were less likely to escape a hunt due to reduced activity as a result of ASF symptoms. The absolute difference in probabilities was 0.6 in phase 2 and 0.1 in phase 3: these were chosen to ensure a high agreement between the observed and simulated dynamics.

4.6.1.2.1.2 Pig herds

We assumed a constant population size in each herd in the absence of ASFV-related mortality, with two population dynamics processes depending on the production type of the herds: birth of susceptible pigs in farrow-to-wean and farrow-to-finish herds, and animals sent to the abattoir in finishing and farrow-to-finish herds.

Thus, in farrow-to-wean and farrow-to-finish herds, the number of pigs leaving the herd (outgoing movements) was compensated by the entry of the same number of pigs (susceptible only). On the other hand, in finishing and farrow-to-finish herds, the number of pigs entering the herd (ingoing movements) was compensated by the same number of pigs leaving to the abattoir. Such processes were considered an acceptable approximation of the population dynamics of the pig herds given the batch system used in swine production (reflected in the herd movement data provided) and the timescale of the simulations.

4.6.1.2.2 Movements

4.6.1.2.2.1 Wild boar

Transmission was modelled between static patch centres. However, patches contained boar which could move within the patch and to other patches, presenting opportunities for transmission. In phase 3, we restricted wild boar movements to a maximum distance from their patch centres to better reflect wild boar movement dynamics. This maximal distance, which we refer to as the MIR, was chosen to be 8 km (Table A4-7), to incorporate information on the maximum distance travelled and estimated home range of wild boar (*Sus scrofa*) (Janeau et al., 1995; Podgórski et al., 2013). Consequently, a patch *A* could infect another patch *B* only if the centre of patch *B* was situated at most MIR km from the centre of patch *A* (Figure A4-7).

Table A4-7: Parameters involved in wild boar movements

Symbol	Description	Value(s)	Source(s)
MIR	Maximum infection range based on maximum distance travelled and estimated home range of wild boar (*Sus scrofa*)	8 km	(Janeau et al., 1995; Podgórski et al., 2013) (Used only in phase 3)

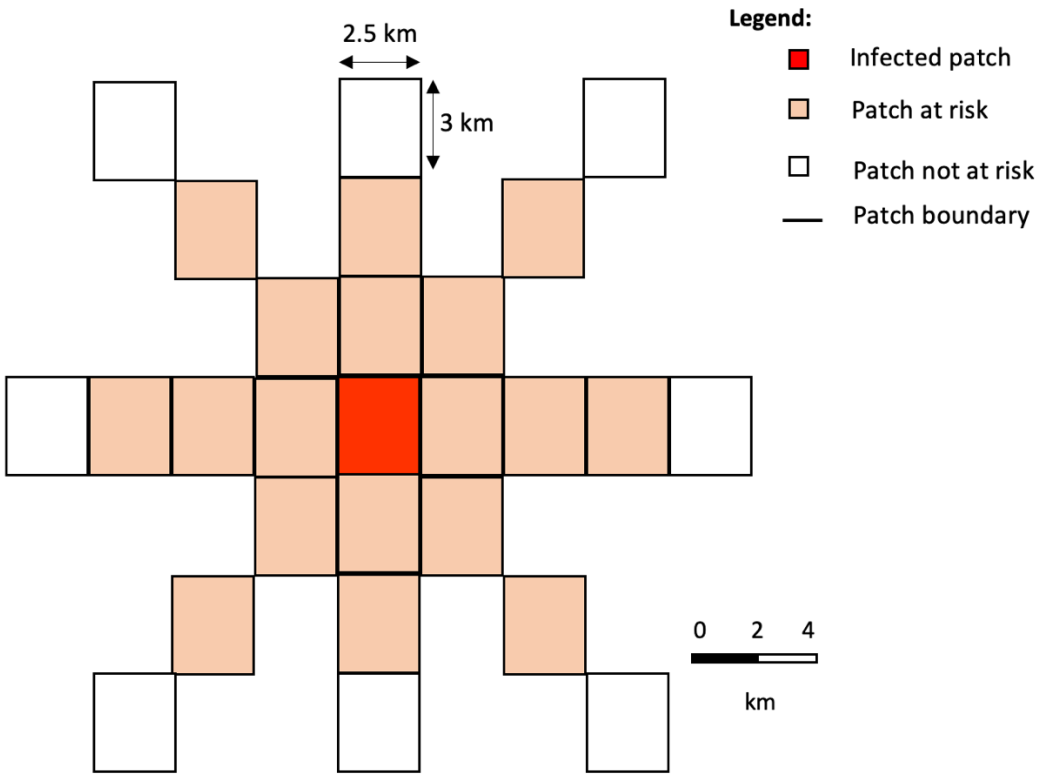


Figure A4-7: Depiction of the effect of the MIR on transmission. Transmission is possible between the infected patch (shaded red) and those patches with centres situated at most 8 km from the centre of the infectious patch (patches at risk of infection are shaded orange). Transmission is not possible between the infected patch and those patches situated more than 8 km from its centre (patches not at risk of infection are shaded white). Patch dimensions are 2.5 km by 3 km.

4.6.1.2.2.2 Pig herds

Data on movements of pigs up to day 110 were provided by the challenge coordinators and therefore pig shipments between herds were fixed (day of the shipment, source and destination herds, number of pigs shipped). Pig movements from day 111 onwards were predicted using ERGMs to determine a pair between a source herd and a destination herd (Relun et al., 2017), and using GLMs with zero-truncated negative binomial distribution to determine the number of pigs exchanged.

Pig movements can be represented as a network of nodes (the pig herds) connected by edges representing movements of pigs between them (Relun et al., 2017). ERGMs specify the probability of any random network Y given a set of n nodes as:

$$P_\theta(Y = y|n) = \left(\frac{1}{c}\right)\exp\left(\theta^T g(y)\right), \quad \text{(A4.6)}$$

where θ is the vector of model coefficients, $g(y)$ is the vector of model covariates and c is a normalizing factor corresponding to the sum of $\exp\left(\theta^T g(y)\right)$ over all possible networks with n

nodes (Hunter et al., 2008). Possible covariates include structure-related network statistics (for example to account for local network configuration: if herd 2 exchanges pigs with both herds 1 and 3, this might encourage herds 1 and 3 to exchange pigs as well), as well as node or edge attributes (e.g., herd characteristics or the distance between two herds). As in Relun et al. (2017), three types of ERGMs were built: (1) a simple Bernoulli model that only includes the number of edges; (2) a model with edges and structure-related network statistics ("edges + structure") and (3) a model with edges and nodal attributes ("edges + attributes"). Unfortunately, models combining both structural network statistics and nodal attributes ("edges + structure + attributes") were not fitted as they did not converge.

Six structural network statistics were used and related to local network topological measures such as number of isolates, clustering, or degree distribution (Table A4-8). Five types of network statistics based on categorical nodal attributes (herd production type: farrow-to-wean/farrow-to-finish/finishing, herd activity: commercial/backyard, herd housing system: outdoor/indoor, multisite farms) were used: number of ingoing and outgoing edges, uniform and differential homophily, and selective mixing. Uniform and differential homophily reflect the tendency of edges to occur between herds belonging to the same category within a given attribute (e.g., between commercial herds) that is uniform or varies between categories, respectively. Selective mixing accounts for variation in the occurrence of edges according to the category of each herd within a given attribute. Finally, two quantitative nodal attributes were used: distance between herds (in meters) and size of the herd (Table A4-8).

Following Relun et al. (2017), for the "edges + structure" and "edges + attributes" models, univariable analyses were performed first, and then the covariates were added one by one, until the best model fit was obtained. ERGMs were fitted using package **ergm** (Handcock et al., 2021; Hunter et al., 2008) in R. Goodness of fit of the models were examined by checking for model convergence and degeneracy, comparing the Akaike information criterion, comparing goodness of fit plots for network statistics (in- and out-degree distributions, geodesic distance distribution, and edgewise shared partner distribution), and finally by comparing plots of simulated vs observed networks.

In the provided data on movements of pigs, outgoing movements from a given herd happened every 30 days, corresponding to regular batch shipments. First, we fitted the model to a training subset of the movement data excluding the last 30 days. Then, predictions of the fitted model were compared to the subset of data corresponding to movements in the last 30 days.

The "edges + attributes" model provided the best fit to the observed data (Table A4-9 and Figure A4-8). Herd size, selective mixing between herds according to their production type and commercial activity, as well as movements happening preferably between herds belonging to the same multisite farm, were the important mechanisms of network organisation (Table A4-9).

Table A4-8: Model covariates used to fit the ERGMs of pig movements.

Model covariates $g(y)$	**Abbreviation**
Number of edges	$\mathrm{L}(y)$
Structural network statistics	
Number of isolates	$\mathrm{I}(y)$
Number of asymmetric links	$\mathrm{A}(y)$
Geometrically weighted dyadwise shared partners	$\mathrm{gwdsp}(y, 0.5)$
Geometrically weighted edgewise shared partners	$\mathrm{gwesp}(y, 0.5)$
Geometrically weighted in-degree distribution	$\mathrm{gwid}(y, 0.5)$
Geometrically weighted out-degree distribution	$\mathrm{gwod}(y, 0.5)$
Nodal attributes	
Number of ingoing edges for each production type, activity, housing system and multisite	$\mathrm{In}(y)$
Number of outgoing edges for each production type, activity, housing system and multisite	$\mathrm{Out}(y)$
Number of edges that are within each production type, activity, housing system and multisite, with uniform homophily	$\mathrm{U}(y)$
Number of edges that are within each production type, activity, housing system and multisite, with differential homophily	$\mathrm{D}(y)$
Number of edges that are within and between each production type, activity, and housing system (selective mixing)	$\mathrm{M}(y)$
Euclidean distance between herds	$\mathrm{E}(y)$
Herd size	$\mathrm{S}(y)$

The coefficients from the best ERGM fitted to the training dataset were used to simulate new movement networks using the movements in the last 30 days as observed data. One simulated network thus corresponded to movements between herds for one month. During phase 3, four months were simulated after day 110 (for projections up to day 230); therefore, four networks were simulated for each of the 500 stochastic repetitions of the model, giving a total of 2000 simulated networks.

These simulated networks performed well, reproducing correctly 88.5% (95% CrI: 87.1%-89.9%) of the existing edges between pairs of herds in the observed data ("sensitivity") and 99.984% (95% CrI: 99.982-99.986) of the pairs of herds that were not connected in the observed data ("specificity").

To keep the shipments regular, movement dates were set every 30 days for herds that had outgoing movements of pigs in the observed data. For herds predicted to have outgoing movements in the future but did not in the observed data, movement dates were randomly sampled from a uniform distribution (between day 51 and day 80 for phase 1, between day 81 and day 110 for phase 2, and between day 111 and days 140 or 230 for phase 3).

Table A4-9: Parameter estimates, standard errors (SEs) and Akaike information criterion for the ERGMs of pig movements in phase 3.

Covariates	Estimated ERGM coefficients (SE)[a]		
	Bernoulli (edges)	Edges structure	+ Edges + attributes
Number of edges $\mathrm{L}(y)$	-7.43 (0.01) ***	-5.85 (0.01) ***	-10.53 (0.18) ***
Structural network statistics			
Number of isolates $\mathrm{I}(y)$		3.95 (0.10) ***	
$\mathrm{gwdsp}(y, 0.5)$		-0.64 (0.01) ***	
$\mathrm{gwesp}(y, 0.5)$		3.47 (0.10) ***	
Nodal attributes			
Differential homophily $\mathrm{D}(y)$: multisite[b]			
0 to 0			Reference
1 to 1			6.35 (0.55) ***
2 to 2			6.46 (0.40) ***
…			…
126 to 126			5.84 (0.62) ***
Selective mixing $\mathrm{M}(y)$: production type[c]			
BF to BF			Reference
B to BF			4.45 (0.04) ***
B to F			4.42 (0.04) ***
BF to F			2.02 (0.05) ***
Selective mixing $\mathrm{M}(y)$: activity[d]			
0 to 0			Reference
1 to 0			1.17 (0.18) ***
0 to 1			0.19 (0.20)
1 to 1			1.30 (0.18) ***
Herd size $S(y)$			0.0001 (0.000001) ***
Akaike information criterion	206,749	186,264	149,075

[a] *** p<0.001

[b] For clarity, only estimates for differential homophily within a few multisite farms were shown; 0: no multisite

[c] B: farrow-to-wean herds; BF: farrow-to-finish herds; F: finishing herds

[d] 0: backyard herds; 1: commercial herds

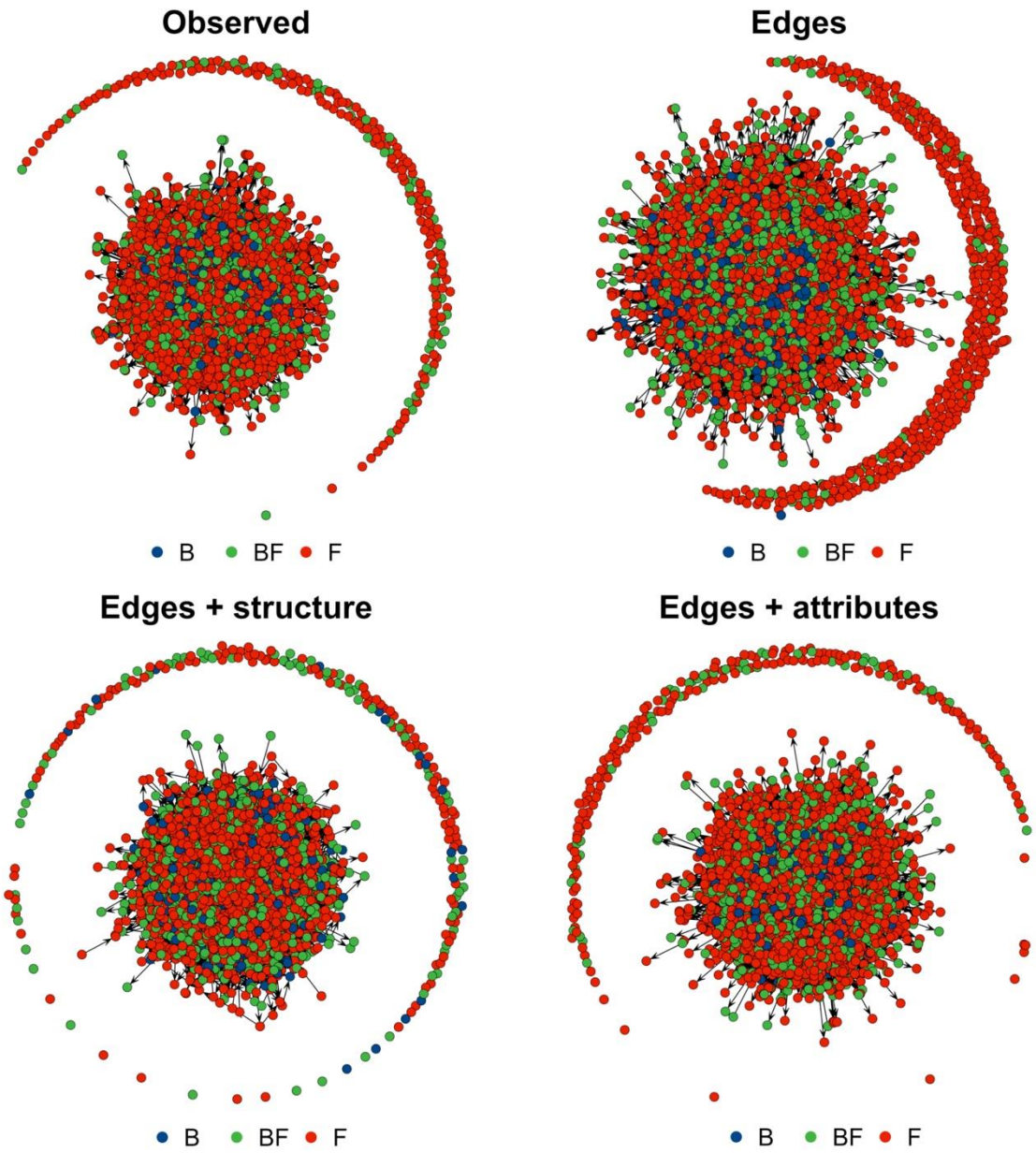


Figure A4-8: Observed and simulated movement networks based on the three ERGMs. Nodes (pig herds) are coloured according to their production type (B: farrow-to-wean herds (blue); BF: farrow-to-finish herds (green); F: finishing herds (red)). The nodes in the outer circle are isolated from the network (no ingoing or outgoing movements). The nodes in the centre belong to the movement network, with arrows representing presence and direction of edges (movements) between two nodes. Node positions were generated using the force-directed placement algorithm of Fruchterman and Reingold (1991), where nodes connected by an edge are placed near each other. Thus, the layout of the nodes in the centre depends on the number of nodes in the network and the number and structure of the edges between them.

Finally, for predicted movements between herds, the number of pigs in the shipment was predicted using a GLM with a zero-truncated negative-binomial distribution (vglm function in the R package **VGAM** (Yee, 2021)). The model was fitted to a training subset containing two-thirds of the number of pigs moved in the observed data. A complete model was built with the following covariates: distance between herds, size, production type, activity and housing system of the source and destination, and whether the source and destination belong to the same multisite. Model selection was performed using Akaike information criterion in a stepwise algorithm.

Estimated coefficients of the model fitted to the training subset with the lowest Akaike information criterion are provided in Table A4-10. Predictions of the model using the remaining third of the data were compared to the observations in Figure A4-9. Given a good match between simulated and observed data, we used the estimated coefficients of the model to predict the number of pigs that were moved between pairs of herds simulated by the ERGM. Overall, at each phase, we simulated 500 sets of data movements (from day 51 to day 80 in phase 1, from day 81 to day 110 in phase 2, and from day 111 to 230 in phase 3) with the same information as in the movement data provided for each phase: the data comprised movement dates, identities of the source and destination herds, and quantity of pigs moved. These simulated data were then used as deterministic events in the mechanistic model of ASFV transmission in pig herds.

Table A4-10: Parameter estimates, standard errors (SEs) and Akaike information criterion for the generalized linear model with zero-truncated negative binomial distribution of number of pigs moved between two herds in phase 3.

Covariates	**Estimated coefficients (SE)**	***p*-value**
Source size	1.56×10^{-5} (1.29×10^{-6})	<0.001
Destination size	8.69×10^{-4} (8.82×10^{-6})	<0.001
Source production type: BF (reference: B)	-2.09 (0.02)	<0.001
Destination production type: F (reference: BF)	0.66 (0.02)	<0.001
Source activity: commercial (reference: backyard)	2.08 (0.09)	<0.001
Destination activity: commercial (reference: backyard)	3.12 (0.03)	<0.001
Source housing system: outdoor (reference: indoor)	-0.04 (0.03)	0.15
Multisite: same (reference: different)	0.14 (0.02)	<0.001
Akaike information criterion	105,616	

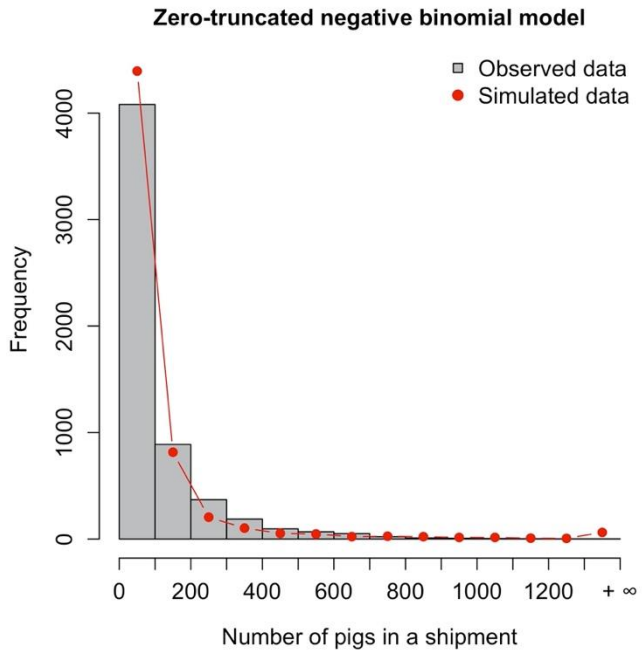


Figure A4-9: Observed and simulated number of pigs in a movement between two herds based on a generalized linear model with zero-truncated negative binomial distribution.

Pigs in each compartment (S, E, Isc, Ic, R) could enter or leave a herd, based on the numbers available in each compartment and the total number of movements. For each herd i with outgoing pig movements, the number of movements for each compartment was determined by generating random two-way tables with given row and column totals using Patefield's algorithm (Patefield, 1981):

$$\begin{pmatrix} m_{i,1,t}^{S} & \cdots & m_{i,i,t}^{S} & \cdots & m_{i,j,t}^{S} & \cdots & m_{i,n,t}^{S} \\ m_{i,1,t}^{E} & \cdots & m_{i,i,t}^{E} & \cdots & m_{i,j,t}^{E} & \cdots & m_{i,n,t}^{E} \\ m_{i,1,t}^{Isc} & \cdots & m_{i,i,t}^{Isc} & \cdots & m_{i,j,t}^{Isc} & \cdots & m_{i,n,t}^{Isc} \\ m_{i,1,t}^{Ic} & \cdots & m_{i,i,t}^{Isc} & \cdots & m_{i,i,t}^{Isc} & \cdots & m_{i,i,t}^{Isc} \\ m_{i,1,t}^{R} & \cdots & m_{i,i,t}^{R} & \cdots & m_{i,j,t}^{R} & \cdots & m_{i,n,t}^{R} \end{pmatrix} \quad \text{(A4.7)}$$

where $m_{i,j,t}^{X}$ is the number of pigs in compartment X in herd i moving to herd j at time t. The row and column totals were fixed, with the total number of pigs moving out from herd i to herd j (or staying in herd i if $j = i$): $m_{i,j,t}^{S} + m_{i,j,t}^{E} + m_{i,j,t}^{Isc} + m_{i,j,t}^{Ic} + m_{i,j,t}^{R}$ (column total) being the observed or predicted number of pigs moving from herd i to herd j at time t, and the number of pigs moving or staying in compartment X (row total: $\sum_{j=1}^{n} m_{i,j,t}^{X}$) equal to the number of pigs available in each compartment.

4.6.1.3 Basic disease management

4.6.1.3.1 Wild boar

For wild boar carcasses, two main types of surveillance were carried out in the island: (1) passive surveillance, which involved the removal and reporting of found wild boar carcasses, and (2) active surveillance, which involved active search for wild boar carcasses around already detected infected carcasses.

We implemented active surveillance in wild boar as follows: for each patch (7.5 km^2 area) where infectious boar was found at day $t-1$, searching was made to take place within that patch at day t, such that carcasses in those patches were more likely to be removed than those in patches where infectious boar had not been found previously. This procedure was not implemented directly in phase 1, as the structure of the model at that stage did not enable tracking of individual boar. However, for phase 1, the larger value of the patch prevalence rate parameter (r_2) used within the prediction period (day 51-day 78), compared to that used before day 28 (r_1) captured the increased likelihood of detecting a positive boar in a patch where infections had been found previously.

For live boar, surveillance was carried out through hunting: if hunting took place before day 60 or outside the fence and buffer zone, the probability of detecting an infection in a boar killed by hunting was 0.2, as only 20% of hunted boar were tested during that period. If hunting took place within the fence and buffer zone, and between (and including) day 60 and day 120, the probability of detecting an infection in a boar killed by hunting was 1, as all boar hunted within the fence and buffer zone during this period were tested (we assume the diagnostic procedures used in the island were perfectly sensitive and specific).

4.6.1.3.2 Pig herds

Based on the description of the disease management measures provided by the challenge coordinators, the following measures were implemented in our model: (1) suspected pig herds were confirmed infected three days after suspicion, assuming perfect ASFV diagnostic tests (100% sensitivity and specificity); (2) all herds with confirmed infection were culled the day after confirmation (four days after suspicion), by setting all compartments to zero (including residues from dead pigs, i.e., assuming cleaning and disinfection were effective immediately); (3) after a herd was confirmed infected, a protection zone (3 km radius for 40 days) and a surveillance zone (10 km radius for 30 days) were defined, and at-risk herds that traded pigs with infected herds (ingoing or outgoing movements) within the previous three weeks were traced; (4) all movements of pigs (ingoing or outgoing) were banned for 40 days in protection zones and in at-risk herds, and for 30 days in surveillance zones; (5) awareness of farmers about ASF was improved in surveillance and protection zones, as well as for at-risk herds; (6) repopulation of a culled herd was allowed 50 days after culling (except if the herd was still in a protection or surveillance zone), assuming that all pigs used for repopulation were susceptible.

Detection by passive surveillance in pig herds was assumed to occur when two conditions were met: (1) when the mortality rate caused by ASFV during the previous 14 days in the herd was more than 6% (Andraud et al., 2019); and (2) the number of clinical or dead animals in the herd reached a minimum value of five during the previous 14 days (Halasa et al., 2016c). The minimum number of clinical or dead animals was introduced to represent more accurately the probability of detecting abnormal events, especially in small herds, where only one death could make the mortality rate exceed the threshold (Halasa et al., 2016c). Increased awareness of farmers in protection and surveillance zones and in at-risk herds was represented in the model by reducing the minimum number of clinical or dead animals required for detection to one (Halasa et al., 2016c).

4.6.1.4 Initial conditions

4.6.1.4.1 Initial populations

4.6.1.4.1.1 Wild boar

For efficiency, we considered only wild boar located within a defined area around the detected cases in wild boar and pig herds. In phase 1, this area measured 40,000 km^2; however, this was reduced in phases 2 and 3 to 18,750 km^2 for the sake of computational efficiency. Our decision to reduce the area was motivated by the observation that the area within which infections had been found by phase 1 was significantly smaller compared to the area we had initially considered, hence an appreciable number of patches in our phase 1 model had a negligible risk of getting infected. Thus, the patch boundaries were updated to exclude such patches without artificially limiting the range of possible transmission events over the time periods we considered.

The abundance of wild boar in each patch was determined based on the hunting bag data from 2019 in each administrative region and the estimate of the hunting rate in 2019. These data were given by the challenge organizers. Given that 50% of the wild boar population was hunted by the end of the 2019 hunting season and assuming the change in the size of the wild boar population from 2019 to 2020 was negligible, we estimated the number of wild boar in each administrative region at the start of the 2020 hunting season by doubling the provided region-level hunting figures for 2019.

Also, given that the hunting season began 36 days before day 0, we estimated the number of boar remaining in the island by day 0 by subtracting the number of boar that had already been removed from day -36 to day 0 from our estimate of the number of boar at the start of the hunting season (day -36). To estimate the number of boar removed from the landscape between day -36 and day 0, we multiplied the fraction of boar f expected to be removed within this period by the number of boar in the island by day -36. For each region, f was a constant given by 36/240: this value was arrived at by considering that the hunting season in Merry Island lasts 240 days and by assuming that the daily hunting rates are constant throughout the season. This estimation was performed separately for each region.

Having obtained an estimate for the number of boar expected to be in the landscape by day 0, we simulated locations for those boar for whom (location and test) data had not been reported by day 50. Boar in this category had either been removed but had not been tested (and hence not reported) or had not been removed (i.e., remaining in the landscape). The number of coordinates to be simulated was determined for each region separately, by subtracting the number of boar that had already been observed in that region (by day 50) from the number of boar estimated to be in that region at day 0. The locations for the simulated coordinates were chosen uniformly at random within region boundaries.

Boar were then assigned to patches, based on their coordinates: a boar belonged to a given patch if its coordinates were within the boundaries of the patch.

4.6.1.4.1.2 Pig herds

Pig herds were distributed according to the coordinates provided. We considered all known herds in the island. Each herd was initialized with susceptible pigs based on its herd size provided by the challenge coordinators.

4.6.1.4.2 Initial infections

4.6.1.4.2.1 Wild boar

For ease of explanation, we will explain the seeding of infections at each of the three phases in reverse chronological order. In phase 3, simulation began at day 60. The model was seeded with:

1. Category A: infections among boar that had been observed as ASFV-positive between day 59 and day 74. (Since death from ASFV infection in wild boar usually occurs within 14 days of infection (Blome et al., 2020; Pietschmann et al., 2015), we assumed that boar observed as positive between day 59 and day 74 were likely to be positive on day 60.)
2. Category B: additional infections among boar that had not been removed by day 60. (Details are as follows.)

The number of additional infections seeded (Category B) was determined by subtracting the number of boar in Category A from the number of positive boar expected to be in the landscape at day 60. The expected number of infections in wild boar by day 60 was estimated by multiplying the prevalence observed by day 59 by the number of boar remaining in the landscape on day 60.

The prevalence by day 59 was estimated by dividing the number of positive boar removed up to day 59 by the total number of boar (positive and negative) removed up to that day. The number of positive boar removed up to day 59, p_{59}, was computed as:

$$p_{59} = p^*_{59} + 5\,\tilde{p}_{59} \qquad \text{(A4.8)}$$

where p_{59}^{*} is the number of infected boar found by active or passive surveillance by day 59 and $\tilde{p}_{59}$ is the number of detected infected boar hunted by day 59. Between day 0 and day 59, only 20% of hunted boar were tested (and reported). Thus, to obtain the total number of positive boar hunted, the observed number of positive boar hunted was multiplied by 5.

In phase 2, the process was similar to that described for phase 3, except these two differences:

1. Simulation began at day 1 (and not day 60).
2. Boar in Category A were positive boar removed between day 0 and day 8.

In phase 1, the model was seeded at day 1 with a 1% random selection of the infections which had been observed by day 50.

4.6.1.4.2.2 Pig herds

At each phase, ASFV was introduced in pig herd 3594 (the first detected infected pig herd) by replacing one susceptible pig (*S*) by an exposed pre-infectious pig (*E*) at day −31, giving a median suspicion date in the model at day 0 and a median detection date at day 3, as observed in the data.

Table A4-11: Clock times elapsed for a single projection implementation at each phase. The projection periods for phases 1 and 2 were days 51-78 and days 81-110, respectively. The projection periods for phase 3 re days 111-140 and days 141-230. To maximize efficiency, the pig herd and wild boar components of the model were run on separate systems. Simulations for the wild boar model were run on a 2.4GHz processor with 62GB of RAM on a Linux 64-bit operating system while simulations for the pig herd model were run on a 2GHz processor with 257GB of RAM on a Linux 64-bit operating system.

Phase	Wild boar component	Pig herd component*
1	0 minutes 8 seconds	1 minute 30 seconds
2	2 minutes 33 seconds	4 minutes 30 seconds
3	By day 140: 3 minutes 17 seconds By day 230: 4 minutes 24 seconds	By day 140: 4 minutes 55 seconds By day 230: 8 minutes

*Includes the time required to compute movements and external infectious pressure from wild boar.

4.7 APPENDIX B: MODEL FIT

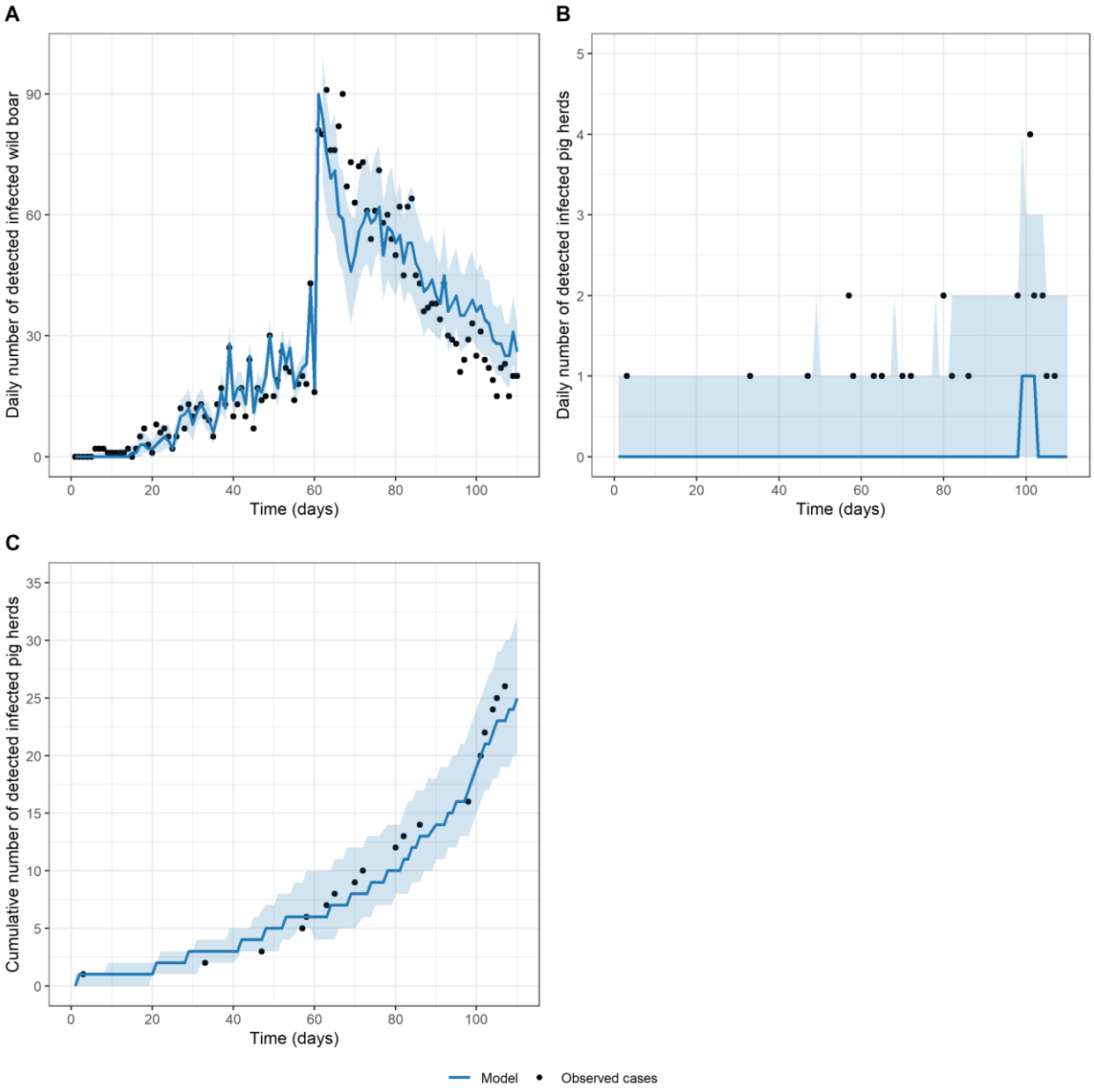


Figure B4-10: Observed and estimated number of detected infected wild boar and detected infected pig herds from day 1 to day 110. **A**: Daily number of detected infected wild boar. **B**: Daily number of detected infected pig herds. **C**: Cumulative number of detected infected pig herds. Median model estimates (blue line) are shown along with 95% CrIs (shaded areas) and observed counts (black dots). Model estimates were computed from 500 stochastic repetitions of the model corresponding to best-fitting parameter values. The model incorporated disease management measures as observed: normal hunting pressure from day 1 to day 59 and increased hunting pressure (within fence) from day 60 to day 110.

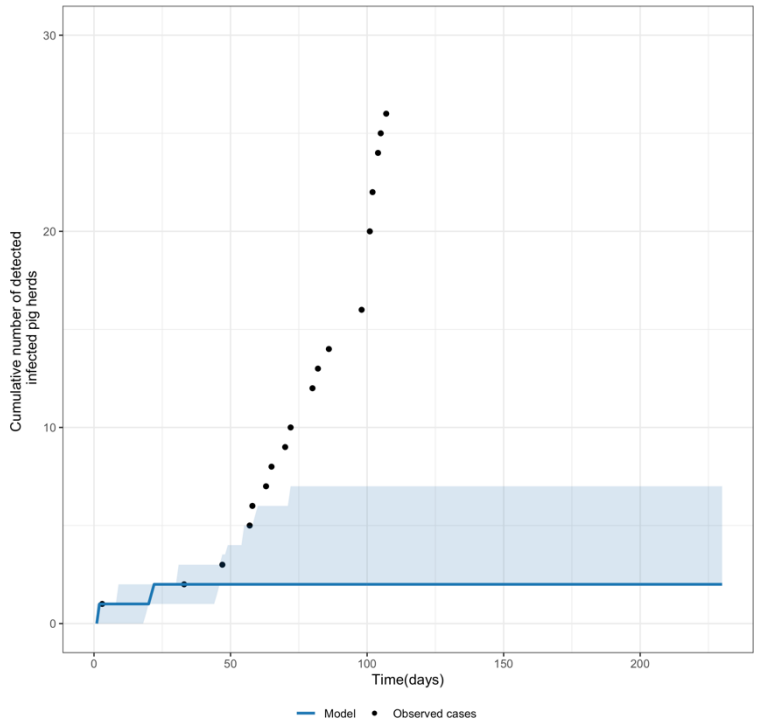


Figure B4-11: Observed cumulative number of detected infected pig herds from day 1 to day 110 (black dots) and estimated cumulative number of detected infected pig herds from day 1 to day 230 in the absence of transmission of ASFV from wild boar to pig herds (blue line). Shaded areas represent 95% CrIs for the model projections, computed from 500 stochastic repetitions of the model corresponding to best-fitting parameter values

4.8 APPENDIX C: SENSITIVITY ANALYSIS

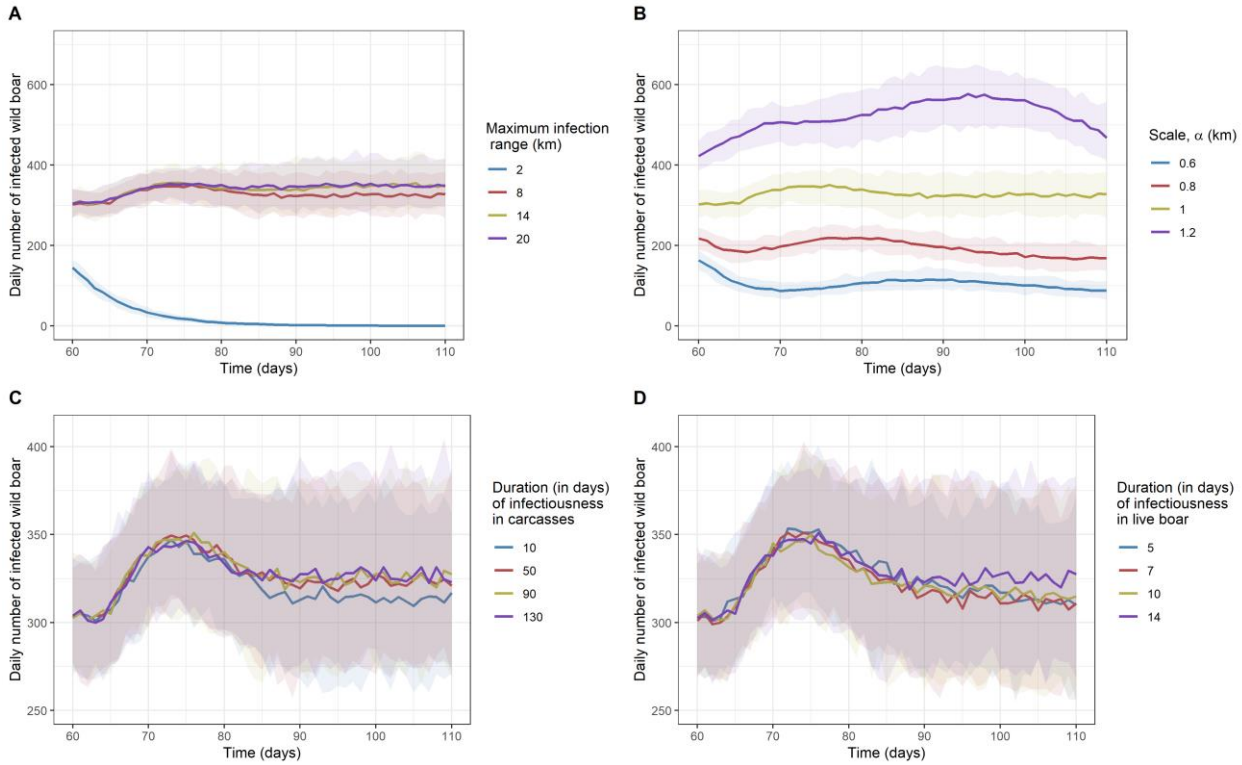


Figure C4-12: Sensitivity of the daily number of infections in wild boar (from day 60 to day 110) to **A**: the MIR (values: 2 km, 8 km, 14 km, 20 km); **B**: the scale parameter α of the dispersal kernel (values: 0.6 km, 0.8 km, 1 km, 1.2 km); **C**: the duration of infectiousness in carcasses (values: 10 days, 50 days, 90 days, 130 days); and **D**: the duration of infectiousness in live boar (values: 5 days, 7 days, 10 days, 14 days). Trajectories are medians computed from 100 stochastic repetitions of the model. Shaded regions are corresponding 95% CrIs. Note that the scale of the y-axis differs across plots. Simulations corresponding to baseline parameter values (MIR = 8 km, α = 1 km, duration of infectiousness in carcasses = 90 days; and duration of infectiousness in live boar = 14 days) are the same across panels.

Chapter 5 Structural identifiability of compartmental models for infectious disease transmission is influenced by data type

This chapter presents the following manuscript:

> **Dankwa, E.A.**, Brouwer, A.F., Donnelly, C.A. Structural identifiability of compartmental models for infectious disease transmission is influenced by data type.

A statement of specific contributions by the author of this thesis is provided at the end of the chapter.

ABSTRACT

If model identifiability is not confirmed, inferences from infectious disease transmission models may not be reliable, so they might lead to misleading recommendations. Structural identifiability analysis characterizes whether it is possible to obtain unique solutions for all unknown model parameters, given the model structure.

In this work, we studied the structural identifiability of some typical deterministic compartmental models for infectious disease transmission, focusing on the influence of the data type considered as model output on the identifiability of unknown model parameters, including initial conditions.

We defined 26 model versions, each having a unique combination of underlying compartmental structure and data type(s) considered as model output(s). Four compartmental model structures (SIR, SLIR (L: Latent), SLIR allowing for relapse and vaccination, and a vector-borne disease model with SLIR for hosts and SLI for vectors) and three common data types in disease surveillance (incidence, prevalence and detected vector counts) were studied.

The structural identifiability of some parameters varied depending on the type of model output. In general, models with multiple data types as outputs had more structurally identifiable parameters, than did models with a single data type as output.

This study highlights the importance of a careful consideration of data types as an integral part of the inference process with compartmental infectious disease transmission models.

Keywords: structural identifiability, infectious disease transmission, compartmental models, data types

5.1 INTRODUCTION

Goals of infectious disease transmission modelling often include making inferences about the underlying transmission process, predicting the future course of an epidemic given a range of interventions, or estimating what would have happened in a counterfactual scenario. A defined model is fitted to a given data set (frequently an incidence time series generated by passive surveillance). This model-fitting process is parameter estimation, where one determines parameter values or distributions corresponding to model outputs that best fit (or at least, approximate) the observed data. Parameter estimation however can only produce robust results if the model is *identifiable* (Audoly et al., 2001; Castro and de Boer, 2020; Cobelli and Distefano III, 1980; Kao and Eisenberg, 2018; Ljung and Glad, 1994; Villaverde et al., 2016; Wieland et al., 2021); that is, if it is possible, in principle, to obtain unique solutions for all unknown model parameters, given the model structure and available data. We note that other properties such as predictability (Castro et al., 2020; Scarpino and Petri, 2019) and uncertainty quantification (Capaldi et al., 2012; McCabe et al., 2021) also affect the reliability of model inferences (Massonis et al., 2021a); these are not treated here, however.

Although the subject of identifiability has received considerable attention in the systems biology and control literature (see (Wieland et al., 2021) for a recent review), it is inconsistently applied in the infectious disease modelling literature. Relatively few studies exist on the identifiability analysis of infectious disease models (Brouwer et al., 2018; Eisenberg et al., 2013; Evans et al., 2005; Kao and Eisenberg, 2018; Massonis et al., 2021a; Tuncer et al., 2016; Tuncer and Le, 2018) and the practice of routinely checking the identifiability of these models *before* parameter estimation is not widespread. Nevertheless, identifiability is required to make meaningful inferences on model parameters and, consequently, to provide reliable evidence to inform public health policymaking.

In a non-identifiable model, parameter sets with similar values may yield considerably different model predictions (Kao and Eisenberg, 2018; Roda et al., 2020). Thus, a failure to consider identifiability could result in misleading recommendations, as has been previously noted (Kao

and Eisenberg, 2018; Massonis et al., 2021; Roda et al., 2020; Roosa and Chowell, 2019), some of which could have serious consequences. For example, Kao and Eisenberg demonstrated using a dengue transmission model that two sets of parameters which fit the incidence data comparably well yielded very different predictions for incidence after an intervention is applied (see Fig. 9 in their paper) (Kao and Eisenberg, 2018). Roda and colleagues also showed that the lack of identifiability in COVID-19 transmission models could lead to extreme variability in predictions (Roda et al., 2020).

A distinction is made between the two types of identifiability: *structural identifiability* and *practical identifiability*. Structural identifiability, a concept first introduced by Bellman and Astrom in 1970, is a property of the model structure and associated measurement function (i.e., the function of model variables that is to be observed) and does not depend on the quantity or quality of the observed data. It addresses the question: Given an error-free model structure, and assuming noise-free, infinite data, do unique solutions exist for the model parameters? Structural identifiability is affected by: 1) the nature of the model parameterization (Muñoz-Tamayo et al., 2018) which influences symmetries, i.e., functional relationships between model parameters (Eisenberg and Hayashi, 2014; Hengl et al., 2007; Massonis et al., 2021a; Villaverde, 2022); and 2) the data type considered as model output (Balsa-Canto et al., 2010; Chis et al., 2011a; Massonis et al., 2021a; Tuncer and Le, 2018). Practical identifiability, on the other hand, is related to the adequacy of the available observed data for the estimation problem at hand (Balsa-Canto et al., 2010; Brouwer et al., 2017; Kao and Eisenberg, 2018; Miao et al., 2011; Raue et al., 2009; Tuncer et al., 2016; Tuncer and Le, 2018). The corresponding question is: Do the data contain enough information to infer the model parameters? Structural identifiability is a necessary, but not sufficient, condition for practical identifiability; that is, a structurally non-identifiable parameter cannot be practically identifiable, and a structurally identifiable parameter could be practically non-identifiable depending on the data available (Cobelli and Distefano III, 1980; Eisenberg et al., 2013). In

this work, we are considering structural identifiability of infectious disease transmission models.

Several studies have demonstrated the influence of the type of observed data on the structural identifiability of infectious disease models. For example, Tuncer and Le studied a Susceptible-Infected-Treated-Recovered epidemic model which becomes structurally identifiable only when both cumulative incidence rates and the number of treated individuals is observed (Tuncer and Le, 2018). In the same work, the authors explored the identifiability of a Susceptible-Exposed-Infected-Recovered model and showed that the type of structural identifiability for two parameters (recovery rate and length of latent period) depended on whether the observed data were cumulative incidence or prevalence. Similar works include (Evans et al., 2005), on the structural identifiability of a seasonally forced SIR model with prevalence and a proportion of the incidence as outputs; (Eisenberg et al., 2013), on the identifiability of parameters of compartmental models for cholera with prevalence as output; (Tuncer et al., 2016), on the identifiability of an immune-epidemiological model for Rift Valley fever with time-series data of viremia levels as output; (Kao and Eisenberg, 2018), on the identifiability of a dengue transmission model with various types of human and mosquito incidence data as outputs; and more recently, (Massonis et al., 2021a), on the structural identifiability of a wide range of COVID-19 transmission models with a variety of surveillance data types as outputs.

However, few of these studies (e.g., (Eisenberg et al., 2013; Evans et al., 2005)) have explicitly studied the identifiability of unknown ICs. Other studies have either assumed known ICs (e.g., (Tuncer and Le, 2018)) or have implicitly considered unknown ICs through assessment of the *observability* of model states (Massonis et al., 2021a); i.e., whether the state variable trajectories can be uniquely determined from observed data. (Structural identifiability has been considered as a particular case of observability (Massonis et al., 2021a; Sedoglavic, 2002; Tunali and Tarn, 1987; Villaverde, 2019).) Here, we explicitly consider ICs as unknown parameters in all models and analyse their structural identifiability given various

data types. Often values are assumed for ICs, but careful analysis often reveals that parameter estimates depend on these IC assumptions. We can ask under what circumstances ICs can be uniquely determined from observed data. Although this question might technically be considered one about observability, when the ICs are reframed as parameters, the question is one of identifiability. Thus, our work adds to the literature by examining how the structural identifiability of ICs of classic compartmental models change with data type. Additionally, we employ a publicly available web-based toolbox, SIAN (Hong et al., 2019), to analyse the structural identifiability of model parameters, allowing us to demonstrate the utility of such tools.

Specifically, we consider four compartmental structures (SIR, SLIR, SLIR with vaccination and relapse and a vector-borne disease model with SLIR for hosts and SLI for vectors) and three common data types in disease surveillance (incidence, prevalence and detected vector counts). Using SIAN, we analyse the structural identifiability of unknown parameters in 26 model versions, each a unique combination of underlying compartmental structure and data type considered as model output. We use the term “model version” to refer to a compartmental structure-output(s) combination; e.g., SIR with incidence, or SLIR with incidence and prevalence.

Although the compartmental structures and data types we consider are by no means exhaustive, our work is intended to demonstrate the importance of identifiability and to be instructive for those seeking to apply these techniques to their own models. We have therefore made available all input codes and output files to facilitate reproducibility: structural-identifiability-epi-models · GitFront.

The chapter is outlined as follows. In Section 5.2, we introduce the general modelling framework and notation and provide formal definitions of relevant structural identifiability concepts. Here, we also introduce the four compartmental structures, briefly introduce the software toolbox utilized, present the model versions examined and finally, outline the

structural identifiability analysis performed. Section 5.3 presents the results and Section 5.4 presents a discussion of results. Concluding remarks are given in Section 5.5.

5.2 METHODS

5.2.1 General modelling framework and formal definitions

Consider a deterministic ordinary differential equation (ODE) infectious disease transmission model $\mathcal{M}$ of the form

$$\mathcal{M} \coloneqq \begin{cases} \dot{\boldsymbol{X}}(t) = f\big(\boldsymbol{X}(t), \boldsymbol{p}, \boldsymbol{u}(t)\big) \\ \boldsymbol{y}(t) = g(\boldsymbol{X}(t), \boldsymbol{p}) \\ \boldsymbol{X}_{t_0} = \boldsymbol{X}(t_0) \end{cases}, \tag{5.1}$$

with observations on the interval $t_0 \leq t \leq T$, where $\dot{\boldsymbol{X}}(\boldsymbol{t})$ is a system of non-linear ODEs, $\boldsymbol{X}(\boldsymbol{t}) \in R^{n_X}$ is a vector of time-varying disease states and the unique solution to the system $\dot{\boldsymbol{X}}(\boldsymbol{t})$, $\boldsymbol{p} \in R^{n_p}$ is a vector of constant unknown model parameters, $\boldsymbol{y}(t) \in R^{n_y}$ is a vector of time-dependent model outputs corresponding to a specific data type (for example, case incidence rates), $\boldsymbol{u}(t) \in R^{n_u}$ is a time-dependent input vector, g is the measurement equation (which defines the relationship between $\boldsymbol{X}(\boldsymbol{t})$, $\boldsymbol{p}$ and $\boldsymbol{y}$), and $\boldsymbol{X}_{t_0} \subset R^{n_X}$ is a vector of known ICs. Note that unknown components of $\boldsymbol{X}_{t_0}$ are included in $\boldsymbol{p}$ and that f and g are vectors of analytic functions of their arguments.

The formal definition of structural identifiability for a model and its parameters is given below. The structural identifiability of a parameter may either be *local* (i.e., holding only within a limited region of the parameter space or about a given point) or *global* (i.e., holding (almost) everywhere within the parameter space) (Ljung and Glad, 1994).

Definition 1 [Parameter structural identifiability] (Cobelli and Distefano III, 1980; Ljung and Glad, 1994)

A parameter $p_i \in \boldsymbol{p}$ *is structurally globally identifiable (s.g.i.) on* $[t_0, T]$ *for a given output choice* $\boldsymbol{y}$ *if a unique solution exists for* p_i*; that is, if and only if for almost any* $\boldsymbol{p}^*$ *and almost any IC (i.e., excluding degenerate values),* $\boldsymbol{y}(\boldsymbol{X}, \hat{\boldsymbol{p}}) = \boldsymbol{y}(\boldsymbol{X}, \boldsymbol{p}^*)$ *implies* $\hat{p}_i = p_i^*$*. Otherwise,* p_i *is structurally globally non-identifiable.*

A parameter $p_i \in \boldsymbol{p}$ *is structurally locally identifiable (s.l.i.) on* $[t_0, T]$ *for a given an output* $\boldsymbol{y}$ *if there exists a neighbourhood* $V(\boldsymbol{p})$ *of the parameter space within which a unique solution exists for* p_i*. Otherwise,* p_i *is structurally non-identifiable (s.n.i.).*

Definition 2 [Model structural identifiability] (Cobelli and Distefano III, 1980; Ljung and Glad, 1994)

The model $\mathcal{M}$ *is s.g.i. for a given output* $\boldsymbol{y}$ *if every* $p_i \in \boldsymbol{p}$ *is s.g.i.* given $\boldsymbol{y}$*.*

The model $\mathcal{M}$ *is s.l.i. for a given output* $\boldsymbol{y}$ *if every* $p_i \in \boldsymbol{p}$ *is s.l.i. given* $\boldsymbol{y}$.

The model M is s.n.i. for a given output $\boldsymbol{y}$ *if at least one* $p_i \in \boldsymbol{p}$ *is s.n.i. given* $\boldsymbol{y}$**.**

5.2.2 Model structures

The most basic model structure we consider is the SIR model. For simplicity, we assume no demography, no migration, homogenous populations and a constant, population size N. In this SIR model, there are three mutually exclusive compartments, each corresponding to a distinct infection state: Susceptible S, Infectious I and Recovered (and immune) R. Susceptible individuals become infected at a rate $\beta I/N$ where β is the transmission rate and is equal to the product of the contact rate and the probability that a contact will successfully result in an infection. Infectious individuals recover at a rate γ. These dynamics can be described by the following set of ODEs:

SIR:
$$\begin{cases} \frac{dS}{dt} = -\frac{\beta \mathrm{SI}}{N} \\ \frac{dI}{dt} = \frac{\beta \mathrm{SI}}{N} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} . \tag{5.2}$$

ICs for the S, I and R states will be denoted by $\mathrm{S}(0), \mathrm{I}(0)$ and $\mathrm{R}(0)$, respectively. At any time $t \geq 0$, $\mathrm{N} = \mathrm{S}(t) + \mathrm{I}(t) + \mathrm{R}(t)$. For the SIR model, we consider two outputs: incidence, $y_1 = \beta \mathrm{SI}/N$, and prevalence, $y_2 = \mathrm{I}/N$.

For diseases with a non-negligible latent period (e.g., COVID-19 (Liu et al., 2020)), the SIR model can be modified to include a latent state L. The modified dynamics are described by the following set of equations:

SLIR:
$$\begin{cases} \frac{dS}{dt} = -\frac{\beta \mathrm{SI}}{N} \\ \frac{dL}{dt} = \frac{\beta \mathrm{SI}}{N} - \alpha L \\ \frac{dI}{dt} = \alpha L - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} , \tag{5.3}$$

where $1/\alpha$ represents the length of the latent period. Let $\mathrm{L}(0)$ denote the IC for the latent state. For all $t \geq 0$, $\mathrm{N} = \mathrm{S}(t) + \mathrm{L}(t) + \mathrm{I}(t) + \mathrm{R}(t)$. We study an equivalent set of outputs as for the SIR model: incidence, $y_3 = \alpha L$, and prevalence, $y_4 = \mathrm{I}/N$.

For diseases where a relapse of symptoms is possible after a period of remission (e.g., hepatitis A), we can include a compartment Q to represent the remission state. In this model,

we also allow for immunity by vaccination. The dynamics of this SLIRQ (Susceptible-Latent-Infectious-Recovered (or immune)-Remission) model as adapted from Dankwa et al. (2021) are as follows. In this model, individuals in the R compartment are immune, either as a result of vaccination or past infection. Susceptible individuals become exposed at a rate $\beta \mathrm{I}/N$ and move to the latent state, where they remain for $1/\alpha$ time units, after which they become infectious. A proportion, $1-\eta$, of infectious individuals recover temporarily, moving to the remission state for a period of $1/\sigma$ time units, after which they experience a relapse of symptoms, becoming infectious. The remaining proportion, η, of infectious individuals recover permanently and become immune. The recovery rate is γ. A number $v(t)$ of individuals are vaccinated at time t and become immune. These dynamics are captured by the following set of ODEs:

SLIRQ:

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta \mathrm{SI}}{N} - v\dfrac{S}{N} \\ \dfrac{dL}{dt} = \dfrac{\beta \mathrm{SI}}{N} - \alpha L \\ \dfrac{dI}{dt} = \alpha L - \gamma I + \sigma Q \\ \dfrac{dR}{dt} = v\dfrac{S}{N} + \eta\gamma I \\ \dfrac{dQ}{dt} = (1-\eta)\gamma I - \sigma Q \end{cases} \tag{5.4}$$

The IC corresponding to the remission state will be denoted by $\mathrm{Q}(0)$. For all $t \geq 0$, $\mathrm{N} = \mathrm{S}(t) + \mathrm{L}(t) + \mathrm{I}(t) + \mathrm{R}(t) + \mathrm{Q}(t)$. We consider the same set of outputs as before: incidence, $y_5 = \alpha L + \sigma Q$, and prevalence, $y_6 = \mathrm{I}/N$.

Finally, we introduce a SLIR/SLI model structure suitable for vector-borne diseases, and adapted from the works of Ngwa and Shu (2000) and Kao and Eisenberg (2018), who apply the model to malaria and dengue, respectively. In the model, infection dynamics within the

host population are explained via a SLIR model, as in equation (5.3), while the dynamics in the vector population are explained via a SLI model, thus a SLIR/SLI model. Transmission can only occur between individuals of different populations, i.e., host-to-vector or vector-to-host. Like in the previous models, we assume constant sizes for both populations: let N_h and N_v represent the sizes of the host and vector populations, respectively. We use subscripts "h" and "v" to represent compartments for hosts and vectors, respectively. Thus, we have $N_h = S_h(t) + L_h(t) + I_h(t) + R_h(t)$ and $N_v = S_v(t) + L_v(t) + I_v(t)$, $\forall t \geq 0$.

The pathogen transmission rate from host to vector β_{hv} is equal to the product of the contact rate between host and vector (in malaria for example, this may be the human biting rate of mosquitoes) and the probability of successful transmission from an infectious host to a susceptible vector. Similarly, the transmission rate from vector to host β_{vh} is equal to the product of the contact rate between vector and host and the probability of successful transmission from an infectious vector to a susceptible host. Infected hosts become infectious after a latency period of $1/\alpha_h$ time units and remain infectious for a period of $1/\gamma_h$ time units before recovery. Recovered hosts become immune to the disease. Infectious hosts transmit the pathogen to susceptible vectors at a rate $\beta_{hv} I_h/N_h$. Infected vectors become infectious after a latency period of $1/\alpha_v$ time units. Infectious vectors transmit the pathogen to susceptible hosts at a rate $\beta_{vh} I_v/N_h$. Within each population, we assume equal birth and death rates: μ_h and μ_v for hosts and vectors, respectively, so no disease-related mortality is incorporated. The SLIR/SLI model is represented by the following system of ODEs:

SLIR/SLI:

$$\begin{cases} \frac{dS_h}{dt} = \mu_h N_h - \frac{\beta_{vh} S_h I_v}{N_h} - \mu_h S_h \\ \frac{dL_h}{dt} = \frac{\beta_{vh} S_h I_v}{N_h} - \alpha_h L_h - \mu_h L_h \\ \frac{dI_h}{dt} = \alpha_h L_h - \gamma_h I_h - \mu_h I_h \\ \frac{dR_h}{dt} = \gamma_h I_h - \mu_h R_h \\ \frac{dS_v}{dt} = \mu_v N_v - \frac{\beta_{hv} S_v I_h}{N_h} - \mu_v S_v \\ \frac{dL_v}{dt} = \frac{\beta_{hv} S_v I_h}{N_h} - \alpha_v L_v - \mu_v L_v \\ \frac{dI_v}{dt} = \alpha_v L_v - \mu_v I_v \end{cases} . \quad (5.5)$$

The ICs for the SLIR/SLI model will be denoted by (listed in order of states): $S_h(0), E_h(0), I_h(0), R_h(0), S_v(0), L_v(0)$ and $I_v(0)$.

The following outputs are studied: 1) incidence in hosts (host incidence), $y_7 = \alpha_h L_h$; 2) prevalence in hosts (host prevalence), $y_8 = I_h/N_h$; 3) incidence in vectors (vector incidence), $y_9 = \alpha_v L_v$; and 4) detected vector counts, $y_{10} = \lambda_v(S_v + L_v + I_v)$, λ_v is an unknown vector detection rate.

5.2.3 Toolbox employed

In this study, we employ the SIAN (Structural Identifiability ANalyser) (Hong et al., 2019) software tool for structural identifiability analysis. The algorithm implemented in SIAN, proposed by Hong et al. (2020), is based on a combination of differential algebra and Taylor series approaches to structural identifiability analysis. SIAN is implemented in Maple and available as a web application: https://maple.cloud/app/6509768948056064.

Here, we are interested in assessing both local and global identifiability of model parameters, including ICs. Therefore, although other toolboxes exist which are capable of assessing the structural identifiability of $\mathcal{M}$, we employ SIAN because it uniquely possesses the following characteristics as desired for this study: 1) it is capable of assessing both local and global identifiability of model parameters; 2) it provides identifiability results for ICs of state variables; and 3) it is available as a web application and accepts a simple text-based input, hence more accessible than toolboxes which require program installation or knowledge of a particular programming language. This latter characteristic is a particularly desirable one for a structural identifiability analysis software, as it addresses a potential barrier to the application of structural identifiability analysis (see discussion in Section 1.2 on other barriers).

For a given model, SIAN typically produces one of the following results for the structural identifiability of each model parameter: s.g.i., s.l.i. or s.n.i. SIAN is also capable of computing identifiable combinations, although we do not employ that functionality here.

5.2.4 Structural identifiability assessments

Structural identifiability analysis was conducted in four stages, each stage designed to reflect a possible scenario as may be encountered when modelling infectious disease transmission. Across these stages, we studied the structural identifiability of model parameters given three common data types as model outputs – incidence, prevalence, and detected vector counts (the latter only applicable to SLIR/SLI). We analysed 26 ODE model versions, assuming in all cases constant, unknown population sizes. For each model, we assessed the structural identifiability of all unknown parameters, including ICs.

Stages are now described.

- **Stage one (single outputs, all parameters unknown)**: Structural identifiability analysis was conducted for models defined with a single data type as output and assuming all parameters were unknown. This scenario is typical in the initial stages of

an outbreak of an emerging pathogen, when little is known of pathogen epidemiology and consequently, natural history parameters or transmission rates. Furthermore, in such scenarios, as data are often limited, only one type of data may be available for parameter estimation. It is therefore of interest to determine which parameters are structurally identifiable in such contexts.

Therefore, for SIR, SLIR and SLIRQ, we assessed the structural identifiability of model parameters given either incidence or prevalence data. For SLIR/SLI, output was host incidence or host prevalence. We do not consider vector data at this stage, as such data are less likely to be available during the early stages of an emerging vector-borne disease outbreak. Thus, at this stage, eight model versions were analysed.

- **Stage two (single outputs, only natural history parameters known)**: In the case of an endemic disease which has been widely studied (e.g., malaria in sub-Saharan Africa), a high level of certainty may be obtained on the values of natural history parameters. In modelling such diseases, knowledge of natural history parameters may be assumed and hence these parameters may be treated as known quantities in the model. Stage two considers this scenario. For the model versions analysed at stage one, we assumed all natural history parameters to be known and re-evaluated the structural identifiability of the other (unknown) model parameters, i.e., all ICs, transmission rate parameters, and for the SLIR/SLI models, the birth rate parameters, additionally. This analysis enabled us to identify how the structural identifiability properties of unknown parameters change once other parameters in the model are assumed known. As in stage one, eight model versions were analysed at this stage.

- **Stage three (multiple outputs, all parameters unknown)**: In instances where surveillance capacities are strengthened in the face of an emerging outbreak, it is possible to observe more than one type of data. For example, in the context of a vector-borne disease outbreak, there may be, in addition to host incidence data, data on the

size of the vector population, as could be obtained through traps in the case of mosquitoes (for mosquito-borne diseases), or field signs, in the case of badgers (for bovine tuberculosis). In stage three, we studied the structural identifiability of model parameters in these "data-rich" scenarios by defining models to have at least two output types. All parameters were treated as unknown, as in stage one. Thus, we were able to compare results obtained at this stage to results at stage one (with single outputs) to assess the influence of additional outputs on parameters' structural identifiability.

For the SIR, SLIR and SLIRQ structures, outputs were incidence and prevalence. For the SLIR/SLI structure, we studied two output combinations. One comprised host incidence and host prevalence, reflecting a scenario in which host infection data are available but vector data are absent, while the other comprised both host and vector data: host incidence, host prevalence, vector incidence and detected vector counts. Thus, five model versions were analysed at this stage.

- **Stage four (multiple outputs, only natural history parameters known)**: Here, we consider the five model versions analysed at stage three, but assuming knowledge of natural history parameters, as in stage two. Thus, we could compare the structural identifiability of parameters at this stage to corresponding results: 1) at stage two, to determine whether additional outputs improved parameters' structural identifiability after some parameters have been assumed known; and 2) at stage three, to determine how structural identifiability of parameters improved with knowledge of natural history parameters, given multiple outputs.

5.3 RESULTS

Structural identifiability results of model parameters assessed at stages one, two, three, and four are presented in Table 5-1, Table 5-2, Table 5-3, and Table 5-4, respectively.

For some models, SIAN was unable to complete global identifiability calculations but provides results for local identifiability. For these model versions, parameters assessed as being s.l.i. by SIAN are referred to in this study as being *at least* s.l.i., given that they may potentially be s.g.i.

Results are now discussed by stage.

Stage one (single outputs, all parameters unknown): See Table 5-1. When all parameters were assumed unknown and single outputs considered, all models except the SLIRQ models are s.n.i. All parameters of the SLIRQ model are s.l.i., irrespective of output type. In the SIR and SLIR models with output as prevalence, the transmission rate β is s.g.i. However, with output as incidence, β becomes s.n.i. The IC for the recovered compartment $R(0)$ is s.n.i. in all SIR and SLIR models studied at stage one, but is at least s.l.i. in both SLIRQ models (i.e., given incidence or prevalence as output). In the SLIR/SIR model with incidence as output, the IC corresponding to the recovered compartment for hosts $R_h(0)$ is at least s.l.i. when output is host incidence but s.n.i. when output is host prevalence. The transmission rate parameter and all ICs corresponding to the vector population are s.n.i. with host prevalence or host prevalence as output, while other parameters associated with the vector population (birth rate μ_v and parameter controlling the length of latent period α_v) are at least s.l.i.

Stage two (single outputs, only natural history parameters known): See Table 5-2. Assuming knowledge of the natural history parameters in the SIR, SLIR and SLIR/SLI models did not lead to an improvement of the structural identifiability of parameters which were s.n.i. at stage one (where all parameters – including natural history parameters – were unknown), irrespective of output type. However, for the SLIRQ models, the structural identifiability of unknown parameters (β and ICs) is seen to improve with the assumption of knowledge of natural history parameters: these parameters are s.g.i. at this stage but were at least s.l.i at stage one.

Stage three (multiple outputs, all parameters unknown): See Table 5-3. When incidence and prevalence data are considered jointly as outputs in the same model, structural identifiability of the SIR, SLIR and SLIRQ models improves considerably compared to stage one. All parameters in these models which were s.n.i. at stage one become s.g.i. For example, β is s.n.i. in the SIR model with incidence only as output; however, with the addition of prevalence data as an output in the model, β becomes s.g.i. Likewise, $R(0)$ is s.n.i. in all SIR and SLIR models with single outputs (either incidence or prevalence; Table 5-1) but becomes s.g.i. when these outputs are considered simultaneously.

For the SLIR/SLI model, all parameters associated with the host population are at least s.l.i. when host incidence and host prevalence data are joint model outputs. However, the ICs and transmission rate parameter associated with the vector population are s.n.i., as in stage one when these outputs were considered separately (Table 5-1).

Stage four (multiple outputs, natural history parameters known): See Table 5-4. Even when natural history parameters are assumed known, the ICs and transmission rate parameter associated with the vector population in the SLIR/SLI model remain s.n.i. with host prevalence and host incidence as joint model outputs. It is only with the addition of vector data (vector incidence and detected vector counts) as outputs that these parameters become s.g.i.

5.4 DISCUSSION

In this work, we have studied the structural identifiability of 26 ODE model versions, each with a unique combination of underlying compartmental structure (SIR, SLIR, SLIRQ or SLIR/SLI) and data type considered as model output (incidence, prevalence or detected vector counts). We gained useful insights on the influence of data types on the structural identifiability of the infectious disease transmission models studied here.

The consideration of multiple data types as outputs generally improved models' structural identifiability. Indeed, when only single outputs were considered (Table 5-1, Table 5-2), all

models except the SLIRQ-structured models were s.n.i. However, when these models were defined to have at least two data types as outputs, all but one model become s.g.i. (Table 5-3, Table 5-4).

The exception – the SLIR/SLI model with outputs as host incidence and host prevalence – had its transmission rate parameter and ICs for the vector population remaining s.n.i. despite having host incidence and host prevalence as model outputs (Table 5-3), and even after all natural history parameters in the model were assumed known (Table 5-4). However, when vector-related data (vector incidence and detected vector counts) were added as outputs in the model, these parameters become s.g.i. (Table 5-3, Table 5-4), suggesting that data on host infection alone (incidence, prevalence or both) are not sufficient to identify these vector-related parameters.

We found it surprising that the other vector-related parameters studied – μ_v, the vector birth rate and α_v, the parameter controlling the length of the latent period – were at least s.l.i. given host incidence or host prevalence (Table 5-1), since we expected vector-related parameters to be non-identifiable in the absence of vector data. We thus checked with other structural identifiability software capable of computing global results – GenSSI 2.0 (Ligon et al., 2018), COMBOS (Meshkat et al., 2014) and DAISY (Bellu et al., 2007) – but none of this were able to complete computations. That these vector-related parameters are identifiable with host data is not yet clear to us and it is a question we continue to explore. We suspect that these parameters are likely not practically identifiable from typically available host incidence data, even if they are structurally identifiable.

Assuming knowledge of the natural history parameters did not seem to improve the structural identifiability of parameters in the majority of single-output models (Table 5-1, Table 5-2), likely because all natural history parameters were at least s.l.i (in those models in which they were treated as unknown parameters; Table 5-1), indicating that they were not in identifiable combinations. Hence, fixing the values of these parameters appeared not to have influenced existing symmetries.

We note that for all SIR and SLIR models with single outputs (incidence or prevalence), the IC corresponding to the recovered compartment $R(0)$ is s.n.i., and its structural identifiability does not improve even when natural history parameters in these models are assumed known (Table 5-1, Table 5-2). Only with the simultaneous analysis of multiple data types as outputs does $R(0)$ become s.g.i. (Table 5-3). It is interesting to observe this "synergy-like" effect: separately, neither incidence nor prevalence is sufficient for the identification of $R(0)$, but considered jointly, these data prove adequate to identify $R(0)$. It is of interest to study further what causes this effect in this case, as resulting insights could potentially aid to *pre-determine* which data types will lead to structural (non-) identifiability. We recommend the development of formal methods to aid such pre-determination: these methods may offer a more efficient paradigm as they may eliminate the need for trials with several data types.

Our results on the IC of the recovered state agree with Massonis et al. (2021) who, in a structural identifiability analysis of several compartmental COVID-19 transmission models, found that the recovered state is "almost never observable". That is, it is not distinguishable from other states given output and input observations, unless measured directly. An important question then arises: what sources of data are useful to inform the IC of the recovered/immune state in scenarios where this state is not directly observed? Expert knowledge or seroprevalence estimates based on representative studies may be helpful in this regard. Where these data are not readily available, this IC has often been set to zero; however, if the true value is different from zero, other parameters need to be interpreted accordingly; the transmission rate, in particular, needs to be understood as an observed transmission (i.e., the transmission estimated from the observed data) rather than a true transmission rate. The distinction may be particularly important when trying to mechanistically interpret the transmission rate as a product of constituent parameters (e.g., contact rate times probability of infection). More broadly, simulation studies and sensitivity analysis may be needed to understand the specific influences of IC values on one's parameter estimates and thus the robustness of one's inferences.

Our study is a relevant contribution to the literature as it explicitly considers ICs and population sizes as unknown in models which have been mostly studied assuming these quantities are known. Data on ICs or population size may not always be available or able to be measured directly, hence the need to study identifiability in such scenarios. Also, as we had complete control over structure-output combinations, we were able to modify model characteristics such that the cause for a change in identifiability results could be precisely identified. In addition, unlike most previous studies, we provide input code for all analyses conducted, to serve as a model to individuals who may be new to structural identifiability analysis. To further facilitate increased adoption of structural identifiability analysis, we chose to use a web-based structural identifiability analysis tool, which accepts simple text-based inputs. This eliminates potential barriers to adoption such as the need for program installation or proficiency in a programming language.

Despite these strengths, some limitations exist. First, when models were complex (i.e., having more than four states, or multiple outputs and several parameters), it was generally challenging for SIAN (and other toolboxes used) to produce complete results. More work is needed on scaling toolboxes to match the increasing complexity of modern epidemic models. Second, it would have been desirable to use multiple toolboxes for all analysis, as that would have facilitated the detection of potentially problematic results; however as stated earlier, SIAN was the only publicly available toolbox – as far as we know – which had the combination of functionalities required for this study: 1) ability to assess both local and global identifiability; 2) ability to assess identifiability of unknown ICs, and 3) possibility to implement without requiring program installation or specialized programming language skills. Work on developing more accessible toolboxes with a range of relevant functionalities is therefore warranted. Third, the selection of compartmental models studied here is limited. Similarly, although the set of data types examined here comprises some of the most commonly measured in disease surveillance, it is not representative of the wide variety of possible data types; for example, we do not consider detected incidence (i.e., incidence allowing for underreporting). Our work is

intended to be primarily illustrative, providing the rationale for assessing structural identifiability and some approaches.

Our work focused on deterministic, compartmental ODE models. It would be desirable to extend our study to cover stochastic models (Browning et al., 2020); models which incorporate population structure (e.g., age-structured or spatial models); time-varying parameters, which have been shown to address structural identifiability issues due to their role in breaking symmetries in the model structure (Massonis et al., 2021a); and additional data types such as the number of recovered individuals and the number of symptomatic individuals in quarantine (for models with a quarantine compartment). A critical caveat exists, however: the available structural identifiability toolboxes only allow for deterministic ODEs, although they could be used to establish proxy identifiability results for stochastic differential equation models (Browning et al., 2020). More research is needed towards developing identifiability analysis tools suited to stochastic models.

So far, we have focused on answering the question: Given a model $\mathcal{M}$, which data types can make model parameters more structurally identifiable? Our discussions have therefore originated from an output (or data type) perspective. Less attention has been paid to the influence of the rest of the model structure on the identifiability of model parameters. The alternative question, therefore, and one that is necessary in data-limited settings, is: Which structural modifications on the system of ODEs $\dot{\boldsymbol{X}}(\boldsymbol{t})$ will improve the structural identifiability of $\mathcal{M}$? Some approaches have been suggested. One approach involves reparameterizing the model with the aim to reduce the number of parameters, concentrating particularly on identifiable combinations (Eisenberg and Hayashi, 2014; Massonis et al., 2021a, 2021b; Meshkat et al., 2014; Wieland et al., 2021). Another approach centers on simplifying model complexity by reducing the number of features/states (Massonis et al., 2021a) and another entails non-dimensionalizing (Kao and Eisenberg, 2018) or scaling some state variables (Brouwer et al., 2018; Eisenberg et al., 2013). These considerations are outside the scope of

the current discussion but are important to the broader goal of developing infectious disease models for useful inference.

It is important to note that although a model may be s.n.i, it may be useful for drawing inferences, if these are limited to the structurally identifiable parameters of the model (Janzén et al., 2016; Massonis et al., 2021a).

In this work, we have demonstrated the influence of data types on structural identifiability of model parameters. A careful consideration of the type of data available for parameter estimation is therefore advised as a relevant initial step in performing inference with infectious disease transmission models.

5.5 CONCLUSIONS

We have studied the structural identifiability of parameters of various compartmental models for infectious disease transmission. We have demonstrated the influence of data types on structural identifiability by considering different data types as model outputs and examining how structural identifiability of unknown parameters, including ICs, varied with varying outputs. The structural identifiability of some parameters varied depending on the type of model output, and single-output models were often not structurally identifiable. In general, the inclusion of additional data types as outputs improved structural identifiability of parameters. Attention ought therefore to be paid to the type(s) of observed data at hand, prior to estimating model parameters, given that data types influence a model's structural identifiability and consequently, the robustness of resulting inferences.

Acknowledgments

The authors thank Gleb Pogudin for helpful support on the implementation of SIAN.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. E.A.D. was supported by a studentship at the Department of Statistics, University of Oxford. A.F.B. was supported by the National Science Foundation (grant DMS1853032) and the National Institutes of Health (grant U01GM110712). C.A.D. was supported by joint Centre funding from the UK Medical Research Council and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union. C.A.D. was funded on grants from the UK National Institute for Health Research (NIHR) [Vaccine Efficacy Evaluation for Priority Emerging Diseases: PR-OD-1017-20007 and HPRU in Emerging and Zoonotic Infections: NIHR200907. The views expressed in this publication are those of the authors and not necessarily those of their funding institutions.

Table 5-1 (stage one): Structural identifiability of parameters and models assuming **all parameters are unknown** and given **single model outputs**: incidence (I) or prevalence (P). For the SLIR/SLI models, outputs corresponding to the host population are annotated with "(h)". Output cells are shaded according to the structural identifiability of the model given that output: a green shade indicates the model is structurally globally identifiable (s.g.i.), a yellow shade indicates the model is structurally locally identifiable (s.l.i.) and a brown shade indicates the model is structurally non-identifiable (s.n.i.).

Model structure	**Output**	**Structural identifiability of parameters**		
		s.g.i.	**s.l.i**	**s.n.i.**
SIR	I	$\gamma, S(0), I(0)$		$\beta, R(0)$
	P	β, γ		$S(0), I(0), R(0)$
SLIR	I	$\alpha, \gamma, S(0), L(0), I(0)$		$\beta, R(0)$
	P	β	α, γ	$S(0), L(0), I(0), R(0)$
SLIRQ	I		$\alpha, \beta, \eta, \gamma, \sigma, S(0), L(0), I(0), R(0), Q(0)$[a]	
	P		$\alpha, \beta, \eta, \gamma, \sigma, S(0), L(0), I(0), R(0), Q(0)$[a]	
SLIR/SLI	I (h)		$\alpha_h, \alpha_v, \beta_h, \gamma_h, \mu_h, \mu_v, S_h(0), L_h(0), I_h(0), R_h(0)$[a]	$S_v(0), L_v(0), I_v(0), \beta_v$
	P (h)		$\alpha_h, \alpha_v, \beta_h, \gamma_h, \mu_h, \mu_v$	$\beta_v, S_h(0), L_h(0), I_h(0), R_h(0),$ $S_v(0), L_v(0), I_v(0)$

[a] Parameters are <u>at least</u> s.l.i. No results were produced for global identifiability: SIAN timed out before global identifiability calculations could be completed.

Table 5-2 (stage two): Structural identifiability of parameters and models assuming **all parameters except natural history parameters are unknown** and given **single model outputs**: incidence (I) or prevalence (P). For the SLIR/SLI models, outputs corresponding to the host population are annotated with "(h)". Output cells are shaded according to the structural identifiability of the model given that output: a green shade indicates the model is structurally globally identifiable (s.g.i.), a yellow shade indicates the model is structurally locally identifiable (s.l.i.) and a brown shade indicates the model is structurally non-identifiable (s.n.i.).

Model structure	**Output**	**Structural identifiability of parameters**		
		s.g.i.	**s.l.i**	**s.n.i.**
SIR	I	$S(0), I(0)$		$\beta, R(0)$
	P	β		$S(0), I(0), R(0)$
SLIR	I	$S(0), L(0), I(0)$		$\beta, R(0)$
	P	β		$S(0), L(0), I(0), R(0)$
SLIRQ	I	$\beta, S(0), L(0), I(0), R(0), Q(0)$		
	P	$\beta, S(0), L(0), I(0), R(0), Q(0)$		
SLIR/SLI	I (h)		$\beta_h, \mu_h, \mu_v, S_h(0), L_h(0), I_h(0), R_h(0)$[a]	$\beta_v, S_v(0), L_v(0), I_v(0)$
	P (h)		β_h, μ_h, μ_v[a]	$\beta_v, S_h(0), L_h(0), I_h(0), R_h(0), S_v(0), L_v(0), I_v(0)$

[a] Parameters are <u>at least</u> s.l.i. No results were produced for global identifiability: SIAN timed out before global identifiability calculations could be completed.

Table 5-3 (stage three): Structural identifiability of parameters and models assuming **all parameters are unknown** and given **multiple model outputs**: outputs are incidence (I), prevalence (P) or detected vector counts (DC). For the SLIR/SLI models, outputs corresponding to the host and vector populations are annotated with "(h)" and "(v)", respectively. Output cells are shaded according to the structural identifiability of the model given that output: a green shade indicates the model is structurally globally identifiable (s.g.i.), a yellow shade indicates the model is structurally locally identifiable (s.l.i.) and a brown shade indicates the model is structurally non-identifiable (s.n.i.).

Model structure	**Output**	**Structural identifiability of parameters**		
		s.g.i.	**s.l.i**	**s.n.i.**
SIR	I, P	$\beta, \gamma, S(0), I(0), R(0)$		
SLIR	I, P	$\alpha, \beta, \gamma, S(0), L(0), I(0), R(0)$		
SLIRQ	I, P	$\alpha, \beta, \eta, \gamma, \sigma, S(0), L(0), I(0), R(0), Q(0)$		
SLIR/SLI	I (h), P (h)		$\alpha_h, \alpha_v, \beta_h, \gamma_h, \mu_h, \mu_v, S_h(0), L_h(0), I_h(0), R_h(0)$[a]	$\beta_v, S_v(0), L_v(0), I_v(0)$
	I (h), P (h), I (v), DC (v)	$\alpha_h, \alpha_v, \beta_h, \beta_v, \gamma_h, \lambda_v, \mu_h, \mu_v, S_h(0), L_h(0), I_h(0), R_h(0), S_v(0), L_v(0), I_v(0)$		

[a] Parameters are <u>at least</u> s.l.i. No results were produced for global identifiability: SIAN timed out before global identifiability calculations could be completed.

Table 5-4 (stage four): Structural identifiability of parameters and models assuming **all parameters except natural history parameters are unknown** and given **multiple model outputs**: outputs are incidence (I), prevalence (P) or detected vector counts (DC). For the SLIR/SLI models, outputs corresponding to the host and vector populations are annotated with "(h)" and "(v)", respectively. Output cells are shaded according to the structural identifiability of the model given that output: a green shade indicates the model is structurally globally identifiable (s.g.i.), a yellow shade indicates the model is structurally locally identifiable (s.l.i.) and a brown shade indicates the model is structurally non-identifiable (s.n.i.).

Model structure	**Output**	**Structural identifiability of parameters**		
		s.g.i.	**s.l.i**	**s.n.i.**
SIR	I, P	$\beta, S(0), I(0), R(0)$		
SLIR	I, P	$\beta, S(0), L(0), I(0), R(0)$		
SLIRQ	I, P	$\beta, S(0), L(0), I(0), R(0), Q(0)$		
SLIR/SLI	I (h), P (h)	$\beta_h, \mu_h, \mu_v, S_h(0), L_h(0), I_h(0), R_h(0)$		$\beta_v, S_v(0), L_v(0), I_v(0)$
	I (h), P (h), I (v), DC (v)	$\beta_h, \beta_v, \lambda_v, \mu_h, \mu_v, S_h(0), L_h(0), I_h(0), R_h(0), S_v(0), L_v(0), I_v(0)$		

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Structural identifiability of compartmental models for infectious disease transmission is influenced by data type
Publication Status	□Published □ Accepted for Publication □Submitted for Publication ☑ Unpublished and unsubmitted work written in a manuscript style
Publication Details	**Dankwa, E.A.**, Brouwer, A.F., Donnelly, C.A. Structural identifiability of compartmental models for infectious disease transmission is influenced by data type.

Student Confirmation

Student Name:	Emmanuelle A. Dankwa		
Contribution to the Paper	Co-conceptualized analysis, performed all analysis, wrote initial version of the manuscript, co-revised the manuscript, documented analysis files, created repository for analysis files		
Signature		Date	15 March 2022

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Professor Christl A. Donnelly		
Supervisor comments Emmanuelle A. Dankwa was the leading author on this manuscript and made the substantial contributions detailed above.		
Signature 	Date	16 March 2022

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 6 Discussion

6.1 RECAP OF CONTRIBUTIONS

In this thesis, we have developed two models of infectious disease transmission – one, a deterministic model for hepatitis A and the other, a stochastic, spatial model for ASF – to demonstrate how mathematical models may be employed for evaluating alternative control interventions in both human and animal contexts. We have also analysed the structural identifiability of 26 ODE model versions, to demonstrate how structural identifiability of parameters in infectious disease transmission models is influenced by the data type(s) considered as model output(s).

From our study on the 2017-2019 HAV outbreak in Louisville (Chapter 3), we gained the following insights (Dankwa et al., 2021).

- People experiencing homelessness were at an increased risk of infection in the initial stages of the outbreak (OR for homeless versus not homeless: 2.62; 95% CI: 1.62–4.25 for PEH/PWUD and 2.39; 95% CI: 1.51–3.78 for all detected cases).
- The critical vaccination threshold needed to achieve herd immunity among PEH/PWUD in Louisville is 76% (95% CI: 72%–80%) assuming a vaccine efficacy of 90%.
- The observed vaccination program was estimated to have averted 30 hospitalizations (95% CI: 19–43), associated with over US$490,000 (95% CI: US$310,000– US$700,000). However, more benefits would have been seen with earlier and faster vaccination rollouts.

From our modelling study on ASFV transmission in a typical European context (Chapter 4), we gained the following insights (Dankwa et al., 2022b).

- Timescale considerations are important when comparing interventions.

- Construction of a fence around the area of high wild boar infection prevalence and increased hunting pressure were effective epidemic control measures.
- Interventions targeting wild boar were more effective at controlling the epidemic than those targeting only pig herds, as
    - the infection risk to pig herds from wild boar was higher than the risk from other herds, and
    - wild boar infection dynamics were the main contributor to epidemic dynamics.

From our structural identifiability analysis of various ODE model versions (Chapter 5), we gained the following insights (Dankwa et al., 2022a).

- Generally, the data type considered as model output influenced structural identifiability of model parameters.
- The inclusion of additional data types as outputs in a model generally led to more structurally identifiable parameters.
- These observations emphasize the importance of carefully considering the data type at hand prior to performing inference with compartmental models for infectious disease transmission.

6.2 STRENGTHS OF THESIS

Each main chapter features a discussion of strengths pertaining to that chapter. Here, we discuss the strengths of the thesis as a whole and highlight additional strengths of individual studies.

One of the major strengths of this thesis is in its diversity. First, we considered both deterministic (Chapter 3 and Chapter 5) and stochastic (Chapter 4) model types, enabling us explore the different ways in which uncertainty is handled in infectious disease models.

Second, we considered both frequentist- and Bayesian-based estimation approaches. The HAV transmission model (Chapter 3) was calibrated using maximum likelihood estimation, a

frequentist approach, while the ASF transmission model (Chapter 4) was calibrated using ABC, a Bayesian approach. We chose to use ABC for the ASF model, since its likelihood function was too complex to be directly accessed, hence precluding MLE and similar approaches. The use of different estimation approaches enabled us explore not only the differences in the principles of these approaches, but also how their implementations differed.

Third, we studied transmission in both humans (Chapter 3) and animals (Chapter 4), allowing us explore intervention evaluation in these different host settings and to learn important differences between settings. One key difference relates to data on population sizes. We found it was less challenging to obtain data on population sizes of animals, compared to humans. This was particularly so, since the human population we studied (PEH/PWUD; Chapter 3) is often stigmatized (Institute of Medicine (US) Committee on Health Care for Homeless, 1988), and hence individuals may be less willing to provide data which portray their inclusion in this population. Indeed, high percentages of missingness (up to 33%; Table 1, Chapter 3) was observed for the variables associated with homelessness and drug use in the surveillance data collected. On the other hand, the data we analysed on animal populations were reflective of European livestock farming and wildlife hunting systems, which are typically characterized by legislations which require up-to-date record-keeping (The Council of the European Union, 1998). A related point, and one pertinent to the implementation of interventions, is that practically, we expect that implementing interventions in animal populations – particularly in a livestock farming context – may be less challenging than in human populations, given the higher degree of autonomy in the latter. Considerable uncertainty may therefore be associated with estimates of intervention effectiveness in human populations, given the potentially high uncertainty in the proportion of a population which may adhere to a given intervention.

The consideration of the savings in hospitalization cost in our comparison of alternative vaccination scenarios in Chapter 3 is a key strength, as it makes our results more relevant to the policy maker, compared to results from similar studies which evaluate vaccine effectiveness only based on epidemiological factors such as the number of cases to be averted

(e.g., Brouwer et al., 2020). As resources are typically in limited supply, there is usually a need, in decision making, to achieve a fine balance between effectiveness (as quantified by epidemiological factors) and cost savings. Estimates like ours which capture both economic and epidemiological factors are therefore more relevant to real-world decision making. Furthermore, in performing our analysis, we worked closely with colleagues at the relevant policy organization in the US – the Centers for Disease Control and Prevention – making our work more readily accessible to support decision-making on the control of HAV infection among PEH/PWUD.

Our ASF model (Chapter 4) was developed in the context of a synthetic epidemic in which detected incidence data were provided at regular intervals and future projections were required under a range of control scenarios. Thus, although ground truth data were available on the full course of the epidemic, we were blind to the part of the data corresponding to the period over which projections were required. However, we were later able to compare our model projections with the ground truth, enabling us to identify the pros and cons of our modelling strategy. Such a setup facilitates iterative model refinement and encourages model designs which are flexible enough to be easily adapted to account for new outbreak data. These represent some key strengths of our ASFV transmission modelling framework.

The structural identifiability analysis in Chapter 5 was limited to familiar model structures and data types commonly used in the infectious disease modelling literature, thus enabling an easier understanding of the insights gained in the study. This feature of our study is an appealing one, as it contributes to addressing a factor which potentially limits the application of structural identifiability analysis in the current infectious disease modelling landscape, namely the scarcity of illustrative examples which apply structural identifiability analysis methods to infectious disease transmission models (as discussed in the “Motivation for this work”; see Section 1.2).

Another major strength of this thesis is its reproducibility. We have provided well-documented code and relevant files to reproduce all results, facilitating the reuse and extension of our ideas.

6.3 LIMITATIONS OF THESIS AND POTENTIAL FUTURE DIRECTIONS

Limitations associated with each study and potential future directions have been detailed in the corresponding main chapter and briefly summarised here. Additional suggestions for future work are also raised.

In our study on HAV transmission (Chapter 3), there was appreciable uncertainty in population size estimates due to underreporting, uncertainty in baseline immunity against HAV due to lack of local seroprevalence estimates, absence of infection risk attributable to MSM status due to high levels of missing data and finally, inadequate data to verify constant case-detection rates, an assumption of the RR analysis (Dankwa et al., 2021).

A potentially interesting extension of the HAV transmission modelling work would be to conduct the model fitting procedure using a Bayesian approach such as MCMC and then compare the resulting parameter estimates (and uncertainty) to the current estimates, which were obtained via maximum likelihood estimation. Zhang and Charlett (2019) performed a similar study on the hepatitis A model originally proposed by Regan et al. (2016) for the 1991-1992 hepatitis A outbreak among MSM in Sydney, Australia. Regan et al. performed parameter estimation using least squares, while Zhang and Charlett did so using MCMC, on the same dataset. The distribution of the fitted model parameters by Zhang and Charlett proved to converge far better than those by Regan et al. Neither study included a formal identifiability analysis; hence, it is not clear the true extent to which the change in estimation approaches affected parameter estimates. It would therefore be instructive to study how parameter estimates change depending on the parameter estimation procedure employed, in the context of a structurally (or practically) identifiable model. This is work in progress.

While the complexity of the ASFV transmission model allowed to evaluate a variety of intervention scenarios at individual-level, it led to a decrease in computational efficiency. This presented a major challenge. Although we took some measures such as cross-language computing and parallelization to increase computational efficiency, the parameter estimation and projection algorithms were still generally slow. We have suggested (Discussion, Chapter 4) that further work is needed on: 1) the design of efficient, yet sufficiently complex models for transmission modelling; and 2) more efficient parameter estimation methods which can be employed on expensive likelihood functions as the one we encountered.

Our structural identifiability analysis (Chapter 5) was only limited to deterministic ODE models, primarily because methods for structural identifiability of stochastic models are either not well-established or properly formalized (Brouwer et al., 2017; Browning et al., 2020). Also, no such method has been implemented in a publicly available software toolbox, as far as we know. Due to these, we were not able to formally assess the structural identifiability of the stochastic ASFV transmission model studied in Chapter 4. There is a clear need for formal methods and software tools for structural identifiability analysis for stochastic models, particularly as such models are becoming widespread in the field.

Finally, a desirable extension of our identifiability analysis would be to demonstrate, for the s.n.i. models, how structural non-identifiability may be addressed. Structural non-identifiability may be addressed by reparameterizing the model, particularly focusing on identifiable combinations, as demonstrated by Brouwer et al. (2018) for a SLIR-type model for polio, Eisenberg et al. (2013) for a SIWR model for cholera (W: compartment representing pathogen in water), and Massonis et al. (2021a) for a SIR model for COVID-19, although a potential challenge is the interpretation of parameter combinations (Massonis et al., 2021a). Other ways to address structural non-identifiability include: 1) fixing the values of those parameters which may be less relevant for prediction or for which prior knowledge is available, thus reducing the number of unknown parameters; 2) modifying the type or number of model outputs, as demonstrated in Chapter 5; and 3) reducing model complexity where possible, by, for

example, decreasing the number of compartments (Balsa-Canto et al., 2010; Chis et al., 2011a; Massonis et al., 2021a, 2021b; Meshkat et al., 2014; Tuncer and Le, 2018). Methods 1) and 2) may however not always be feasible in a disease transmission context due to the typical scarcity of data. For example, in the emergence of a new pathogen, it may be difficult to obtain data on natural history parameters such as the latent period, or on complete incidence rates, in the presence of asymptomatic infections or poor surveillance. In such cases, one faces challenges in determining suitable values to which parameters should be fixed. Additionally, fixing parameter values may result in the loss of important scale-level information on some model components, potentially compromising the model's predictive ability (Wieland et al., 2021).

Chapter 7 Conclusion

In summary:

- **We have developed a dynamic, deterministic model for HAV transmission (Chapter 3)**. We calibrated this model to observed surveillance data from the 2017-2019 HAV outbreak in Louisville, to estimate the critical vaccination threshold required for herd immunity. We also estimated the number of cases averted and associated savings in hospitalization costs for alternative vaccination scenarios among PEH/PWUD. These findings are useful to inform vaccination policies in this and similar populations.
- **We have developed a stochastic, spatiotemporal model for ASFV transmission (Chapter 4)**. Our model allowed the estimation of the impacts of a range of interventions on the epidemic trajectory in wild boar and pig herd populations. An important finding was that it is important to consider the timescale over which interventions are evaluated, as interventions which appear to have limited impact in the short term may have considerable long-term benefits. Our modelling framework could be applied, in real time, to inform understanding of the impacts of alternative interventions on an ASF epidemic, particularly one occurring in a similar wildlife-livestock interface as studied here.
- **We have analysed the structural identifiability of parameters of 26 ODE model versions for infectious disease transmission (Chapter 5)**. We found that generally, the type of data considered as model output influenced parameters' structural identifiability. Models with multiple data types as outputs tended to have more structurally identifiable parameters, compared to models with a single data type as output. Our results emphasize the importance of carefully considering the type of observed data prior to conducting inference with deterministic ODE models of infectious disease transmission.

For all results presented in this thesis, we have provided the relevant code to facilitate reproducibility and the extension of our ideas.

The models, methods and results presented in Chapter 3 and Chapter 4 of this thesis are a valuable addition to the literature on comparing alternative interventions in the contexts of hepatitis A and ASF, respectively. Our studies not only shed light on the efficacy of various interventions but also demonstrate how estimates of intervention efficacy may be obtained using different types of mathematical models and statistical parameter estimation techniques. Additionally, the study presented in Chapter 5 is valuable for demonstrating the influence of the type(s) of model output(s) on structural identifiability of model parameters and serves to increase awareness of the importance of structural identifiability in infectious disease modelling. Such awareness is both timely and critical, given that although structural identifiability is a prerequisite for successful parameter estimation, only little attention has been paid to the subject in the infectious disease modelling literature.

References

Allen, L.J., 2017. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. Infect Model 2, 128–142.

Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M., Rohani, P., 2006. Seasonality and the dynamics of infectious diseases. Ecol Lett 9, 467–484. https://doi.org/10.1111/j.1461-0248.2005.00879.x

Alves da Silva, A., Santos, P., Bento, P., Alves, J., Soares, A., Fonseca, C., Petrucci-Fonseca, F., Monzón, A., Silvério, A., 2004. Reproduction in the wild boar (Sus scrofa Linnaeus, 1758) populations of Portugal. Galemys 16, 53–65.

Anderson, R.M., 1991. Discussion: the Kermack-McKendrick epidemic threshold theorem. Bull Math Biol 53, 1.

Anderson, R.M., May, R.M., 1991. Infectious diseases of humans: dynamics and control. Oxford University Press.

Anderson, R.M., May, R.M., 1982. Directly transmitted infectious diseases: control by vaccination. Science 215, 1053–1060.

Andersson, H., Britton, T., 2000. Stochastic epidemic models and their statistical analysis. Springer Science & Business Media.

Andraud, M., Halasa, T., Boklund, A., Rose, N., 2019. Threat to the French swine industry of African swine fever: surveillance, spread, and control perspectives. Front Vet Sci 6, 248. https://doi.org/10.3389/fvets.2019.00248

Anstett-Collin, F., Denis-Vidal, L., Millérioux, G., 2020. A priori identifiability: An overview on definitions and approaches. Annu. Rev. Control 50, 139–149. https://doi.org/10.1016/j.arcontrol.2020.10.006

ASF Challenge coordinators, 2020. Modelling as a decision-support tool to control the ASF epidemic in Merry Island.

Audoly, S., Bellu, G., D'Angio, L., Saccomani, M.P., Cobelli, C., 2001. Global identifiability of nonlinear models of biological systems. IEEE Trans. Biomed. Eng. 48, 55–65.

Bailey, N.T., 1975. The mathematical theory of infectious diseases and its applications, 2nd ed. Charles Griffin & Company Ltd.

Balsa-Canto, E., Alonso, A.A., Banga, J.R., 2010. An iterative identification procedure for dynamic modeling of biochemical networks. BMC Syst. Biol. 4, 1–18.

Barongo, M.B., Bishop, R.P., Fèvre, E.M., Knobel, D.L., Ssematimba, A., 2016. A mathematical model that simulates control options for African swine fever virus (ASFV). PLoS One 11, e0158658. https://doi.org/10.1371/journal.pone.0158658

Barongo, M.B., Ståhl, K., Bett, B., Bishop, R.P., Fèvre, E.M., Aliro, T., Okoth, E., Masembe, C., Knobel, D., Ssematimba, A., 2015. Estimating the basic reproductive number (R0) for African Swine Fever Virus (ASFV) transmission between pig herds in Uganda. PLoS One 10, e0125842. https://doi.org/10.1371/journal.pone.0125842

Bartlett, M., 1949. Some evolutionary stochastic processes. J R Stat Soc Ser. B Stat Methodol 11, 211–229.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162, 2025–2035.

Begon, M., Bennett, M., Bowers, R.G., French, N.P., Hazel, S., Turner, J., 2002. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. Epidemiol. Infect. 129, 147–153.

Bellman, R., Åström, K.J., 1970. On structural identifiability. Math. Biosci. 7, 329–339.

Bellu, G., Saccomani, M.P., Audoly, S., D'Angiò, L., 2007. DAISY: A new software tool to test global identifiability of biological and physiological systems. Comput. Methods Programs Biomed. 88, 52–61.

Beltrán-Alcrudo, D., Lubroth, J., Depner, K., De La Rocque, S., 2008. African swine fever in the Caucasus. FAO Empres Watch 1, 1–8.

Blome, S., Franzke, K., Beer, M., 2020. African swine fever – A review of current knowledge. Virus Res 287, 198099. https://doi.org/10.1016/j.virusres.2020.198099

Blome, S., Gabriel, C., Beer, M., 2013. Pathogenesis of African swine fever in domestic pigs and European wild boar. Virus Res 173, 122–130. https://doi.org/10.1016/j.virusres.2012.10.026

Blome, S., Gabriel, C., Dietze, K., Breithaupt, A., Beer, M., 2012. High virulence of African swine fever virus caucasus isolate in European wild boars of all ages. Emerg Infect Dis 18, 708. https://doi.org/10.3201/eid1804.111813

Brisson, M., Kim, J.J., Canfell, K., Drolet, M., Gingras, G., Burger, E.A., Martin, D., Simms, K.T., Bénard, É., Boily, M.-C., others, 2020. Impact of HPV vaccination and cervical screening on cervical cancer elimination: a comparative modelling analysis in 78 low-income and lower-middle-income countries. Lancet 395, 575–590.

Britton, T., 2010. Stochastic epidemic models: a survey. Math Biosci 225, 24–35.

Brooks-Pollock, E., Roberts, G.O., Keeling, M.J., 2014. A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature 511, 228–231. https://doi.org/10.1038/nature13529

Brouwer, A.F., Eisenberg, J.N., Pomeroy, C.D., Shulman, L.M., Hindiyeh, M., Manor, Y., Grotto, I., Koopman, J.S., Eisenberg, M.C., 2018. Epidemiology of the silent polio outbreak in Rahat, Israel, based on modeling of environmental surveillance data. Proc. Natl. Acad. Sci. 115, E10625–E10633.

Brouwer, A.F., Meza, R., Eisenberg, M.C., 2017. Parameter estimation for multistage clonal expansion models from cancer incidence data: A practical identifiability analysis. PLoS Comput. Biol. 13, e1005431.

Brouwer, A.F., Zelner, J.L., Eisenberg, M.C., Kimmins, L., Ladisky, M., Collins, J., Eisenberg, J.N., 2020. The impact of vaccination efforts on the spatiotemporal patterns of the hepatitis A outbreak in Michigan, 2016–2018. Epidemiology 31, 628–635.

Browning, A.P., Warne, D.J., Burrage, K., Baker, R.E., Simpson, M.J., 2020. Identifiability analysis for stochastic differential equation models in systems biology. J. R. Soc. Interface 17, 20200652.

Buchberger, B., 1970. An algorithmical criteria for the solvability of algebraic systems of equations. Aequationes Math 4, 374–383.

Buchino, S., Fosl, C., Haynes, L., Kinahan, K., Omer, L., Zero, D., 2019. Solving street homelessness in Louisville, KY: improving the climate of care for individuals experiencing homelessness. University of Louisville, Louisville, Kentucky, US.

Camacho, A., Kucharski, A., Aki-Sawyerr, Y., White, M.A., Flasche, S., Baguelin, M., Pollington, T., Carney, J.R., Glover, R., Smout, E., others, 2015. Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. PLoS Curr 7.

Capaldi, A., Behrend, S., Berman, B., Smith, J., Wright, J., Lloyd, A.L., 2012. Parameter estimation and uncertainty quantification for an epidemic model. Math. Biosci. Eng. 9, 553–576.

Castro, M., Ares, S., Cuesta, J.A., Manrubia, S., 2020. The turning point and end of an expanding epidemic cannot be precisely forecast. Proc. Natl. Acad. Sci. 117, 26190–26196.

Castro, M., de Boer, R.J., 2020. Testing structural identifiability by a simple scaling method. PLOS Comput. Biol. 16, e1008248.

Centers for Disease Control and Prevention, 2021. Epidemiology and prevention of vaccine-preventable diseases, 14th ed. Public Health Foundation, Washington, D.C.

Centers for Disease Control and Prevention, 2020. Widespread outbreaks of hepatitis A across the United States [WWW Document]. URL https://www.cdc.gov/hepatitis/outbreaks/2017March-HepatitisA.htm

Centers for Disease Control and Prevention, 2012. Hepatitis A, acute. 2012 case definition [WWW Document]. URL https://wwwn.cdc.gov/nndss/conditions/hepatitis-a-acute/case-definition/2012/

Chappell, M.J., Godfrey, K.R., Vajda, S., 1990. Global identifiability of the parameters of nonlinear systems with specified inputs: a comparison of methods. Math. Biosci. 102, 41–73.

Chenais, E., Ståhl, K., Guberti, V., Depner, K., 2018. Identification of wild boar-habitat epidemiologic cycle in African swine fever epizootic. Emerg Infect Dis 24, 810–812. https://doi.org/10.3201/eid2404.172127

Chis, O.-T., Banga, J.R., Balsa-Canto, E., 2011a. Structural identifiability of systems biology models: a critical comparison of methods. PloS One 6, e27755.

Chis, O.-T., Banga, J.R., Balsa-Canto, E., 2011b. GenSSI: a software toolbox for structural identifiability analysis of biological models. Bioinformatics 27, 2610–2611.

Chodick, G., Heymann, A., Ashkenazi, S., Kokia, E., Shalev, V., 2008. Long-term trends in hepatitis A incidence following the inclusion of Hepatitis A vaccine in the routine nationwide immunization program. J Viral Hepat 15, 62–65.

Chowell, G., Rothenberg, R., 2018. Spatial infectious disease epidemiology: on the cusp. BMC Med 16, 1–5.

Coalition for the Homeless, 2017. 2017 homeless census [WWW Document]. URL https://www.louhomeless.org/
Cobelli, C., Distefano III, J.J., 1980. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. Am. J. Physiol.-Regul. Integr. Comp. Physiol. 239, R7–R24.
Collier, M.G., Drobeniuc, J., Cuevas-Mota, J., Garfein, R.S., Kamili, S., Teshale, E.H., 2015. Hepatitis A and B among young persons who inject drugs—vaccination, past, and present infection. Vaccine 33, 2808–2812.
Cori, A., Ferguson, N.M., Fraser, C., Cauchemez, S., 2013. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol 178, 1505–1512.
Costard, S., Mur, L., Lubroth, J., Sanchez-Vizcaino, J.M., Pfeiffer, D.U., 2013. Epidemiology of African swine fever virus. Virus Res 173, 191–197. https://doi.org/10.1016/j.virusres.2012.10.030
Cox, D.R., 2006. Principles of statistical inference. Cambridge University Press.
Croft, S., Massei, G., Smith, G.C., Fouracre, D., Aegerter, J.N., 2020. Modelling spatial and temporal patterns of African swine fever in an isolated wild boar population to support decision-making. Front Vet Sci 7, 154. https://doi.org/10.3389/fvets.2020.00154
Dankwa, E.A., Brouwer, A.F., Donnelly, C.A., 2022a. Structural identifiability of compartmental models for infectious disease transmission is influenced by data type.
Dankwa, E.A., Donnelly, C.A., Brouwer, A.F., Zhao, R., Montgomery, M.P., Weng, M.K., Martin, N.K., 2021. Estimating vaccination threshold and impact in the 2017–2019 hepatitis A virus outbreak among persons experiencing homelessness or who use drugs in Louisville, Kentucky, United States. Vaccine 39, 7182–7190. https://doi.org/10.1016/j.vaccine.2021.10.001
Dankwa, E.A., Lambert, S., Hayes, S., Thompson, R.N., Donnelly, C.A., 2022b. Stochastic modelling of African swine fever in wild boar and domestic pigs: epidemic forecasting and comparison of disease management strategies.
Danzetta, M.L., Marenzoni, M.L., Iannetti, S., Tizzani, P., Calistri, P., Feliziani, F., 2020. African swine fever: Lessons to learn from past eradication experiences. A systematic review. Front Vet Sci 7, 296. https://doi.org/10.3389/fvets.2020.00296
de Carvalho Ferreira, H.C., Backer, J.A., Weesendorp, E., Klinkenberg, D., Stegeman, J.A., Loeffen, W.L.A., 2013. Transmission rate of African swine fever virus under experimental conditions. Vet Microbiol 165, 296–304. https://doi.org/10.1016/j.vetmic.2013.03.026
Dellicour, S., Desmecht, D., Paternostre, J., Malengreaux, C., Licoppe, A., Gilbert, M., Linden, A., 2020. Unravelling the dispersal dynamics and ecological drivers of the African swine fever outbreak in Belgium. J Appl Ecol 57, 1619–1629. https://doi.org/10.1111/1365-2664.13649
Demicheli, V., Tiberti, D., 2003. The effectiveness and safety hepatitis A vaccine: a systematic review. Vaccine 21, 2242–2245.
Denis-Vidal, L., Joly-Blanchard, G., 2000. An easy to check criterion for (un) identifiability of uncontrolled systems and its applications. IEEE Trans. Autom. Control 45, 768–771.
Department for Environment, Food and Rural Affairs, 2022. African swine fever in Europe [WWW Document]. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1047016/ASF_in_Eastern_Europe_24.pdf
Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. John Wiley & Sons.
Diekmann, O., Heesterbeek, J.A.P., Metz, J.A., 1990. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. J Math Biol 28, 365–382.
Dietz, K., 1975. Transmission and control of arbovirus diseases. Epidemiology 104, 104–121.
Distefano, J., Cobelli, C., 1980. On parameter and structural identifiability: Nonunique observability/reconstructibility for identifiable systems, other ambiguities, and new definitions. IEEE Trans. Autom. Control 25, 830–833.
Dixon, L.K., Sun, H., Roberts, H., 2019. African swine fever. Antivir. Res 165, 34–41. https://doi.org/10.1016/j.antiviral.2019.02.018
Dong, R., Goodbrake, C., Harrington, H.A., Pogudin, G., 2021. Differential elimination for dynamical models via projections with applications to structural identifiability. arXiv:1302.5484.
Donnelly, C.A., Wei, G., Johnston, W.T., Cox, D., Woodroffe, R., Bourne, F.J., Cheeseman, C., Clifton-Hadley, R.S., Gettinby, G., Gilks, P., others, 2007. Impacts of widespread badger culling on cattle tuberculosis: concluding analyses from a large-scale field trial. Int J Infect Dis 11, 300–308.
Duncan, A.B., Gonzalez, A., Kaltz, O., 2013. Stochastic environmental fluctuations drive epidemiology in experimental host–parasite metapopulations. Proc R Soc Lond B Biol Sci 280, 20131747.

Eisenberg, M.C., Hayashi, M.A., 2014. Determining identifiable parameter combinations using subset profiling. Math. Biosci. 256, 116–126.
Eisenberg, M.C., Robertson, S.L., Tien, J.H., 2013. Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease. J. Theor. Biol. 324, 84–102.
Elam-Evans, L.D., Yankey, D., Singleton, J.A., Sterrett, N., Markowitz, L.E., Williams, C.L., Fredua, B., McNamara, L., Stokley, S., 2020. National, regional, state, and selected local area vaccination coverage among adolescents aged 13–17 years—United States, 2019. Morb. Mortal. Wkly. Rep. 69, 1109.
European Commission, 2020a. Commission Delegated Regulation (EU) 2020/687 of 17 December 2019 supplementing Regulation (EU) 2016/429 of the European Parliament and the Council, as regards rules for the prevention and control of certain listed diseases. J Eur Union L 174, 64–139.
European Commission, 2020b. Strategic approach to the management of African Swine Fever for the EU (Working Document No. SANTE/7113/2015 – Rev 12), Directorate G - Crisis management in food, animals and plants Unit G3 – Official controls and eradication of diseases in animals. Brussels.
European Commission, 2016. Regulation on transmissible animal diseases and amending and repealing certain acts in the area of animal health ('Animal Health Law'). J Eur Union L 84, 1–208.
European Commission, 2002. Proposal for a Council Directive laying down specific provisions for the control of African swine fever and amending Directive 92/119/EEC as regards Teschen disease and African swine fever (COM(2002)51 final). J Eur Union 181 E, 0142—0159.
European Food Safety Authority, 2015. Scientific opinion on African swine fever. EFSA J 13, 4163. https://doi.org/10.2903/j.efsa.2015.4163
European Food Safety Authority, Anette, Boklund, Anette, Bøtner, Theodora, C.V., Klaus, D., Daniel, D., Vittorio, G., Georgina, H., Daniela, K., Annick, L., others, 2020. Epidemiological analyses of African swine fever in the European Union (November 2018 to October 2019). EFSA J 18, e05996.
European Food Safety Authority, Boklund, A., Cay, B., Depner, K., Földi, Z., Guberti, V., Masiulis, M., Miteva, A., More, S., Olsevskis, E., others, 2018. Epidemiological analyses of African swine fever in the European Union (November 2017 until November 2018). EFSA J 16, 5494. https://doi.org/10.2903/j.efsa.2018.5494
European Food Safety Authority, Desmecht, D., Gerbier, G., Gortázar Schmidt, C., Grigaliuniene, V., Helyes, G., Kantere, M., Korytarova, D., Linden, A., Miteva, A., others, 2021. Epidemiological analysis of African swine fever in the European Union (September 2019 to August 2020). EFSA J 19, 6572. https://doi.org/10.2903/j.efsa.2021.6572
Evans, N.D., Chapman, M.J., Chappell, M.J., Godfrey, K.R., 2002. Identifiability of uncontrolled nonlinear rational systems. Automatica 38, 1799–1805. https://doi.org/10.1016/S0005-1098(02)00094-8
Evans, N.D., Chappell, M.J., 2000. Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems. Math. Biosci. 168, 137–159.
Evans, N.D., White, L.J., Chapman, M.J., Godfrey, K.R., Chappell, M.J., 2005. The structural identifiability of the susceptible infected recovered model with seasonal forcing. Math. Biosci. 194, 175–197.
Fan, V.Y., Jamison, D.T., Summers, L.H., 2016. The inclusive cost of pandemic influenza risk (No. Working paper 22137). National Bureau of Economic Research.
Fares, A., 2015. Seasonality of hepatitis: a review update. J. Fam. Med. Prim. Care 4, 96.
Fazel, S., Geddes, J.R., Kushel, M., 2014. The health of homeless people in high-income countries: descriptive epidemiology, health consequences, and clinical and policy recommendations. Lancet 384, 1529–1540.
Fine, P., 2005. The spread of bacterial infection, the problem of herd immunity: Topley WWC, Wilson GS. J Hyg 1923; 21: 243–249. Epidemiol Infect 133, S35–S36.
Fine, P., Eames, K., Heymann, D.L., 2011. "Herd immunity": a rough guide. Clin Infect Dis 52, 911–916.
Finger, F., Funk, S., White, K., Siddiqui, M.R., Edmunds, W.J., Kucharski, A.J., 2019. Real-time analysis of the diphtheria outbreak in forcibly displaced Myanmar nationals in Bangladesh. BMC Med 17, 58.
Fiore, A.E., Wasley, A., Bell, B.P., 2006. Prevention of hepatitis A through active or passive immunization: recommendations of the Advisory Committee on Immunization Practices (ACIP). Morb. Mortal. Wkly. Rep. Recomm. Rep. 55, 1-CE.

Fischer, M., Hühr, J., Blome, S., Conraths, F.J., Probst, C., 2020. Stability of African swine fever virus in carcasses of domestic pigs and wild boar experimentally infected with the ASFV “Estonia 2014” isolate. Viruses 12, 1118. https://doi.org/10.3390/V12101118

Flaxman, S., Mishra, S., Gandy, A., Unwin, H.J.T., Mellan, T.A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J.W., others, 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 584, 257–261.

Food and Agriculture Organization, 2020. FAO Regional Conference for Asia and the Pacific. Report on African Swine Fever (ASF) in Asia and the Pacific [WWW Document]. URL http://www.fao.org/3/nb742en/NB742EN.pdf

Foster, M., Ramachandran, S., Myatt, K., Donovan, D., Bohm, S., Fiedler, J., Barbeau, B., Collins, J., Thoroughman, D., McDonald, E., others, 2018. Hepatitis A virus outbreaks associated with drug use and homelessness—California, Kentucky, Michigan, and Utah, 2017. MMWR Morb Mortal Wkly Rep 67, 1208.

Fruchterman, T.M.J., Reingold, E.M., 1991. Graph drawing by force-directed placement. Softw Pr. Exp 21, 1129–1164. https://doi.org/10.1002/spe.4380211102

Gallardo, C., Soler, A., Nieto, R., Cano, C., Pelayo, V., Sánchez, M.A., Pridotkas, G., Fernandez-Pinero, J., Briones, V., Arias, M., 2017. Experimental infection of domestic pigs with African swine fever virus Lithuania 2014 genotype II field isolate. Transbound Emerg Dis 64, 300–304. https://doi.org/10.1111/tbed.12346

Gervasi, V., Guberti, V., 2021. African swine fever endemic persistence in wild boar populations: Key mechanisms explored through modelling. Transbound Emerg Dis 68, 2812–2825. https://doi.org/10.1111/tbed.14194

Gervasi, V., Marcon, A., Bellini, S., Guberti, V., 2020. Evaluation of the efficiency of active and passive surveillance in the detection of African swine fever in wild boar. Vet Sci 7, 5. https://doi.org/10.3390/vetsci7010005

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81, 2340–2361.

Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys 22, 403–434.

Glikson, M., Galun, E., Oren, R., Tur-Kaspa, R., Shouval, D., 1992. Relapsing hepatitis A. Review of 14 cases and literature survey. Med. Baltim. 71, 14–23.

Global Burden of Disease 2019 Diseases and Injuries Collaborators, 2020. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 396, 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

Gogin, A., Gerasimov, V., Malogolovkin, A., Kolbasov, D., 2013. African swine fever in the North Caucasus region and the Russian Federation in years 2007-2012. Virus Res 173, 198–203. https://doi.org/10.1016/j.virusres.2012.12.007

Greenwood, P.E., Gordillo, L.F., 2009. Stochastic epidemic modeling, in: Mathematical and Statistical Estimation Approaches in Epidemiology. Springer, pp. 31–52.

Guinat, C., Gogin, A., Blome, S., Keil, G., Pollin, R., Pfeiffer, D.U., Dixon, L., 2016a. Transmission routes of African swine fever virus to domestic pigs: Current knowledge and future research directions. Vet Rec. https://doi.org/10.1136/vr.103593

Guinat, C., Gubbins, S., Vergne, T., Gonzales, J.L., Dixon, L., Pfeiffer, D.U., 2016b. Experimental pig-to-pig transmission dynamics for African swine fever virus, Georgia 2007/1 strain. Epidemiol Infect 144, 25–34. https://doi.org/10.1017/S0950268815000862

Guinat, C., Porphyre, T., Gogin, A., Dixon, L., Pfeiffer, D.U., Gubbins, S., 2018. Inferring within-herd transmission parameters for African swine fever virus using mortality data from outbreaks in the Russian Federation. Transbound Emerg Dis 65, e264–e271. https://doi.org/10.1111/tbed.12748

Gulenkin, V.M., Korennoy, F.I., Karaulov, A.K., Dudnikov, S.A., 2011. Cartographical analysis of African swine fever outbreaks in the territory of the Russian Federation and computer modeling of the basic reproduction ratio. Prev Vet Med 102, 167–174. https://doi.org/10.1016/j.prevetmed.2011.07.004

Gunn, R.N., Chappell, M.J., Cunningham, V.J., 1997. Reparameterisation of unidentifiable systems using the Taylor series approach. IFAC Proc. Vol. 30, 247–252.

Halasa, T., Boklund, A., Bøtner, A., Mortensen, S., Kjær, L.J., 2019. Simulation of transmission and persistence of African swine fever in wild boar in Denmark. Prev Vet Med 167, 68–79. https://doi.org/10.1016/j.prevetmed.2019.03.028

Halasa, T., Boklund, A., Bøtner, A., Toft, N., Thulke, H.-H., 2016a. Simulation of spread of African swine fever, including the effects of residues from dead animals. Front Vet Sci 3, 6. https://doi.org/10.3389/fvets.2016.00006
Halasa, T., Bøtner, A., Mortensen, S., Christensen, H., Toft, N., Boklund, A., 2016b. Control of African swine fever epidemics in industrialized swine populations. Vet Microbiol 197, 142–150. https://doi.org/10.1016/j.vetmic.2016.11.023
Halasa, T., Bøtner, A., Mortensen, S., Christensen, H., Toft, N., Boklund, A., 2016c. Simulating the epidemiological and economic effects of an African swine fever epidemic in industrialized swine populations. Vet Microbiol 193, 7–16. https://doi.org/10.1016/j.vetmic.2016.08.004
Halasa, T., Bøtner, A., Mortensen, S., Christensen, H., Wulff, S.B., Boklund, A., 2018. Modeling the effects of duration and size of the control zones on the consequences of a hypothetical African swine fever epidemic in Denmark. Front Vet Sci 5, 49. https://doi.org/10.3389/fvets.2018.00049
Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Krivitsky, P., Morris, M., 2021. ergm: fit, simulate and diagnose exponential-family models for networks [WWW Document]. URL https://cran.r-project.org/package=ergm
Hayes, B.H., Andraud, M., Salazar, L.G., Rose, N., Vergne, T., 2021. Mechanistic modelling of African swine fever: A systematic review. Prev. Vet. Med. 191, 105358.
Heesterbeek, J.A.P., 2002. A brief history of R0 and a recipe for its calculation. Acta Biotheor 50, 189–204.
Heesterbeek, J.A.P., Dietz, K., 1996. The concept of R0 in epidemic theory. Stat Neerl 50, 89–110.
Heffernan, J.M., Smith, R.J., Wahl, L.M., 2005. Perspectives on the basic reproductive ratio. J R Soc Interface 2, 281–293.
Hendrix, K.S., Sturm, L.A., Zimet, G.D., Meslin, E.M., 2016. Ethics and childhood vaccination policy in the United States. Am J Public Health 106, 273–278.
Hengl, S., Kreutz, C., Timmer, J., Maiwald, T., 2007. Data-based identifiability analysis of non-linear dynamical models. Bioinformatics 23, 2612–2618.
Hennessey, K.A., Bangsberg, D.R., Weinbaum, C., Hahn, J.A., 2009. Hepatitis A seroprevalence and risk factors among homeless adults in San Francisco: should homelessness be included in the risk-based strategy for vaccination? Public Health Rep 124, 813–817.
Hermann, R., Krener, A., 1977. Nonlinear controllability and observability. IEEE Trans. Autom. Control 22, 728–740.
Herrero, J., García-Serrano, A., García-González, R., 2008. Reproductive and demographic parameters in two Iberian wild boar Sus scrofa populations. Mammal Res 53, 355–364. https://doi.org/10.1007/BF03195196
Hofmeister, M.G., Yin, S., Aslam, M.V., Teshale, E.H., Spradling, P.R., 2020. Hepatitis A hospitalization costs, United States, 2017. Emerg Infect Dis 26, 1040.
Hong, H., Ovchinnikov, A., Pogudin, G., Yap, C., 2020. Global identifiability of differential models. Commun. Pure Appl. Math. 73, 1831–1879.
Hong, H., Ovchinnikov, A., Pogudin, G., Yap, C., 2019. SIAN: software for structural identifiability analysis of ODE models. Bioinformatics 35, 2873–2874.
Hu, B., Gonzales, J.L., Gubbins, S., 2017. Bayesian inference of epidemiological parameters from transmission experiments. Sci Rep 7, 1–13. https://doi.org/10.1038/s41598-017-17174-8
Huber, C., Finelli, L., Stevens, W., 2018. The economic and social burden of the 2014 Ebola outbreak in West Africa. J Infect Dis 218, S698–S704.
Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, M., 2008. ergm: A package to fit, simulate and diagnose exponential-family models for networks. J Stat Softw 24. https://doi.org/10.18637/jss.v024.i03
Institute of Medicine (US) Committee on Health Care for Homeless, 1988. Homelessness, health, and human needs. Washington (DC): National Academies Press (US).
International Society for Infectious Diseases, 2022. African swine fever - Europe (02): Italy (Piedmont) wild boar. ProMED-Mail Post Arch Number 202201108700778.
International Society for Infectious Diseases, 2021. African swine fever - Europe (15): Germany (BB) 1st report in domestic pig. ProMED-Mail Post Arch Number 202107168522973.
Isham, V., 2005. Stochastic models for epidemics, in: AC Davison, Y.D., Wermuth, N. (Eds.), Celebrating Statistics: Papers in Honour of Sir David Cox on His 80th Birthday. Oxford University Press, pp. 27–54.
Janeau, G., Cargnelutti, B., Cousse, S., Hewison, M., Spitz, F., 1995. Daily movement pattern variations in wild boar (Sus scrofa L.). Ibex JME 3, 98–101.
Janzén, D.L., Bergenholm, L., Jirstrand, M., Parkinson, J., Yates, J., Evans, N.D., Chappell, M.J., 2016. Parameter identifiability of fundamental pharmacodynamic models. Front. Physiol. 7, 590.

Jezierski, W., 1977. Longevity and mortality rate in a population of wild boar. Acta Theriol. (Warsz.) 22, 337–348.
Jo, Y.-S., Gortázar, C., 2021. African swine fever in wild boar: Assessing interventions in South Korea. Transbound Emerg Dis 68, 2878–2889. https://doi.org/10.1111/TBED.14106
Johansson, M.A., Apfeldorf, K.M., Dobson, S., Devita, J., Buczak, A.L., Baugher, B., Moniz, L.J., Bagley, T., Babin, S.M., Guven, E., others, 2019. An open challenge to advance probabilistic forecasting for dengue epidemics. Proc Natl Acad Sci U A 116, 24268–24274.
Jori, F., Vial, L., Penrith, M.L., Pérez-Sánchez, R., Etter, E., Albina, E., Michaud, V., Roger, F., 2013. Review of the sylvatic cycle of African swine fever in sub-Saharan Africa and the Indian ocean. Virus Res 173, 212–227. https://doi.org/10.1016/j.virusres.2012.10.005
Kalman, R.E., 1960. On the general theory of control systems, in: Proceedings First International Conference on Automatic Control, Moscow. pp. 491–502.
Kao, Y.-H., Eisenberg, M.C., 2018. Practical unidentifiability of a simple vector-borne disease model: Implications for parameter estimation and intervention assessment. Epidemics 25, 89–100.
Karlsson, J., Anguelova, M., Jirstrand, M., 2012. An efficient method for structural identifiability analysis of large dynamic systems. IFAC Proc. Vol. 45, 941–946.
Keeling, M.J., Rohani, P., 2011. Modeling infectious diseases in humans and animals. Princeton University Press.
Kentucky Department for Public Health, 2019. Acute Hepatitis A Outbreak Weekly Report, Weeks 51 and 52, 2019, December 15, 2019 – December 28, 2019, Morbidity and Mortality Weekly Report (MMWR). Kentucky Department for Public Health.
Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc R Soc Lond A 115, 700–721.
Keuling, O., Stier, N., Roth, M., 2008. Annual and seasonal space use of different age classes of female wild boar Sus scrofa L. Eur J Wildl Res 54, 403–412. https://doi.org/10.1007/s10344-007-0157-4
Koekpe, R., Sill, D., Akhtar, W., Mitchell, K., Guilfoyle, S., Westergaard, R., Schauer, S.L., Vergeront, J.M., 2019. Hepatitis A and hepatitis B vaccination coverage among persons who inject drugs and have evidence of hepatitis C infection. Public Health Rep 134, 651–659.
Kreutz, C., Raue, A., Timmer, J., 2012. Likelihood based observability analysis and confidence intervals for predictions of dynamic models. BMC Syst. Biol. 6, 1–9.
Lambert, S., Ezanno, P., Garel, M., Gilot-Fromont, E., 2018. Demographic stochasticity drives epidemiological patterns in wildlife with implications for diseases and population management. Sci Rep 8, 16846.
Lange, M., 2015. Alternative control strategies against ASF in wild boar populations. EFSA Support. Publ. 12, 843E.
Lange, M., Guberti, V., Thulke, H.-H., 2018. Understanding ASF spread and emergency control concepts in wild boar populations using individual-based modelling and spatio-temporal surveillance data. EFSA Support Publ 15, 1521E. https://doi.org/10.2903/sp.efsa.2018.en-1521
Lange, M., Thulke, H.-H., 2017. Elucidating transmission parameters of African swine fever through wild boar carcasses by combining spatio-temporal notification data and agent-based modelling. Stoch Env. Res Risk Assess 31, 379–391. https://doi.org/10.1007/s00477-016-1358-8
Lange, M., Thulke, H.-H., 2015. Mobile barriers as emergency measure to control outbreaks of African Swine Fever in wild boar, in: Proc. Annu. Meet. Soc. Vet. Epidemiol. Prev. Med. pp. 122–132.
Lee, H.S., Thakur, K.K., Bui, V.N., Pham, T.L., Bui, A.N., Dao, T.D., Thanh, V.T., Wieland, B., 2021. A stochastic simulation model of African swine fever transmission in domestic pig farms in the Red River Delta region in Vietnam. Transbound Emerg Dis 68, 1384–1391. https://doi.org/10.1111/tbed.13802
Lee, J.-W., McKibbin, W.J., 2004. Estimating the global economic costs of SARS, in: Learning from SARS: Preparing for the next Disease Outbreak: Workshop Summary. National Academies Press Washington, DC, pp. 92–109.
Lemon, S.M., Ott, J.J., Van Damme, P., Shouval, D., 2018. Type A viral hepatitis: a summary and update on the molecular virology, epidemiology, pathogenesis and prevention. J Hepatol 68, 167–184.
Li, J., Blakeley, D., Smith?, R.J., 2011. The failure of R0. Comput Math Methods Med 2011.
Ligon, T.S., Fröhlich, F., Chiş, O.T., Banga, J.R., Balsa-Canto, E., Hasenauer, J., 2018. GenSSI 2.0: multi-experiment structural identifiability analysis of SBML models. Bioinformatics 34, 1421–1423.

Liu, Z., Magal, P., Seydi, O., Webb, G., 2020. A COVID-19 epidemic model with latency period. Infect. Dis. Model. 5, 323–337.
Ljung, L., Glad, T., 1994. On global identifiability for arbitrary model parametrizations. Automatica 30, 265–276.
Lloyd, A.L., 2001. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. Theor Popul Biol 60, 59–71.
Lloyd-Smith, J.O., 2007. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. PloS One 2, e180.
Loi, F., Cappai, S., Laddomada, A., Feliziani, F., Oggiano, A., Franzoni, G., Rolesu, S., Guberti, V., 2020. Mathematical approach to estimating the main epidemiological parameters of African swine fever in wild boar. Vaccines 8, 1–20. https://doi.org/10.3390/vaccines8030521
Lu, P.-J., Hung, M.-C., Srivastav, A., Grohskopf, L.A., Kobayashi, M., Harris, A.M., Dooling, K.L., Markowitz, L.E., Rodriguez-Lainz, A., Williams, W.W., 2021. Surveillance of vaccination coverage among adult populations—United States, 2018. MMWR Surveill. Summ. 70, 1.
Luskin, M.S., Meijaard, E., Surya, S., Sheherazade, Walzer, C., Linkie, M., 2020. African swine fever threatens Southeast Asia's 11 endemic wild pig species. Conserv Lett 14, e12784. https://doi.org/10.1111/conl.12784
Marcon, A., Linden, A., Satran, P., Gervasi, V., Licoppe, A., Guberti, V., 2020. R0 estimation for the African swine fever epidemics in wild boar of Czech Republic and Belgium. Vet Sci 7, 2. https://doi.org/10.3390/vetsci7010002
Maselli, V., Rippa, D., Russo, G., Ligrone, R., Soppelsa, O., D'Aniello, B., Raia, P., Fulgione, D., 2014. Wild boars' social structure in the Mediterranean habitat. Ital J Zool 81, 610–617. https://doi.org/10.1080/11250003.2014.953220
Massonis, G., Banga, J.R., Villaverde, A.F., 2021a. Structural identifiability and observability of compartmental models of the COVID-19 pandemic. Annu. Rev. Control 51, 441–459.
Massonis, G., Banga, J.R., Villaverde, A.F., 2021b. AutoRepar: A method to obtain identifiable and observable reparameterizations of dynamic models with mechanistic insights. Int. J. Robust Nonlinear Control 1–19.
Massonis, G., Villaverde, A.F., 2020. Finding and breaking lie symmetries: implications for structural identifiability and observability in biological modelling. Symmetry 12, 469.
Mazur-Panasiuk, N., Woźniakowski, G., 2020. Natural inactivation of African swine fever virus in tissues: Influence of temperature and environmental conditions on virus survival. Vet Microbiol 242, 108609. https://doi.org/10.1016/j.vetmic.2020.108609
McCabe, R., Kont, M.D., Schmit, N., Whittaker, C., Løchen, A., Walker, P.G., Ghani, A.C., Ferguson, N.M., White, P.J., Donnelly, C.A., Watson, O.J., 2021. Communicating uncertainty in epidemic models. Epidemics 37, 100520.
McGowan, C.J., Biggerstaff, M., Johansson, M., Apfeldorf, K.M., Ben-Nun, M., Brooks, L., Convertino, M., Erraguntla, M., Farrow, D.C., Freeze, J., others, 2019. Collaborative efforts to forecast seasonal influenza in the United States, 2015–2016. Sci Rep 9, 1–13. https://doi.org/10.1038/s41598-018-36361-9
McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, 239–245.
McMahon, B.J., Beller, M., Williams, J., Schloss, M., Tanttila, H., Bulkow, L., 1996. A program to control an outbreak of hepatitis A in Alaska by using an inactivated hepatitis A vaccine. Arch. Pediatr. Adolesc. Med. 150, 733–739.
Medley, G., Anderson, R., Cox, D., Billard, L., 1987. Incubation period of AIDS in patients infected via blood transfusion. Nature 328, 719–721.
Mellor, P.S., Kitching, R.P., Wilkinson, P.J., 1987. Mechanical transmission of capripox virus and African swine fever virus by Stomoxys calcitrans. Res Vet Sci 43, 109–112. https://doi.org/10.1016/s0034-5288(18)30753-7
Merkt, B., Timmer, J., Kaschek, D., 2015. Higher-order Lie symmetries in identifiability and predictability analysis of dynamic models. Phys. Rev. E 92, 012920.
Meshkat, N., Anderson, C., DiStefano III, J.J., 2012. Alternative to Ritt's pseudodivision for finding the input-output equations of multi-output models. Math. Biosci. 239, 117–123.
Meshkat, N., Anderson, C., DiStefano III, J.J., 2011. Finding identifiable parameter combinations in nonlinear ODE models and the rational reparameterization of their input–output equations. Math. Biosci. 233, 19–31.
Meshkat, N., Eisenberg, M., DiStefano III, J.J., 2009. An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner Bases. Math. Biosci. 222, 61–72.

Meshkat, N., Kuo, C.E., DiStefano III, J., 2014. On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and COMBOS: a novel web implementation. PLoS One 9, e110261.
Miao, H., Xia, X., Perelson, A.S., Wu, H., 2011. On identifiability of nonlinear ODE models and applications in viral dynamics. SIAM Rev. 53, 3–39.
Mighell, E., Ward, M.P., 2021. African swine fever spread across Asia, 2018–2019. Transbound Emerg Dis 68, 2722–2732. https://doi.org/10.1111/tbed.14039
Minter, A., Retkute, R., 2019. Approximate Bayesian computation for infectious disease modelling. Epidemics 29, 100368. https://doi.org/10.1016/j.epidem.2019.100368
Muñoz-Tamayo, R., Puillet, L., Daniel, J.-B., Sauvant, D., Martin, O., Taghipoor, M., Blavy, P., 2018. To be or not to be an identifiable model. Is this a relevant question in animal science modelling? Animal 12, 701–712.
Mur, L., Sánchez-Vizcaíno, J.M., Fernández-Carrión, E., Jurado, C., Rolesu, S., Feliziani, F., Laddomada, A., Martínez-López, B., 2018. Understanding African swine fever infection dynamics in Sardinia using a spatially explicit transmission model in domestic pig farms. Transbound Emerg Dis 65, 123–134. https://doi.org/10.1111/tbed.12636
Nelson, N.P., Weng, M.K., Hofmeister, M.G., Moore, K.L., Doshani, M., Kamili, S., Koneru, A., Haber, P., Hagan, L., Romero, J.R., others, 2020. Prevention of hepatitis A virus infection in the United States: Recommendations of the Advisory Committee on Immunization Practices, 2020. MMWR Recomm. Rep. 69, 1–38.
Ngwa, G.A., Shu, W.S., 2000. A mathematical model for endemic malaria with variable human and mosquito populations. Math. Comput. Model. 32, 747–763.
Nielsen, J.P., Larsen, T.S., Halasa, T., Christiansen, L.E., 2017. Estimation of the transmission dynamics of African swine fever virus within a swine house. Epidemiol Infect 145, 2787–2796. https://doi.org/10.1017/S0950268817001613
Nishiura, H., Chowell, G., 2009. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends, in: Mathematical and Statistical Estimation Approaches in Epidemiology. Springer, pp. 103–121.
Nunes, M.A., Balding, D.J., 2010. On optimal selection of summary statistics for approximate Bayesian computation. Stat Appl Genet Mol Biol 9.
Oganesyan, A.S., Petrova, O.N., Korennoy, F.I., Bardina, N.S., Gogin, A.E., Dudnikov, S.A., 2013. African swine fever in the Russian Federation: spatio-temporal analysis and epidemiological overview. Virus Res 173, 204–211. https://doi.org/10.1016/j.virusres.2012.12.009
OIE, 2022. African swine fever (ASF) - Situation Report 7 [WWW Document]. URL https://www.oie.int/en/document/african-swine-fever-asf-situation-report-7/?msclkid=c93fae5fc17311ec96fef3fdf9322146 (accessed 4.21.22).
OIE, 2021. African swine fever [WWW Document]. URL https://www.oie.int/en/disease/african-swine-fever/ (accessed 7.16.21).
Olesen, A.S., Lohse, L., Hansen, M.F., Boklund, A., Halasa, T., Belsham, G.J., Rasmussen, T.B., Bøtner, A., Bødker, R., 2018. Infection of pigs with African swine fever virus via ingestion of stable flies (Stomoxys calcitrans). Transbound Emerg Dis 65, 1152–1157. https://doi.org/10.1111/tbed.12918
O'Neill, X., White, A., Ruiz-Fons, F., Gortázar, C., 2020. Modelling the transmission and persistence of African swine fever in wild boar in contrasting European scenarios. Sci Rep. 2020 101 10, 1–10. https://doi.org/10.1038/s41598-020-62736-y
Ovchinnikov, A., Pillay, A., Pogudin, G., Scanlon, T., 2021. Computing all identifiable functions of parameters for ODE models. Syst. Control Lett. 157, 105030.
Patefield, W.M., 1981. Efficient method of generating random R x C tables with given row and column totals: algorithm AS 159. Appl Stat 30, 91–7.
Penrith, M.L., Bastos, A.D., Etter, E.M.C., Beltrán-Alcrudo, D., 2019. Epidemiology of African swine fever in Africa today: Sylvatic cycle versus socio-economic imperatives. Transbound Emerg Dis 66, 672–686. https://doi.org/10.1111/tbed.13117
Pepin, K.M., Golnar, A., Podgórski, T., 2021. Social structure defines spatial transmission of African swine fever in wild boar. J R Soc Interface 18, 20200761. https://doi.org/10.1098/RSIF.2020.0761
Pepin, K.M., Golnar, A.J., Abdo, Z., Podgórski, T., 2020. Ecological drivers of African swine fever virus persistence in wild boar populations: Insight for control. Ecol Evol 10, 2846–2859. https://doi.org/10.1002/ece3.6100

Picault, S., Vergne, T., Mancini, M., Bareille, S., Ezanno, P., 2021. The African swine fever modelling challenge: objectives, model description and synthetic data generation. bioRxiv 2021.12.20.473417. https://doi.org/10.1101/2021.12.20.473417
Pietschmann, J., Guinat, C., Beer, M., Pronin, V., Tauscher, K., Petrov, A., Keil, G., Blome, S., 2015. Course and transmission characteristics of oral low-dose infection of domestic pigs and European wild boar with a Caucasian African swine fever virus isolate. Arch Virol 160, 1657–1667. https://doi.org/10.1007/s00705-015-2430-2
Podgórski, T., Baś, G., Jędrzejewska, B., Sönnichsen, L., Śnieżko, S., Jędrzejewski, W., Okarma, H., 2013. Spatiotemporal behavioral plasticity of wild boar (Sus scrofa) under contrasting conditions of human pressure: primeval forest and metropolitan area. J. Mammal. 94, 109–119.
Podgórski, T., Lusseau, D., Scandura, M., Sönnichsen, L., Jędrzejewska, B., 2014. Long-lasting, kin-directed female interactions in a spatially structured wild boar social network. PLoS One 9, e99875.
Podgórski, T., Śmietanka, K., 2018. Do wild boar movements drive the spread of African swine fever? Transbound Emerg Dis 65, 1588–1596. https://doi.org/10.1111/tbed.12910
Pohjanpalo, H., 1978. System identifiability based on the power series expansion of the solution. Math. Biosci. 41, 21–33.
Pollock, L.A., Newton, E.J., Koen, E.L., 2021. Predicting high-risk areas for African swine fever spread at the wild-domestic pig interface in Ontario. Prev. Vet. Med. 191, 105341.
Porter, A.T., Oleson, J.J., 2016. A spatial epidemic model for disease spread over a heterogeneous spatial support. Stat Med 35, 721–733.
Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., Feldman, M.W., 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16, 1791–1798.
Probst, C., Globig, A., Knoll, B., Conraths, F.J., Depner, K., 2017. Behaviour of free ranging wild boar towards their dead fellows: Potential implications for the transmission of African swine fever. R Soc Open Sci 4. https://doi.org/10.1098/rsos.170054
Puy, A., Piano, S.L.P., Satelli, A., Levin, S.A., 2021. sensobol: Computation of Variance-Based Sensitivity Indices. arXiv:2101.10103.
R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Raue, A., Karlsson, J., Saccomani, M.P., Jirstrand, M., Timmer, J., 2014. Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics 30, 1440–1448.
Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923–1929.
Raue, A., Kreutz, C., Theis, F.J., Timmer, J., 2013. Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 371, 20110544.
Regan, D., Wood, J., Benevent, C., Ali, H., Smith, L.W., Robertson, P., Ferson, M., Fairley, C., Donovan, B., Law, M., 2016. Estimating the critical immunity threshold for preventing hepatitis A outbreaks in men who have sex with men. Epidemiol Infect 144, 1528–1537.
Reich, N.G., McGowan, C., Yamana, T., Tushar, A., Ray, E., Osthus, D., Kandula, S., Brooks, L.C., Crawford-Crudell, W., Gibson, G.C., others, 2019. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. PLoS Comput Biol 15, e1007486. https://doi.org/10.1371/JOURNAL.PCBI.1007486
Relun, A., Grosbois, V., Alexandrov, T., Sánchez-Vizcaíno, J.M., Waret-Szkuta, A., Molia, S., Charles Etter, E.M., Martínez-López, B., 2017. Prediction of pig trade movements in different European production systems using exponential random graph models. Front Vet Sci 4, 27. https://doi.org/10.3389/fvets.2017.00027
Riley, S., Fraser, C., Donnelly, C.A., Ghani, A.C., Abu-Raddad, L.J., Hedley, A.J., Leung, G.M., Ho, L.-M., Lam, T.-H., Thach, T.Q., others, 2003. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. Science 300, 1961–1966.
Robert, C., 2007. The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media.
Roda, W.C., Varughese, M.B., Han, D., Li, M.Y., 2020. Why is it difficult to accurately predict the COVID-19 epidemic? Infect. Dis. Model. 5, 271–281.

Roosa, K., Chowell, G., 2019. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. Theor. Biol. Med. Model. 16, 1–15.
Rosell, C., Navàs, F., Romero, S., 2012. Reproduction of wild boar in a cropland and coastal wetland area: implications for management. Anim Biodivers Conserv 35, 209–217. https://doi.org/10.32800/abc.2012.35.0209
Rowlands, R.J., Michaud, V., Heath, L., Hutchings, G., Oura, C., Vosloo, W., Dwarka, R., Onashvili, T., Albina, E., Dixon, L.K., 2008. African swine fever virus isolate, Georgia, 2007. Emerg Infect Dis 14, 1870–1874. https://doi.org/10.3201/eid1412.080591
Sabrina, S., Jean-Michel, G., Carole, T., Serge, B., Eric, B., 2009. Pulsed resources and climate-induced variation in the reproductive traits of wild boar under high hunting pressure. J Anim Ecol 78, 1278–1290. https://doi.org/10.1111/j.1365-2656.2009.01579.x
Saccomani, M.P., Audoly, S., D'Angiò, L., 2003. Parameter identifiability of nonlinear systems: the role of initial conditions. Automatica 39, 619–632.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global sensitivity analysis. The primer. John Wiley & Sons.
Sánchez-Cordón, P.J., Nunez, A., Neimanis, A., Wikström-Lassa, E., Montoya, M., Crooke, H., Gavier-Widén, D., 2019. African swine fever: Disease dynamics in wild boar experimentally infected with ASFV isolates belonging to genotype I and II. Viruses 11, 852. https://doi.org/10.3390/v11090852
Sauter-Louis, C., Conraths, F.J., Probst, C., Blohm, U., Schulz, K., Sehl, J., Fischer, M., Forth, J.H., Zani, L., Depner, K., others, 2021a. African swine fever in wild boar in Europe—a review. Viruses 13, 1717. https://doi.org/10.3390/v13091717
Sauter-Louis, C., Forth, J.H., Probst, C., Staubach, C., Hlinak, A., Rudovsky, A., Holland, D., Schlieben, P., Göldner, M., Schatz, J., others, 2021b. Joining the club: First detection of African swine fever in wild boar in Germany. Transbound Emerg Dis 68, 1744–1752. https://doi.org/10.1111/tbed.13890
Scarpino, S.V., Petri, G., 2019. On the predictability of infectious disease outbreaks. Nat. Commun. 10, 1–8.
Schiff, E.R., 1992. Atypical clinical manifestations of hepatitis A. Vaccine 10, S18–S20.
Sedgwick, P., 2014. Relative risks versus odds ratios. BMJ 348.
Sedoglavic, A., 2002. A probabilistic algorithm to test local algebraic observability in polynomial time. J. Symb. Comput. 33, 735–755.
Seymour, R.G., Kypraios, T., O'Neill, P.D., 2022. Bayesian nonparametric inference for heterogeneously mixing infectious disease models. Proc. Natl. Acad. Sci. 119, e2118425119. https://doi.org/10.1073/pnas.2118425119
Shi, R., Li, Y., Wang, C., 2020. Stability analysis and optimal control of a fractional-order model for African swine fever. Virus Res 288, 198111. https://doi.org/10.1016/j.virusres.2020.198111
Sobol', I.M., 1993. Sensitivity estimates for non-linear mathematical models. Math. Model. Comput. Exp. 1, 407–414.
Sobol', I.M., 1967. On the distribution of points in cube and the approximate evaluation of integrals. Comput. Math. Math. Phys. 7, 86–112.
Taylor, R.A., Podgórski, T., Simons, R.R., Ip, S., Gale, P., Kelly, L.A., Snary, E.L., 2021. Predicting spread and effective control measures for African swine fever—Should we blame the boars? Transbound. Emerg. Dis. 68, 397–416.
The Council of the European Union, 1998. Council Directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes. Off. J. 221 0023–0027.
Thulke, H.-H., Lange, M., 2017. Simulation-based investigation of ASF spread and control in wildlife without consideration of human non-compliance to biosecurity. EFSA Support Publ 14, 1312E. https://doi.org/10.2903/sp.efsa.2017.en-1312
Tian, X., von Cramon-Taubadel, S., 2020. Economic consequences of African swine fever. Nat Food 1, 196–197. https://doi.org/10.1038/s43016-020-0061-6
Tomley, F.M., Shirley, M.W., 2009. Livestock infectious diseases and zoonoses. Philos Trans R Soc Lond B Biol Sc 364, 2637–2642.
Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface 6, 187–202.
Topley, W., Wilson, G., 1923. The spread of bacterial infection. The problem of herd-immunity. Epidemiol Infect 21, 243–249.

Tunali, E., Tarn, T.-J., 1987. New results for identifiability of nonlinear systems. IEEE Trans. Autom. Control 32, 146–154.
Tuncer, N., Gulbudak, H., Cannataro, V.L., Martcheva, M., 2016. Structural and practical identifiability issues of immuno-epidemiological vector–host models with application to rift valley fever. Bull. Math. Biol. 78, 1796–1827.
Tuncer, N., Le, T.T., 2018. Structural and practical identifiability analysis of outbreak models. Math. Biosci. 299, 1–18.
US Census Bureau, Population Division, 2019. Subcounty population estimates: April 1, 2010 to July 1, 2018 [WWW Document]. URL https://data.census.gov/
US Department of Health and Human Services, 2019. 2017-2018 national survey on drug use and health: model-based prevalence estimates (50 States and the District of Columbia) [WWW Document]. URL https://datafiles.samhsa.gov/
US Department of Health and Human Services, 2017. Key substance use and mental health indicators in the United States: results from the 2016 national survey on drug use and health (No. HHS Publication No. SMA 17-5044), NSDUH Series H-52. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.
Vajda, S., Godfrey, K.R., Rabitz, H., 1989. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. Math. Biosci. 93, 217–248.
Van Effelterre, T.P., Zink, T.K., Hoet, B.J., Hausdorff, W.P., Rosenthal, P., 2006. A mathematical model of hepatitis A transmission in the United States indicates value of universal childhood immunization. Clin. Infect. Dis. 43, 158–164.
Vergne, T., Andraud, M., Bonnet, S., De Regge, N., Desquesnes, M., Fite, J., Etore, F., Garigliany, M.M., Jori, F., Lempereur, L., others, 2021. Mechanical transmission of African swine fever virus by Stomoxys calcitrans: Insights from a mechanistic model. Transbound Emerg Dis 68, 1541–1549. https://doi.org/10.1111/tbed.13824
Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q., Chowell, G., Simonsen, L., Vespignani, A., others, 2018. The RAPIDD Ebola forecasting challenge: Synthesis and lessons learnt. Epidemics 22, 13–21.
Villaverde, A.F., 2022. Symmetries in Dynamic Models of Biological Systems: Mathematical Foundations and Implications. Symmetry 14, 467.
Villaverde, A.F., 2019. Observability and structural identifiability of nonlinear biological systems. Complexity 2019. https://doi.org/10.1155/2019/8497093
Villaverde, A.F., Barreiro, A., Papachristodoulou, A., 2016. Structural identifiability of dynamic systems biology models. PLoS Comput. Biol. 12, e1005153.
Villaverde, A.F., Raimúndez, E., Hasenauer, J., Banga, J.R., 2019. A comparison of methods for quantifying prediction uncertainty in systems biology. IFAC-Pap. 52, 45–51.
Viltrop, A., Boinas, F., Depner, K., Jori, F., Kolbasov, D., Laddomada, A., Ståhl, K., Chenais, E., 2021. African swine fever epidemiology, surveillance and control, in: Underst. Combat. African Swine Fever A Eur. Perspect. Wageningen Academic Publishers, pp. 229–261. https://doi.org/10.3920/978-90-8686-910-7_9
Vynnycky, E., White, R., 2010. An introduction to infectious disease modelling. Oxford University Press.
Walker, E.R., Pratt, L.A., Schoenborn, C.A., Druss, B.G., 2017. Excess mortality among people who report lifetime use of illegal drugs in the United States: A 20-year follow-up of a nationally representative survey. Drug Alcohol Depend. 171, 31–38.
Walter, E., Lecourtier, Y., 1982. Global approaches to identifiability testing for linear and nonlinear state space models. Math. Comput. Simul. 24, 472–482.
Walter, E., Lecourtier, Y., 1981. Unidentifiable compartmental models: what to do? Math. Biosci. 56, 1–25.
WHO, 2020. Hepatitis A fact sheet [WWW Document]. URL https://www.who.int/news-room/fact-sheets/detail/hepatitis-a
WHO, 2015. Immunization, vaccines and biologicals [WWW Document]. URL https://www.who.int/immunization/diseases/hepatitisA/en/
Wieland, F.-G., Hauber, A.L., Rosenblatt, M., Tönsing, C., Timmer, J., 2021. On structural and practical identifiability. Curr. Opin. Syst. Biol. 25, 60–69.
Wooten, D.A., 2019. Forgotten but not gone: learning from the hepatitis A outbreak and public health response in San Diego. Top Antivir Med 26, 117.
Worby, C.J., Kenyon, C., Lynfield, R., Lipsitch, M., Goldstein, E., 2015. Examining the role of different age groups, and of vaccination during the 2012 Minnesota pertussis outbreak. Sci Rep 5, 13182.

World Animal Health Information System, 2020. World Animal Health Information Database (WAHID) Interface from July 2020 - August 2020 [WWW Document]. URL https://wahis.oie.int/#/events
World Health Organization (WHO), 2020. Global health estimates: Leading causes of death [WWW Document]. URL https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates
World Organisation for Animal Health (OIE), 2020. Situational updates of ASF in Asia and the Pacific [WWW Document]. URL https://rr-asia.oie.int/en/projects/asf/situational-updates-of-asf/
Yee, T.W., 2021. VGAM: Vector Generalized Linear and Additive Models [WWW Document]. URL https://cran.r-project.org/package=VGAM
Yoo, D.S., Kim, Y., Lee, E.S., Lim, J.S., Hong, S.K., Lee, I.S., Jung, C.S., Yoon, H.C., Wee, S.H., Pfeiffer, D.U., Fournié, G., 2021. Transmission dynamics of African swine fever virus, South Korea, 2019. Emerg. Infect. Dis. 27, 1909.
Zhang, X.-S., Charlett, A., 2019. Bayesian modelling of a hepatitis A outbreak in men who have sex with men in Sydney, Australia, 1991/1992. Epidemiol. Infect. 147.
Zimmermann, P., Curtis, N., 2019. Factors that influence the immune response to vaccination. Clin Microbiol Rev 32, e00084-18.