Optimisation within Epidemiological Systems: Exploring the Impact and Mitigation of Disease Outbreaks

Ewan McTaggart

Department of Mathematics and Statistics

University of Strathclyde

A thesis submitted for the degree of

Doctor of Philosophy

2024

Copyright Declaration

Chapters 2, 4 and 5 were a conjoint work with Itamar Megiddo and Adam Kleczkowski. Chapter 3 was a conjoint work with Matthew Baister, Paul McMenemy, Itamar Megiddo and Adam Kleczkowski.

Date: 03/04/2024

Signed: Ewan McTaggart

Acknowledgements

I thank my primary supervisor, Prof. Adam Kleczkowski, and co-supervisor, Dr Itamar Megiddo, for their patience, guidance and unwavering support over the last four years. You have both been superb role models.

My PhD project was made possible by funding from the University of Strathclyde Student Excellence Award, for which I am sincerely thankful. I am grateful to the University of Strathclyde for providing the necessary resources and environment to foster my development. In particular, the access to the ARCHIE-WeSt High-Performance Computer (www.archie-west.ac.uk) based at the University of Strathclyde was invaluable for the computational aspects of my research.

To my family, especially my partner, Iona, thank you for all your love, encouragement and support.

To my colleagues and friends, thank you for the discussions and support that enriched my academic experience.

Abstract

Disease outbreaks pose significant global challenges, impacting public health, ecosystems, and economies. Globalisation, population growth, urbanisation, and climate change have heightened the frequency and impact of diseases, necessitating effective management strategies to control outbreaks. There is a growing need for mathematical models, particularly epi-economic and bio-economic models, to help understand disease dynamics and evaluate interventions. By integrating economics and epidemiology, these models offer a comprehensive understanding of disease spread, considering individual behaviour and ecological factors.

This doctoral thesis explores the use of epidemiological models in understanding disease dynamics, assessing impact, and identifying effective mitigation strategies for different systems. Four paper drafts contribute to this objective. Paper 1 presents a bioeconomic model investigating pests and pathogens' effect on forest harvesting regimes, offering insights for forest managers in designing effective control strategies. Paper 2 develops a compartmental metapopulation model to analyse COVID-19 transmission in care homes, identifying mitigation strategies for vulnerable communities. Paper 3 explores COVID-19-related sickness absence

rates among NHS England staff, guiding resource planning and interventions. Paper 4 introduces a mechanistic compartmental model to estimate COVID-19 sickness absence, evaluating cost-effective interventions and informing workforce management decisions.

Several methodological approaches are employed, including; differential equations (compartmental modelling), autoregressive time series models, multivariate regression, and the net present value analysis.

Contents

Contents **v**

List of Figures **ix**

List of Tables **xiii**

1 Introduction **1**

1.1 Background 1

1.2 The Need for Effective Management Strategies 3

1.3 The Role of Mathematical Models 4

1.4 The Objective of this Doctoral Thesis 8

1.5 Overview of the Chapters 9

1.6 Thesis Outline 13

2 The effect of pests and pathogens on forest harvesting regimes: a bioeconomic model **15**

2.1 Abstract 17

2.2 Introduction 18

2.3 Model framework 24

2.3.1 Forest dynamics — the compartmental timber production model . . . 27
2.3.2 Economic model . . . 30
2.4 Results . . . 33
2.4.1 No disease . . . 36
2.4.2 Disease . . . 39
2.5 Discussion . . . 59
2.6 Conclusion . . . 67

Appendices **69**

2.A Heatmaps showing optimal strategies . . . 69
2.B Calculations for the optimal rotation length when there is no disease 75

Bibliography **77**

3 COVID-19 in Scottish care homes: A metapopulation model of spread among residents and staff **87**

3.1 Abstract . . . 89
3.2 Introduction . . . 90
3.3 Materials and methods . . . 93
3.3.1 Mathematical model . . . 93
3.3.2 Model calibration process . . . 103
3.3.3 Sensitivity analysis . . . 108
3.4 Results . . . 109
3.4.1 Data fit . . . 109

3.4.2 Sensitivity analysis 111
3.5 Discussion 112
3.6 Conclusion 119
3.7 Acknowledgements 120

Appendices **121**
3.A Parameter assumptions and estimates 121
3.B Supplementary data fitting figures 124
3.C Supplementary sensitivity analysis figures 129

Bibliography **133**

4 Sickness Absence Rates in NHS England Staff during the COVID-19 Pandemic **144**
4.1 Abstract 145
4.2 Introduction 147
4.3 Methods 151
4.4 Results 155
4.4.1 Workforce level absence trends 155
4.4.2 COVID-19 related absence 161
4.4.3 Mental health-related absence 172
4.5 Discussion 174

Bibliography **182**

5 A Mechanistic Model of COVID-19 Sickness Absence in NHS England Staff **190**

5.1 Abstract 191

5.2 Introduction 192

5.3 Model framework 197

5.3.1 Dynamic model of absence 199

5.3.2 Economic model: net cost of COVID-19 sickness absence to NHS England 204

5.3.3 Results 207

5.4 Discussion 228

Bibliography **235**

6 Conclusions and Future Work **248**

Bibliography for Introduction and Conclusion **256**

List of Figures

2.1	Timber volume trajectories in the absence of disease.	38
2.2	Disease progress curves under the thinning and clear-felling regime	40
2.3	Impact of transmission rate on optimised management strategies when no revenue comes from infected timber ($\rho = 0$). . .	43
2.4	NPV differences between strategies in a $P - \beta$ parameter space	45
2.5	Impact of transmission rate on optimised management strategies when there is revenue from infected timber ($\rho > 0$). . . .	49
2.6	NPV differences between strategies in a $\rho - \beta$ parameter space.	53
2.A.1	Optimal strategy for the rotation only regime in a P-β parameter space when there is no revenue from infected timber ($\rho = 0$)	69
2.A.2	Optimal strategy for the thinning and rotation regime in a P-β parameter space when there is no revenue from infected timber ($\rho = 0$)	70
2.A.3	Optimal strategy for the rotation only regime in a ρ-β parameter space	71

2.A.4 Optimal strategy for the thinning and rotation regime in a ρ-β parameter space 72

2.A.5 Optimal strategy for the thinning and rotation regime in a P-β parameter space when infected timber does not grow ($\varepsilon = 0$) 73

2.A.6 Optimal strategy for the thinning and rotation regime in a β-ε parameter space 74

2.A.7 Optimal strategy for the thinning and rotation regime in a β-δ parameter space when there is no revenue from infected timber ($\rho = 0$) 75

3.1 Schematics for the compartmental and metapopulation structure. 94

3.2 Surveillance data and fitted model. 110

3.3 Sensitivity of the final resident deaths to the time-share/mixing parameters ($\delta, \varepsilon, \omega^{\gamma}_{high}$). 112

3.B.1 Fitted time-dependent parameters. 125

3.B.2 Quality of fit as a function of homes seeded. 126

3.B.3 Distribution of fitted parameters as a function of homes seeded. 128

3.B.4 Natural log of aggregated sum of squared error (SSE) in a $\omega^{C}_{end} - \omega^{G}_{end}$ parameter space. 129

3.C.1 Sensitivity of the final deaths in each population to perturbations in model parameters. 132

4.1 COVID-19 surveillance data for England 154

4.2 Sickness absence rates for NHS England staff by month . . . 157
4.3 Sickness absence rates for NHS England staff by month and reason . . . 158
4.4 Sickness absence rates for NHS England staff by month and reason: grouped . . . 159
4.5 Sickness absence rates for NHS England staff by month and specific reason . . . 160
4.6 COVID-19 related absence as univariate model of confirmed PCR-positive COVID-19 tests . . . 164
4.7 COVID-19 related absence as univariate model of the ONS estimated COVID-19 incidence . . . 165
4.8 COVID-19 related absence as univariate model of new COVID-19 hospitalisations . . . 166
4.9 COVID-19 related absence as univariate model of the ONS average COVID-19 positivity . . . 168
4.10 COVID-19 related absence as a multivariate regression model of new hospitalisations and ONS estimated COVID-19 incidence . . . 171
4.11 Seasonal ARIMA models of mental health-related sickness absence. . . . 173

5.1 COVID-19 surveillance data for England . . . 202
5.2 Dynamic model of COVID-19 sickness absence fitted to NHS England. . . . 209

5.3 Sensitivity analysis of the modelled COVID-19-related absence rate to the transmission parameters 212
5.4 Sensitivity of the optimal intervention strategy to intervention costs relative to absence cost 217
5.5 Sensitivity of the total cost of COVID-19 absence to changes in hospital intervention effort and cost 223
5.6 Sensitivity of the total cost of COVID-19 absence to changes in hospital intervention effort and cost 225
5.7 Sensitivity of the total cost of COVID-19 absence to changes in intervention efforts 227

List of Tables

2.1 Description of authors' contributions to the manuscript. . . . 16

2.2 Parameter definitions, alongside their base case values. . . . 26

3.1 Description of authors' contributions to the manuscript. . . . 88

3.2 Parameter definitions, alongside their base case values and source. . . . 104

3.3 Parameters used for the data fit and the sets of values simulated over. . . . 108

3.C.1 Parameters involved in the sensitivity analysis. . . . 130

4.1 Description of authors' contributions to the manuscript. . . . 145

4.2 Regression models of the COVID-19 related sickness absence rate. . . . 162

5.1 Description of authors' contributions to the manuscript. . . . 191

5.2 Parameter definitions, alongside their base case values. . . . 198

Chapter 1

Introduction

1.1 Background

Infectious disease outbreaks are global challenges that significantly impact public health, ecosystems, and economies [1–5]. They can cause substantial morbidity and mortality in plants, animals and humans, directly threatening healthcare systems, food security and species survival while disrupting agriculture, productivity, and trade. The estimated global economic burden of eight major infectious diseases (HIV/AIDS, malaria, measles, hepatitis, dengue fever, rabies, tuberculosis and yellow fever) in 2016 was over $6 trillion USD and 162 million person-years were estimated to be lost due to ill health, disability or early death (i.e., DALYs) [3, 6].

The last two decades alone have witnessed many severe infectious disease out-

breaks. The first new and highly transmissible disease to emerge at the turn of the 21st century was Severe Acute Respiratory Syndrome (SARS), first identified in February 2002 and causing 8,096 infections and 774 deaths [7] in humans. Similarly, the H1N1 influenza pandemic caused between 123,000 and 203,000 deaths worldwide in 2009 [8]. The 2014-2016 Ebola outbreak in West Africa resulted in 15,261 confirmed cases and 11,325 deaths [9]. Meanwhile, coffee leaf rust caused substantial yield losses in coffee production [10], with estimated annual costs of $1 - 2 billion USD to the coffee industry [11]. Tree pandemics, such as Dutch elm disease, cypress canker, and pine wilt disease, have had devastating impacts on timber and biodiversity conservation [12, 13]. The COVID-19 pandemic that emerged in late 2019 has been an unprecedented global health and economic crisis, causing over 750 million confirmed cases and almost 7 million deaths [14]. Over the last two decades, H5N1 avian influenza has caused panzootic outbreaks in poultry, leading to the culling of hundreds of millions of birds. This virus can cross species barriers, with fatal outcomes in mammals recorded as recently as 2023 [15, 16].

The emergence of infectious diseases has increased significantly over the past century [2, 17, 18], despite advances to understand, prevent and treat them [19, 20]. Globalisation, population growth, climate change, and urbanisation or land use changes are some of the interconnected factors amplifying the risks associated with disease outbreaks [1, 5, 21, 22]. The movement of people, animals, and goods (i.e., hosts) across borders enabled by globalisation has opened pathways

for the rapid and broad spread of diseases. Population growth and urbanisation have created densely populated cities, conducive environments for disease transmission. The accompanied agricultural demand to feed a growing population increases the probability of interspecies transmission by expanding the contacts between wildlife, crops, livestock, and people [23, 24]. Furthermore, climate change alters ecological systems alongside animal or plant metabolism and physiology, increasing their susceptibility to, and the persistence of, pests, parasites, and vector-borne diseases [5, 25–27].

1.2 The Need for Effective Management Strategies

Given the significant challenges of disease outbreaks, developing effective management strategies is of paramount importance. In particular, the COVID-19 pandemic's impact on public health systems, economies, and societies is a stark reminder of the importance of developing effective and proactive prevention, preparedness, and mitigation strategies to respond to disease outbreaks. Furthermore, given the multifaceted mechanisms and impact of disease outbreaks, these strategies should consider the complex interactions between human behaviour, ecological or environmental factors, and economic considerations [22, 28, 29].

1.3 The Role of Mathematical Models

Mathematical models play a crucial role in this process as powerful tools for understanding disease dynamics and evaluating control strategies in epidemiology [30, 31]. These models can simulate different scenarios and analyse factors influencing disease transmission, e.g., heterogeneity in population demographics [32] or contact structures [33]. Additionally, they provide a framework for quantifying uncertainty and variability inherent in epidemiological processes. By incorporating stochasticity and parameter uncertainty into their structures, models can generate probabilistic forecasts and assess decision-making under uncertainty, as well as the robustness of control measures under various conditions [34, 35]. This section will illustrate models' utility in predicting potential disease impact, helping understand the underlying mechanisms of disease transmission, and providing insights for developing targeted prevention, preparedness, and mitigation strategies.

The Foot-and-Mouth Disease outbreak at the start of the 21st century was among the first times that models were used during an outbreak to support veterinary authorities and policymakers in the decision-making process. Spatiotemporal metapopulation models were employed to highlight the importance of rapid and focused implementation of livestock ring culling strategies in reducing outbreak size and duration [36] in the UK, and suggested that culling would be more effective than vaccination in mitigating spread [37]. With concerns of smallpox bioattacks in the UK in the early 2000s, models demonstrated that outbreaks could

be contained with sufficient levels of contact tracing and vaccination [38, 39]. In response to the 2009 H1N1, models illustrated the effectiveness of targeting vaccinations to children and those at risk of severe complications [40]. The ongoing COVID-19 pandemic has further highlighted the importance of epidemiological models in understanding and managing disease outbreaks. Modelling shaped the early pandemic response in the UK [41], notably the decision to implement a strict nationwide lockdown [42], followed by restrictions on social contacts (household bubbles). Furthermore, they demonstrated the benefits of PPE and regular testing in reducing nosocomial transmission and absenteeism in hospitals [43].

Compartmental models, in particular, are powerful tools that can capture the complexities of host-pathogen-environment interactions [44]. These models are used to understand and predict the spread of infectious diseases within populations by dividing the population into different compartments based on their epidemiological status (e.g., susceptible, infected, and recovered or removed individuals) [45, 46]. Importantly, compartmental models can incorporate behavioural factors to study their impact on disease transmission, making them ideal candidates for epi/bio-economic modelling [47] (discussed below). For example, individuals may change their behaviour in response to influenza or COVID-19 outbreaks by practising social distancing or wearing masks. These behavioural changes can be modelled by introducing additional compartments or modifying the transmission rates in the existing compartments [48]. However, these models have limits in the depth of spatial, environmental, and social dimensions of

epidemiology they can capture [46]. Particularly in small stochastic populations, agent or individual-based models are argued to be more suitable [49].

In recent decades, there has been a growing recognition of the utility of integrating epidemiological and economic approaches in disease modelling [47, 48, 50]. This led to the development of epi-economic and bio-economic models, which merge economics and epidemiology to provide a more comprehensive understanding of the feedback loops between disease dynamics and control strategies. Epi-economic models explicitly analyse individual behavioural choices in response to disease risk [50], while bio-economic models incorporate ecological factors into the analysis [48, 51]. Bio-economic approaches typically focus on land-use decision-making, with an emphasis on ecosystem services, for example, through agriculture, livestock, and forestry [52]

The COVID-19 pandemic has also emphasised this need for integration [53–55]. As Basurto et al. [53] argued, while well-established epidemiological models can address the health-related aspects of the pandemic, there is a need for rigorous approaches that consider the economic effects of containment measures. By simulating and comparing the effectiveness of various control measures, these types of models can help (i) identify the most efficient and cost-effective interventions to control disease transmission, (ii) optimise resource allocation, and (iii) estimate the economic costs of disease outbreaks and strategies [50].

The study of epi-economic models in the context of epidemics began with the spread of AIDS. Philipson and Posner (1993) [56] were among the first to incorporate behavioural elements into models of epidemics. They developed an SI model (susceptible-infected) that considered the risk attitudes of forward-looking rational agents. They demonstrated how increased prevalence or subsidies (e.g., education, condoms, clean needles) could impact health and economic outcomes, highlighting the danger of fatalism. Furthermore, researchers have used epi-economic models to analyse the cost-effectiveness of different control measures, such as treatment campaigns and preventative measures or policies, in reducing disease transmission and impact. For example, with influenza vaccines [57], COVID-19 stay at home orders (and subsidies) [57], measles eradication [58], and chronic livestock diseases [47].

Optimisation in mathematical epidemiology, particularly in epi-economic and bio-economic models, plays a vital role in understanding the relationship between behaviour and disease dynamics. Optimisation guides the identification of effective interventions, timing, and resource allocation to enhance social and economic outcomes [59–61]. As well as recommendations for optimal management strategies, the method provides insights into the relationship and feedback between behaviour, disease dynamics and economic factors. For example, optimal control models have been used in forestry economics to inform modifications to planting and harvesting strategies in the face of pest and pathogen outbreaks and storms. [62–64]. Petucco et al. [64] use a net present value analysis to demon-

strate how frequent harvesting of timber forests can hedge against the risks posed by storms and Pine Processionary Moth outbreaks, and how these risks have different implications for how long stands should be left to grow before clearing and replanting. Similarly, Macpherson et al. [63] highlight the benefits of tree species diversification in production forests. They show how the sensitivity of the optimal planting composition thresholds changes with variations in disease risk and damage.

In conclusion, the increasing risks associated with disease outbreaks highlight the need for effective management strategies. Mathematical models, including epi-economic and bio-economic models, can provide valuable predictions and insights into disease dynamics, optimal management strategies, and interventions' economic and ecological consequences. When models can explain these host-pathogen mechanisms well, they become increasingly attractive to policymakers [65]. Their insights can assist stakeholders and decision-makers in developing risk assessments or preparedness plans while providing an evidence-based framework for evaluating policies and interventions before implementation.

1.4 The Objective of this Doctoral Thesis

The main objective of this doctoral thesis is to explore the use of epidemiological models in understanding disease dynamics, assessing their impact on different systems, and developing effective mitigation strategies. Specifically, this thesis

comprises four paper drafts at different stages of the publication process, each focusing on a different aspect of disease outbreaks and their management. We investigate the effects of pests and pathogens on forest harvesting regimes, the transmission dynamics of COVID-19 in care homes, and the sickness absence rates among NHS England staff during the COVID-19 pandemic. The findings from these papers contribute to the field of epidemiology by providing evidence-based insights for effective disease control and management.

1.5 Overview of the Chapters

Chapter 2 (Paper 1): The effect of pests and pathogens on forest harvesting regimes: a bioeconomic model

In the first paper, we use a bioeconomic model to investigate the impact of pests and pathogens on forest harvesting regimes. Our study addresses the significant threat of pests and diseases to forest ecosystems and timber production [66]. The developed bioeconomic model integrates ecological, epidemiological, and economic factors and provides insights for forest managers to make informed decisions regarding control and mitigation strategies. Our analysis of optimal harvesting strategies and their interactions with disease uncovers trade-offs between maximising harvest yields and mitigating pest and pathogen spread. Specifically, we highlight the key role of thinning as an effective intervention within harvesting strategies, showing how it can significantly improve investments' net present

value. Our findings contribute to designing effective forest management strategies in the presence of disease, with practical implications for decision-makers seeking to minimise economic losses.

Chapter 3 (Paper 2): COVID-19 in Scottish care homes: A metapopulation model of spread among residents and staff

In the second paper, we use a compartmental metapopulation model to investigate the transmission dynamics of COVID-19 within care homes in Scotland. The vulnerability of care homes was highlighted by the COVID-19 pandemic, with a disproportionate number of deaths occurring within these facilities [67]. Understanding disease spread within these communities is crucial for minimising casualties and alleviating the strain on healthcare systems. Our study analyses the intra-subpopulation mixing patterns between care home residents, staff, and the general population across a network of care homes. Through this analysis, we provide insights into the factors influencing disease transmission and identify potential mitigation strategies for future outbreaks. We demonstrate the limitations of single control measures, such as staff sharing restrictions or community visitation bans, and highlight the effectiveness of more comprehensive measures, like staff living at the care homes where they work. These findings contribute to developing targeted interventions and management strategies to mitigate the spread of COVID-19 in care homes. Furthermore, they highlight the importance of robust planning and support for care homes and their staff during pandemics.

Chapter 4 (Paper 3): Sickness Absence Rates in NHS England Staff during the COVID-19 Pandemic

In the third paper, we investigate the impact of the COVID-19 pandemic on sickness absence rates among NHS England staff. The pandemic pressured NHS England significantly, leading to increased staff illness-related absences amid surging treatment demands [68, 69]. Our research aims to provide insights into the pandemic's effect on healthcare workers and identify factors influencing sickness absence rates. Our analysis of sickness absence trends highlights substantial increases coinciding with the arrival of COVID-19 in England, with COVID-19-related or mental health absences being the major drivers of these dynamics. We estimate sickness absences in these two categories using multivariate regression and time series models. We provide information for targeted interventions and policies to reduce staff sickness absence, improve workforce health, and enhance productivity. Additionally, our findings contribute to understanding the pandemic's impact on healthcare systems, guiding workforce management and public health preparedness while identifying key areas for future research.

Chapter 5 (Paper 4): A Mechanistic Model of COVID-19 Sickness Absence in NHS England Staff and Assessment of Interventions

In the fourth and final paper, we develop a mechanistic compartmental model that uses publicly available COVID-19 surveillance data to estimate COVID-19 sickness absence among the NHS England workforce. Understanding the relationship between COVID-19 outbreaks and sickness absence is crucial for effective workforce management, ensuring high-quality patient care, and alleviating strain on healthcare staff [70]. Our study integrates an economic model with an epidemiological system, creating a comprehensive framework for evaluating the cost-effectiveness of interventions to control sickness absence. We identify strategies to reduce disease-related absenteeism while considering resource constraints and epidemiological dynamics by analysing staff-related transmission dynamics and applying optimisation techniques. Our model emphasises the importance of intervention costs compared to the savings from preventing staff absence and highlights the most cost-efficient measures to mitigate transmission rates. These findings contribute to evidence-based decision-making in public health and offer practical guidelines for managing future disease outbreaks.

Summary of the Overall Contribution

The findings of each paper contribute to addressing the broader research objectives of this thesis - exploring the use of epidemiological models in understanding disease dynamics, assessing their impact on different systems, and developing effective mitigation strategies. By integrating ecological, economic, and epidemiological factors, our research sheds light on the complexities of disease management in different systems while helping bridge the gap between health-related aspects of disease outbreaks and their economic implications. From forest ecosystems to care homes and healthcare settings, our models provide valuable insights for decision-makers and stakeholders seeking to optimise disease control strategies and minimise economic losses.

1.6 Thesis Outline

The subsequent chapters will delve into the specific details of each research paper draft. Each paper begins with an introduction that outlines the problem being addressed and then gives an overview of the methods used to investigate the problem. The subsequent section presents the results of the analysis and is followed by a discussion of the findings. Each paper concludes with a summary of the key contributions, limitations, and future research directions. The last item in each chapter is the corresponding bibliography and appendices.

Chapter 2 (Paper 1): The effect of pests and pathogens on forest harvesting

regimes: a bioeconomic model

Chapter 3 (Paper 2): COVID-19 in Scottish care homes: A metapopulation model of spread among residents and staff

Chapter 4 (Paper 3): Sickness Absence Rates in NHS England Staff during the COVID-19 Pandemic

Chapter 5 (Paper 4): A Mechanistic Model of COVID-19 Sickness Absence in NHS England Staff and Assessment of Interventions

In the final chapter, we give our concluding remarks.

Publications arising from this Thesis

Paper 1 was published in the Ecological Economics journal (Elsevier) in April 2023 [71]. Plos Computational Biology and Plos One rejected previous versions of Paper 2; the version shown has recently been submitted to the Epidemics journal (Elsevier). Paper 3 will be submitted shortly to BMC Public Health, and Paper 4 is in the later stages of preparation.

Chapter 2

The effect of pests and pathogens on forest harvesting regimes: a bioeconomic model

This chapter is a corrected version of a paper produced conjointly with Itamar Megiddo (IM) and Adam Kleczkowski (AK). The original version was published in the Ecological Economics journal in April 2023 [1]. We developed a bioeconomic model to explore the impact of pests and pathogens on forest harvesting regimes. Each author's contributions are outlined briefly below and shown in Table 2.1.

AK and IM played supervisory roles and provided feedback on the manuscript. The PhD author (EM) worked on; the literature review, writing and editing the

manuscript, developing the model equations, developing the source code to simulate the model, and on formal model analysis and investigation (parameterisation, sensitivity analysis, and data visualisation).

Table 2.1: Description of authors' contributions to the manuscript.

Contributor Role	**Role Definition**	**Name**
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	AK
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	EM
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	EM
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	AK
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	EM
Methodology	Development or design of methodology; creation of models	AK, EM, IM
Project Administration	Management and coordination responsibility for the research activity planning and execution.	AK, IM, EM
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	AK, EM
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	EM
Supervision	Oversight and leadership responsibility for the research activity planning and, execution including mentorship external to the core team.	AK, IM
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.	EM
Visualisation	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	EM
Writing- Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	EM
Writing- Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.	AK, IM, EM

2.1 Abstract

Pests and diseases are an existential threat to trees in forests and woodlands. There is, therefore, a pressing need to use ecological and bioeconomic models to inform forest managers on control and mitigation strategies. For example, the incidence of Dothistroma needle blight in the UK has increased rapidly since the 1990s, and it is a significant threat to the productivity of commercial forestry. Climatic changes are expected to exacerbate this problem further. Control of the disease in the UK primarily focuses on good stand management through pre-commercial thinning; similar practices are widely used in commercial forests worldwide. Forest managers would benefit from evidence on the effectiveness of this precautionary strategy (in comparison to its alternatives) to reduce disease impacts and increase the value extracted from timber. In this paper, we develop a bioeconomic model to determine the economically optimal harvesting regime – in terms of thinning and rotation – of an even-aged plantation under the risk of an invading pest. We extend a Schaefer–Faustmann model to include a compartmental epidemiological system that governs timber growth and disease spread. We analyse a set of management regimes, including the timing of the final clear-felling of the forest and the timing and level of earlier thinning. Thus, in our approach, forest managers decide whether and when to thin and must balance (i) harvesting before infection destroys the timber's value and (ii) exploiting the forest's density-dependent growth.

We use a sensitivity analysis with respect to the disease spread and impact on the tree dynamics to demonstrate that, in the presence of disease, thinning can significantly improve the net present value of the plantation if applied correctly. Furthermore, if thinning reduces the transmission rate significantly, the priority is to protect the final harvest, and rotations extend while the thinning time shortens. Our study provides a framework to help design appropriate forest management strategies in the presence of disease.

2.2 Introduction

Outbreaks of invasive pests and pathogens disrupt forest services and cause significant ecological, economic, and social losses [2, 3]. The incidence of such outbreaks worldwide has grown with the globalisation of trade and climatic changes [4] — a trend expected to continue [5]. In the last few decades, the arrival and establishment of Chalara ash dieback (*Hymenoscyphus fraxineus*), European spruce bark beetle (*Ips typographus*) and Dothistroma needle blight (*Dothistroma septosporum*) has stressed UK forests [6–8]. These pests could devastate the flow of woodland ecosystem services, which contribute £3.3 billion to the UK economy annually [9]. In this paper, we consider a commercial forest where the manager is interested in minimising economic losses due to disease in terms of the timber benefit. We investigate how the optimal harvesting strategy – in terms of thinning and rotation – changes under the risk of an invading pest.

For some diseases, the focus is on arrival prevention in the UK, and for others, it is on early detection, eradication or containment to prevent economic impacts and loss of valued habitats and landscapes [6]. In many cases, eradication of pests or pathogens after their detection, through measures such as widespread clear-felling or chemical treatment, is not an option [2, 10, 11]. Management options revolve around mitigation to reduce disease impacts, and secondary infection pathways [2]. Silvicultural practices such as diseased tree removal, changing species after rotation, pruning, coppicing, and thinning are deployed [2]. In this paper, we consider thinning and clear-felling as the two options for the plantation, after which the land lays bare.

In established stands, thinning is the primary method of influencing the growth and development of trees [12]. Leaving forests unthinned would lead to the premature mortality of trees due to competition for light and other resources [13]. Whereas, thinning frees up growing space and reduces this competition among closely spaced stems, accelerating the growth of the remaining trees [14]. This allows forest managers to achieve target merchantable products (in terms of tree diameters and stand uniformity) and increases the value of timber extracted at rotation [15].

Thinning a stand can manipulate environmental conditions to diminish secondary infection pathways [2], and can also increase the resilience of the remaining trees

to diseases (reducing the stress on trees, leading to healthier and stronger trees less susceptible to diseases) [16]. In the UK, control of Dothistroma Needle Blight focuses on good stand management and "learning to live with the disease" [17, 18]. This includes thinning to improve airflow and make conditions less conducive to fungus development [19]. Thinning increases the distance between trees (reducing tree density) and reduces the effectiveness of rain-splashed spores [17]. After a chemical treatment in response to spruce budworm outbreak, thinning may be used to increase tree and stand resistance to the pest through increased foliage production [20]. Similarly, thinning acts as a preventative tactic for bark beetles by improving tree vigour [21].

Ecological and bioeconomic models that explore thinning as a forest management strategy are largely based on the classic Faustmann model (Martin Faustmann, 1849) [22]. The Faustmann model is a net present value (NPV) framework that determines the optimal rotation age for an even-aged stand under the assumption of periodical regeneration and rotation. One approach that explores a thinning strategy operates at the tree level and determines individual tree harvesting times within a forest [23]. In another approach, Clark and De Pree [24] let the growth of the total timber stock in the Faustmann model react to annual harvests, becoming a Schaefer–Faustmann model. They analysed this model with optimal control to show that optimal annual thinning follows a bang–bang strategy [24]. Halbritter and Deegen [25] build on this and perform a deterministic analysis of the combined optimal planting density, thinning and rotation for an even-aged

stand. Similarly,Tahvonen [26] adapts [24] to investigate the conditions under which continuous cover production forestry is optimal in place of rotation forestry.

The Faustmann model has also been extended to consider disease risk. Reed [27] first explored the impact of natural disturbances on optimal forest management (rotation length and net present value). He adapted the infinite rotation Faustmann formula to include the risk of catastrophic loss from wildfires with a homogeneous Poisson distribution. However, as Macpherson et al. [28–30] argue, tree diseases exhibit key differences compared to other natural disturbances like storms or wildfires, and models should account for these. In particular, disease progresses at a slower speed. While it can progress at variable time scales, the likely units are years. Further, the symptoms (cryptic infection) can result in the disease remaining undetected for long periods. Lastly, the long-term persistence of many pathogens following their invasion is often irreversible. Given these differences, Macpherson et al. [28–30] link a single rotation NPV framework/Faustmann model to a deterministic susceptible-infected compartmental model, to estimate pest/pathogen density impact on impact timber revenue. They investigate the effect of disease on the optimal rotation length (with 29 and without 28 the inclusion of non-timber benefits attributed to the forest) and the optimal mix of planted species [30]. An et al. [31] recently adapted the Macpherson framework and used a structural damage function [a generalised linear model with probit link function, from 32] to capture disease dynamics and their impact. They find the optimal rotation age when pest damage depends on

their density and climatic variables, and predict Pine Wilt disease's damage rate and economic impact under different disease conditions, climate scenarios, and interventions.

Models that build on the Faustmann Model to integrate disease (or natural disasters) and thinning capture some of the interactions between tree growth, thinning, and outbreak progression/risk. Halbritter et al. [33] accompany a Schaefer–Faustmann model [25] with an age and density-dependent hazard function to represent the arrival of catastrophic natural risks (fire, storms, pests). Using optimal control theory, they analytically explore the optimal annual thinning regime and rotation length under different risks, interest rates, and timber prices. Petucco and Andrés-Domenech [34] extend the Faustmann model by adding a fixed thinning regime and considering the combined impact of storms and a defoliator pest (Pine Processionary Moth). Their timber growth model accounts for the number of trees, their heights, and the basal area. These factors are impacted in unique ways by windstorms (modelled as random Poisson events) and the annual pest density (given by a sinusoidal statistical model). Thinning has no effect on the frequency or severity of any of the risks. Although Petucco et al. do not optimise thinning, they investigate how these disturbances change the optimal rotation length and the land's expectation value. Staupendahl and Mühring [35] model an even-aged spruce stand with fixed pre-commercial thinning intervals, and estimate the optimal rotation length and annuity under risk (under the assumption of infinite rotations). They use an exogenous age-dependent survival function

(Weibull distribution) to model the area of forest remaining after damage from "natural risks" (storms or pests) and perform a sensitivity analysis to different risk distributions.

In this paper, we develop a bioeconomic model to determine economically optimal harvesting regimes – in terms of thinning and rotation – of an even-aged plantation under the risk of an invading pest. We extend a Schaefer–Faustmann model [26] to include a compartmental epidemiological system that governs timber production and disease spread/dynamics. Using a dynamic compartmental model of disease separates our approach from previous bioeconomic studies that considered thinning and disease [33–35], while expanding the Macpherson et al. framework [28–30] to include thinning. The compartmental model can represent diseases that are unique from catastrophic events in their speed of progression, symptoms, and management response when detected [28]. Furthermore, it allows for interactions between thinning, tree growth and disease progression. We consider thinning's effect on both forest growth and disease spread: it increases growth by reducing the forest volume and thus reducing competition between trees; it reduces pest or pathogen spread by increasing the space and airflow between trees and tree resilience e.g., with *Dothistroma septosporum* [2]. Thinning and rotation thus effect both forest growth and disease dynamics. We provide insight into the system dynamics and sensitivity to key parameters (controlling infection spread rate secondary and severity/impact). Furthermore, we optimise the thinning timing and intensity during the rotation (unlike [34, 35]), and ex-

plore the optimal strategy — particularly the role of thinning.

This paper is structured as follows. In Section 2.3 we detail the general economic model and the underlying epidemiological system. The results corresponding to our model are given in Section 2.4, and discussed in Section 2.5. We give our concluding remarks in Section 2.6.

2.3 Model framework

We build on a Schaefer–Faustmann model [26] to determine the economically optimal harvesting regime – in terms of thinning and rotation – of an even-aged plantation under the risk of an invading pest. The model depends on a compartmental system that governs timber production and disease dynamics. Disease impact on the stands value is included by scaling the revenue obtained from the timber of infected trees or the growth of infected trees. Our timber production model is age and density-dependent, and operates at the stand volume level. It captures thinning reduces the volume of the forest, thinning resulting in increased tree growth (by lowering competition between trees), and thinning can reduce the spread of a pest or pathogen.

In the first subsection, we introduce this ecological system governing timber pro-

duction and disease dynamics. In the following subsection, we derive the maximisation problems to optimise management regimes for the plot. All relevant parameter definitions for the applied model are in Table 2.2.

Table 2.2: Parameter definitions, alongside their base case values.

Parameter	Description	Base case value [a]
ECOLOGICAL		
T_F	Rotation length/clear-felling time (years)	88.74 [b]
T_1	Thinning time (years)	60.40 [c]
γ	Proportion of trees thinned/thinning intensity	0.74 [d]
t	Time (years)	
$g(t)$	Age-dependent growth function	$0.13e^{-0.01t}$
K	Carrying capacity of plot (m^3)	378
EPIDEMIOLOGICAL		
$x(t)$	Susceptible (uninfected) timber volume (m^3)	
$y(t)$	Infected timber volume (m^3)	
$B(t)$	Transmission rate ($m^{-3}t^{-1}$)	Eq. (2.3)
β	Initial transmission rate ($m^{-3}t^{-1}$)	0.004
P	Primary infection rate (t^{-1})	0.0003
ε	Growth of infected timber relative to uninfected timber	1
δ	Impact of thinning on transmission rate	0
ECONOMIC		
$\hat{J}$	Net Present Value, NPV (£)	3.34
p_1	Price of uninfected thinned timber ($£m^{-3}$)	30.87
p_2	Price of uninfected clear-felled timber ($£m^{-3}$)	30.87
r	Discount rate	0.03
W_p	Establishment cost (£)	1000
ρ	Revenue from a unit of infected timber relative to uninfected timber	1

The base case values (a) represent a model without disease ($\varepsilon = \rho = 1$). The management variables (b) (c) (d) are the corresponding optimal strategy in a thinning and rotation regime (solving Eq. (2.7) for the base case values) — the "disease-free" strategy. The age-dependent growth increment, $g(t)$, was adapted from the numerical example in Tahvonen [26]. The establishment cost W_p comes from the same source. The timber prices p_1, p_2 are assumed equal and were taken from the Coniferous Standing Sales UK Price Index [36].

2.3.1 Forest dynamics — the compartmental timber production model

We now develop the two-state compartmental model of forest dynamics. We use Dothistroma needle blight, a foliar disease caused by the fungal pathogen *Dothistroma septosporum* [37], as an example to build the model. However, the assumptions we make are generic enough to represent other tree diseases with similar effects.

We compartmentalise timber on a hectare of even-aged monoculture into two states of infection ($N = 2$), infected with Dothistroma needle blight or not. We let $x(t) \geq 0$ and $y(t) \geq 0$ be the susceptible and infected stand volumes (m^3). Initially bare land is purchased and susceptible trees planted, with $x(0) = 1$ and $y(0) = 0$. The following (Susceptible-Infected) system governs the evolution of timber volumes in each state,

$$\begin{aligned} \frac{dx}{dt} &= g(t)x\left(1 - \frac{(x+y)}{K}\right) - (P + B(t)y)\,x \\ \frac{dy}{dt} &= \varepsilon g(t)y\left(1 - \frac{(x+y)}{K}\right) + (P + B(t)y)\,x. \end{aligned} \tag{2.1}$$

Additionally, at the thinning ($t = T_1$) and rotation times ($t = T_F$) both timber volumes are instantaneously reduced by the proportion $h(t)$ to represent harvests. We discuss this further in the following paragraphs.

We assume generally that the annual growth increment for timber in each state is

the product of an age-dependent function $g(t)$ and a density/volume-dependent function as in Clark and De Pree [24, 25] and Tahvonen [26]. Furthermore, the infected timber growth increment is multiplied by a constant ε to represent reduced growth from infection. The age-dependent function, $g(t)$, is positive and decreasing with time, representing the growth potential of a stand decreasing with age. The density-dependent growth function is a concave down quadratic with respect to the total forest volume [25, 26]. When competition (timber volume) is high, tree growth is limited, and reducing density can increase growth [14].

Specifically, we assume susceptible timber grows annually at a rate of $g(t)(1 - \frac{x+y}{K})$, which we adapted from the numerical example in Tahvonen [26] modelling the growth of Norway Spruce. Norway Spruce is susceptible to Dothistroma needle blight (DNB), but the susceptibility is low and requires high infection pressure [38]. The carrying capacity of the plot is K (m^3). We assume infected timber grows at the same rate as susceptible if $\varepsilon = 1$ and at a scaled rate if $0 < \varepsilon < 1$. Furthermore, if $\varepsilon = 1$, the total annual growth of all timber equals the age-dependent effect $g(t)$ multiplied by a total density/volume-dependent effect $(x + y)\left(1 - \frac{x+y}{K}\right)$, similarly to Clark and De Pree [24], Tahvonen [26] and Halbritter and Deegen [25]. We selected the functions $g(t)$ and $(1-\frac{x+y}{K})$ so growth represents plantation forestry. They result in logistic growth which tends to zero independently of stand density. After a clear-felling the volume does not grow back, but after a thinning it recovers. This represents the remaining trees after thinning (which reduces forest density) experiencing increased growth, resulting

in larger diameters [39].

Susceptible timber becomes infected with primary infection rate P. This represents external infection pressure, such as long range dispersal of spores in clouds or mist, or by movement of infected planting stock [40]. Within a forest containing infection, spread occurs primarily during periods of damp weather, which is conducive to fungal spore production. Spores are spread between trees by rain splashes and wind [37, 40]. Forest Research states that in the UK, control of the disease typically involves planting resistant species after rotation, and good stand management. In particular, this includes thinning in accordance with good silvicultural practice, to improve air flow and make conditions less conducive to fungus development [19]. Thinning also increases the distance between trees and, thus, reduces the effectiveness of rain-splashed spores [17] as a transmission route. Therefore, we assume that thinning removes a proportion of trees indiscriminately of their infection state. This assumption is reasonable for other diseases for which detection of infection is costly/difficult or the trees exhibit few outward signs of the disease until the later stages of infection. We denote the proportion of timber volume harvested at year t by $h(t)$, with

$$h(t) = \begin{cases} \gamma, & t = T_1 \\ 1, & t = T_F \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

We assume that a single thinning occurs (instead of annual [26]) at time $T_1 \geq 0$ and $\gamma \in [0, 1)$ of the total timber volume is removed, indiscriminate of its infection state. After T_F years, the plot is clear-felled ($x(t) = y(t) = 0$ for $t > T_F$) and lays bare.

We also assume that the transmission (secondary infection) rate, $B(t)$, which controls the spread of disease within the forest, is a step function,

$$B(t) = \begin{cases} \beta, & t \leq T_1 \\ \beta e^{-\delta\gamma}, & t > T_1. \end{cases} \tag{2.3}$$

The transmission rate is initially β. Thinning at $t = T_1$ can reduce the transmission rate, with the strength of this thinning effect determined by the factor $\delta \geq 0$. If there is an effect ($\delta > 0$), its impact increases with the proportion of trees thinned (γ). We assume a density-dependent transmission term [41] in Eq. (2.1), and therefore increased total forest volume (and therefore density in our model) results in increased spread within the forest.

2.3.2 Economic model

We now develop the single rotation Schaefer–Faustmann model for the even-aged forest, where the NPV includes the establishment cost, and the benefits from harvesting the timber. Initially, a hectare-sized plot is purchased and susceptible trees are planted, amounting to W_p (£'s) in establishment costs. Thinning at T_1

and clear-felling at T_F produces the timber benefits $H_{T_1}(\gamma, T_1)$ and $H_{T_F}(\gamma, T_1, T_F)$, both measured in £'s. After clear-felling, the land lays bare ($x(t) = y(t) = 0$ for $t > T_F$). We discount the revenue from harvesting at the rate r. Therefore, the NPV of the plot (the objective function) is

$$\hat{J}(\mathbf{s}) = -W_p + H_{T_1}(\gamma, T_1)e^{-rT_1} + H_{T_F}(\gamma, T_1, T_F)e^{-rT_F}, \tag{2.4}$$

where $\mathbf{s} = (\gamma, T_1, T_F)$.

Let p_1 and p_2 be the constant prices (£'s per m^3) for thinned and clear-felled susceptible timber, respectively. Assume $p_1 \leq p_2$: thinned timber is generally not as mature and therefore valuable as when clear-felled. Following the approach by Macpherson et al. [28, 30], we assume that the disease causes a reduction in the value of timber (e.g., through reduced quality or yield). In particular, DNB causes defoliation of the needles of an infected tree. This gradually weakens the tree, significantly reducing timber yields and eventually causing mortality [19]. Therefore, we let ρ be the revenue from a unit of infected timber relative to susceptible timber, where $0 \leq \rho \leq 1$.

We write the benefit from harvested timber in the each state as the product of the price per m^3 of standing timber, the volume of timber produced, and if timber is infected, then also the scaling parameter ρ. Therefore, the total timber

benefit at the thinning time, $t = T_1$, is

$$\begin{aligned} H_{T_1}(\gamma, T_1) &= p_1 h(T_1) x(T_1) + p_1 h(T_1) \rho y(T_1) \\ &= p_1 \gamma \left\{ x(T_1) + \rho y(T_1) \right\}, \end{aligned} \tag{2.5}$$

and the total timber benefit at the rotation time, $t = T_F$, is

$$\begin{aligned} H_{T_F}(\gamma, T_1, T_F) &= p_2 h(T_F) x(T_F) + p_2 h(T_F) \rho y(T_F) \\ &= p_2 \left\{ x(T_F) + \rho y(T_F) \right\}. \end{aligned} \tag{2.6}$$

Where Eq. (2.6) is a function of thinning proportion (γ) and time (T_1) because the volumes at rotation, $x(T_F)$ and $y(T_F)$, depend on these variables.

The forest manager's task is to maximise the NPV of the stand over a single rotation. We investigate two different management regimes. The first is thinning and rotation, where a single thinning occurs before clear-felling. There are three control variables; the thinning time (T_1), thinning proportion (γ), and rotation length (T_F). Their optimal values that maximise the objective function Eq. (2.4) are given by

$$\begin{aligned} \mathbf{s}^*_{tr} = \ \arg\max_{\mathbf{s}} \quad & \hat{J}(\mathbf{s}), \quad \text{where} \quad \mathbf{s} = (\gamma, T_1, T_F) \\ \text{subject to} \quad & \gamma \in [0, 1) \\ & T_F \geq T_1 \geq 0 \end{aligned} \tag{2.7}$$

The second management regime is rotation only, a Faustmann model. The maximisation problem is choosing the rotation length that maximises the NPV, as explored by Macpherson et al. [28]. The optimal strategy is

$$\mathbf{s}_r^* = \arg\max_{T_F \geq 0} \quad \hat{J}(\mathbf{s}), \quad \text{where} \quad \mathbf{s} = (0, 0, T_F) \tag{2.8}$$

We solved the optimisations problems Eq. (2.7) and Eq. (2.8) numerically in R using the Optim package and the L-BFGS-B algorithm. For Eq. (2.7) we ran the algorithm with three different start points and chose the best solution.

2.4 Results

We compare the optimal management strategies without disease in Subsection 2.4.1. In Subsection 2.4.2 we introduce disease, and in its first two subsections, we let susceptible and infected timber growth be the same ($\varepsilon = 1$), and assume that thinning does not effect transmission ($\delta = 0$). We first compare optimal management strategies for the case with no revenue from infected timber ($\rho = 0$) in Section 2.4.2. Sensitivity analysis of the revenue from a unit of infected timber relative to susceptible (i.e., $\rho > 0$) is undertaken in Section 2.4.2. Finally, we introduce and investigate the impact of the thinning effect ($\delta > 0$) and the reduction in infected timber growth ($\varepsilon < 1$) on the optimal strategy, Section 2.4.2.

To guide our intuition for these future sections, we will first find the optimal rotation length that maximises the NPV Eq. (2.4), assuming that the thinning

proportion (γ) and time (T_1) are fixed. We derive the first order condition by differentiating Eq. (2.4) with respect to T_F,

$$\frac{\partial \hat{J}}{\partial T_F} = p_2 \left\{x'(T_F) + \rho y'(T_F)\right\} e^{-rT_F} - r p_2 \left\{x(T_F) + \rho y(T_F)\right\} e^{-rT_F}. \tag{2.9}$$

Setting Eq. (2.9) equal to zero and rearranging gives the optimal rotation length condition

$$p_2 x'(T_F) + \rho p_2 y'(T_F) = r p_2 \left\{x(T_F) + \rho y(T_F)\right\}. \tag{2.10}$$

The optimal rotation length, T_F, which satisfies the above is the point when the rate of return of timber production equals the opportunity cost (rate of return of clear-felling and storing cash in the bank). We can show the rotation length T_F that satisfies this condition, Eq. (2.10), is a maximum by plotting the NPV as a function of T_F. Furthermore, by substitution of $x'(T_F)$ and $y'(T_F)$ from Eq. (2.1) into Eq. (2.10) and rearranging we can simplify the condition further to

$$\left(\frac{x + \rho\varepsilon y}{x + \rho y}\right) g(T_F) \left(1 - \frac{(x+y)}{K}\right) = r + (1-\rho)\left(\frac{(P + B(T_F)y)\, x}{x + \rho y}\right) \tag{2.11}$$

where we have used the shorthand $x = x(T_F)$ and $y = y(T_F)$.

We conclude that the NPV is maximised (w.r.t rotation length) when the rate of increase in the forest's clear-felled timber value from an additional year of growth equals the discount rate plus the loss in clear-felled timber value from the spread

of infection. This formulation shows that the impact of introducing disease on rotation length is effectively to increase the discount rate. It also highlights the trade-off between waiting for trees to grow and infection spreading further.

Similarly, we will now find the optimal thinning time that maximises the NPV, Eq. (2.4), assuming that the thinning proportion (γ) and rotation time (T_F) are fixed. We derive the first order condition by differentiating Eq. (2.4) with respect to T_1,

$$\begin{aligned}\frac{\partial \hat{J}}{\partial T_1} &= p_1\gamma\left\{x'(T_1)+\rho y'(T_1)\right\}e^{-rT_1} - rp_1\gamma\left\{x(T_1)+\rho y(T_1)\right\}e^{-rT_1} \\ &\quad + p_2\frac{\partial}{\partial T_1}\left\{x(T_F)+\rho y(T_F)\right\}e^{-rT_F}. \end{aligned} \tag{2.12}$$

Setting Eq. (2.12) equal to zero and rearranging after substitution of $x'(T_F)$ and $y'(T_F)$ from Eq. (2.1) into Eq. (2.12) gives the optimal thinning time condition

$$\begin{aligned}&\left(\frac{x(T_1)+\rho\varepsilon y(T_1)}{x(T_1)+\rho y(T_1)}\right)g(T_F)\left(1-\frac{x(T_1)+y(T_1)}{K}\right) \\ &= r+(1-\rho)\left(\frac{(P+B(T_1)y(T_1))\,x(T_1)}{x(T_1)+\rho y(T_1)}\right) \\ &\quad - \frac{p_2\frac{\partial}{\partial T_1}\left\{x(T_F)+\rho y(T_F)\right\}e^{-r(T_F-T_1)}}{p_1\gamma\left(x(T_1)+\rho y(T_1)\right)}.\end{aligned} \tag{2.13}$$

Therefore, we conclude that the NPV is maximised (w.r.t thinning time) when the rate of increase in the thinned timber benefit from an additional year of growth equals the discount rate, plus the loss rate from the spread of infection, minus the discounted change in timber benefit at rotation relative to the thinned timber

benefit. Eq. (2.13) shows that introducing disease has an impact on thinning time similar to the optimal rotation length (Eq. (2.11)) — disease effectively adds to the discount rate. However, it stresses that the choice of thinning time must account for the impact of thinning on growth and disease dynamics and, therefore, the timber benefit at rotation.

2.4.1 No disease

We begin with the simplified non-disease version of the model. We can derive this model by setting $P = \beta = 0$ in Eq. (2.1). Alternatively, by setting $\varepsilon = 1$ in Eq. (2.1) and $\rho = 1$ in Eq. (2.5), the disease has no impact on the timber benefit or timber growth. The equation below governs timber dynamics when there is no disease,

$$x'(t) = g(t)x(1 - \frac{x}{K}). \tag{2.14}$$

At the thinning ($t = T_1$) and rotation times ($t = T_F$) the susceptible timber volume is instantaneously reduced by the proportion $h(t)$.

Without disease, thinning extends the optimal clear-felling time compared to the optimal in a clear-felling only regime. This result is a consequence of density-dependent growth, which implies the forest growth rate increases after thinning. We then need to wait longer for the rate of return of timber production to slow and eventually equal the opportunity cost (rate of return of clear-felling and storing cash in the bank). This result can be deduced using the optimal rotation

length condition (Eq. (2.11)), which simplifies to

$$g(T_F)\left(1-\frac{x(T_F)}{K}\right)=r \tag{2.15}$$

when there is no disease. First, note that the left-hand side of Eq. (2.15) is a decreasing function of rotation length, T_F, and the right-hand side a constant. Let the optimal rotation length without thinning be T_F^*. If the timber volume without thinning ($\gamma = 0$) is always larger than the timber volume with thinning ($\gamma > 0$) at $t = T_F^*$, i.e., if $x_{\text{no thin}}(T_F^*) \geq x_{\text{thin}}(T_F^*)$, then from Eq. (2.15) it is clear that optimal rotation with thinning is greater than without. We can show that this is true by solving Eq. (2.14) using separation of variables for $\gamma = 0$ and $\gamma > 0$ and considering the cases $T_F^* > T_1$ and $T_F^* \leq T_1$. See Section 2.B for the details.

Undisturbed growth (no thinning or rotation in Eq. (2.1)) of the forest is shown in Figure 2.1. The density dependence of the forest is highlighted by the optimised thinning and rotation (dotted) line in Figure 2.1, as after thinning reduces the volume the growth rate increases.

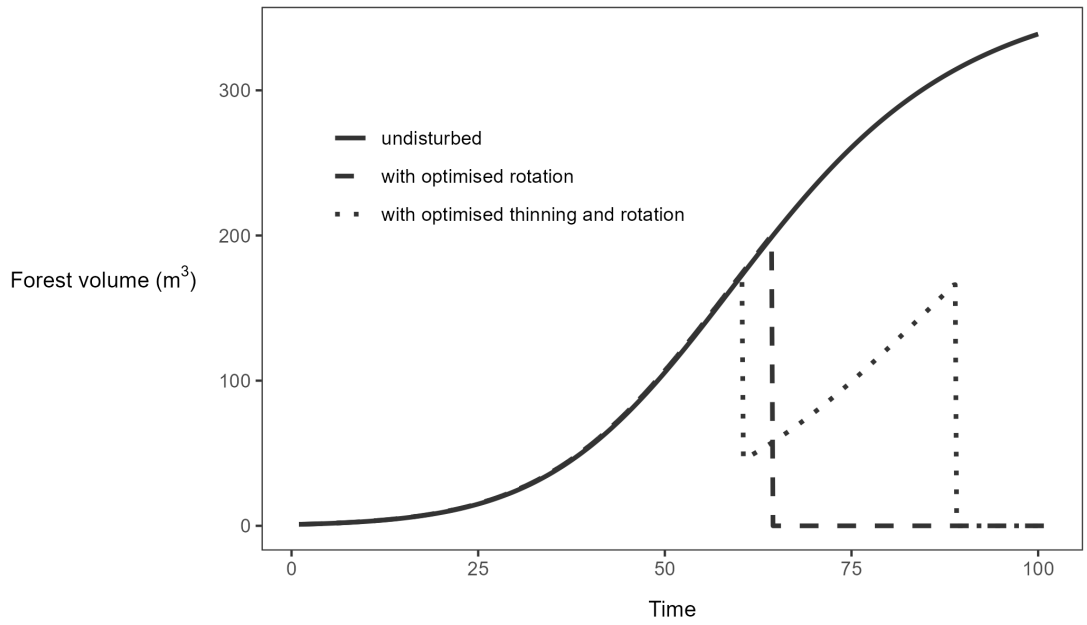

Fig. 2.1: Timber volume trajectories in the absence of disease. Undisturbed timber growth is the thick black line (Eq. (2.1) with $T_1 = \gamma = 0$, $T_F = \infty$, $\beta = P = 0$ and other parameter values as in Table 2.2). The dashed line is the timber volume trajectory ($x(t)$ from Eq. (2.1)) under an optimised clear-felling only management regime (T_1, γ, and T_F given by Eq. (2.8) for $\beta = P = 0$ and other parameter values as in Table 2.2). The dotted line is the timber volume trajectory ($x(t)$ from Eq. (2.1)) under an optimised clear-felling and thinning management regime (solving Eq. (2.8) for $\beta = P = 0$ and other parameter values as in Table 2.2).

We highlight that thinning extends the rotation time in Figure 2.1, where we compare the two optimised management regimes for the base case economic and ecological parameters (Table 2.2). The rotation only regime has an optimal rotation length of $T_F \approx 64$, when it is clear-felled. In the optimal thinning and rotation regime, approximately 74% of the total volume is thinned after $T_1 \approx 60$ years, which is just before the optimal rotation in the rotation only regime. Then, after $T_F \approx 89$ years, the plot is clear-felled. We refer to this thinning and rotation

strategy as the "disease-free" strategy and use it as a baseline for comparison in later sections of this paper.

We also performed a sensitivity analysis of the optimised thinning and rotation regime to the parameters controlling growth and the price difference between clear-felled and thinned timber. We generally find two optimal strategies; (i) rotation only (no thinning), or (ii) thin $\sim 70\%$ of the trees late in the rotation. These parameters have tipping points where the strategy switches between the two strategies. When the growth rate parameters result in quick timber growth, or the price difference between clear-felled and thinned timber is small, thinning is optimal. The optimal strategy switches to no thinning for a large price difference or a small growth rate.

2.4.2 Disease

No revenue from infected timber

In this section, we investigate the sensitivity of optimal strategies to two epidemiological parameters: the primary infection rate (P) and transmission rate (β). We assume that thinning does not affect transmission ($\delta = 0$) and that the growth rate of infected volume is identical to susceptible ($\varepsilon = 1$); our base case parameter values. With these assumptions the infected volume takes up capacity, limiting the growth of susceptible timber. Therefore, the disease does not impact the density-dependent growth. Instead, we represent a reduced yield by assuming

there is no revenue from the infected timber volume ($\rho = 0$). In this scenario, if we fix the strategy in a thinning and rotation regime to be the "disease-free" one, then the total timber volume follows the dashed line in Figure 2.1. An increased transmission rate (β) or primary infection rate (P) will speed up the spread of disease and therefore the prevalence of infection at each harvest, Figure 2.2.

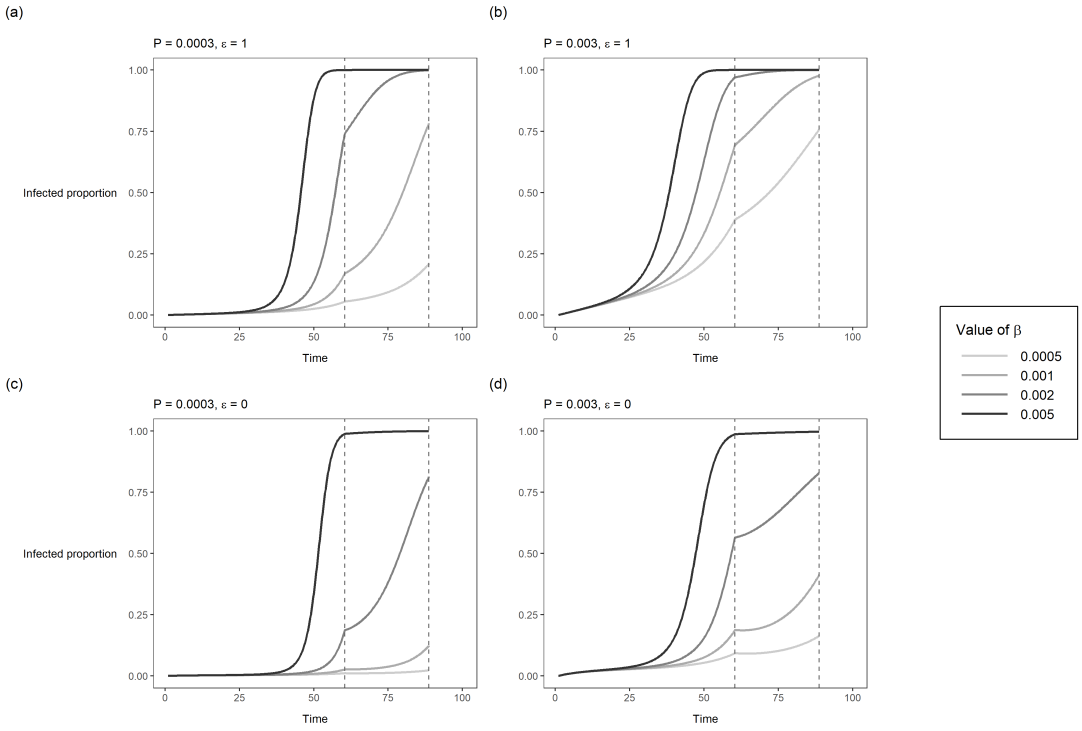

Fig. 2.2: Disease progress curves under the thinning and clear-felling regime. Epidemiological parameters are varied between panels, but the management regime is fixed to the optimal thinning and clear-felling one in absence of disease. In each panel, the annual cumulative proportion of timber volume infected ($\frac{y(t)}{x(t)+y(t)}$ after solving Eq. (2.1)) is shown for four values of the transmission rate (β). Each panel shows results for a different combination of the primary infection rate (P) and growth of infected timber relative to susceptible timber (ε). The thinning and clear-felling times are the dotted vertical lines in each panel (values shown in Table 2.2, calculated by solving Eq. (2.7) for $\beta = P = 0$). All other ecological and economic parameters are given in Table 2.2.

In Figure 2.2(a) and for sensitivity analyses in the following sections of this paper, we set the primary infection $P = 0.0003$, the base case value (Table 2.2). This value ensures a disease outbreak of some form during a typical rotation (64 years — the optimal without thinning or disease in our model). Furthermore, by sweeping through transmission rates (β) in the range $[0, 0.005]$, we capture a large variation in disease progress curves, see Figure 2.2(a). For example, with a low transmission rate $\beta = 0.0005$, the infection spreads very slowly after arrival, and after 60 years less than 5% of the forest has been infected. Then for the transmission rate $\beta = 0.001$, after 60 years approximately 15% of the forest has been infected. Whereas for higher secondary infection rates, $\beta = 0.002, 0.005$ after 60 years approximately 75% and close to 100% of the forest will be infected.

Increasing the primary infection (P) or transmission rate (β) will shorten the optimal rotation length for each management regime when all other parameters are fixed. To see this, we note that the optimal rotation length condition (Eq. (2.11)) assuming the thinning proportion (γ) and time (T_1) are fixed and no revenue from infected timber ($\rho = 0$) becomes

$$g(t)\left(1 - \frac{x(T_F) + y(T_F)}{K}\right) = r + \left(P + B(T_F)y(T_F)\right). \tag{2.16}$$

As the LHS of Eq. (2.16) is a positive decreasing function of T_F, and the RHS is a positive increasing function of T_F, the LHS and RHS functions will intersect once if plotted. This allows us to deduce the high-level impact of increasing the primary infection (P) or transmission rate (β — which increases $B(T_F)$). Below

we use numerical methods to further explore and visualise the interaction between these parameters on the full optimal strategy for each regime.

In a rotation/clear-felling only regime, if the transmission rate (β) is close to zero, a long rotation ($\approx$ 60 years) is optimal, similar to the disease-free length (see Figure 2.3(a)). If the external infection pressure or spread rate of disease is higher (increased P or β) then the optimal rotation length (T_F) decreases, as qualitatively shown in Figure 2.3(a). Shortening the rotation length allows timber to be salvaged before infection comes (Figure 2.3(c)) and destroys its value. The optimal rotation length and maximum NPV are more sensitive to changes in the transmission rate (β) than the primary infection rate (P), see Figure 2.A.1 in Section 2.A which shows the complete breakdown of the optimal strategy in the $\beta - P$ parameter space.

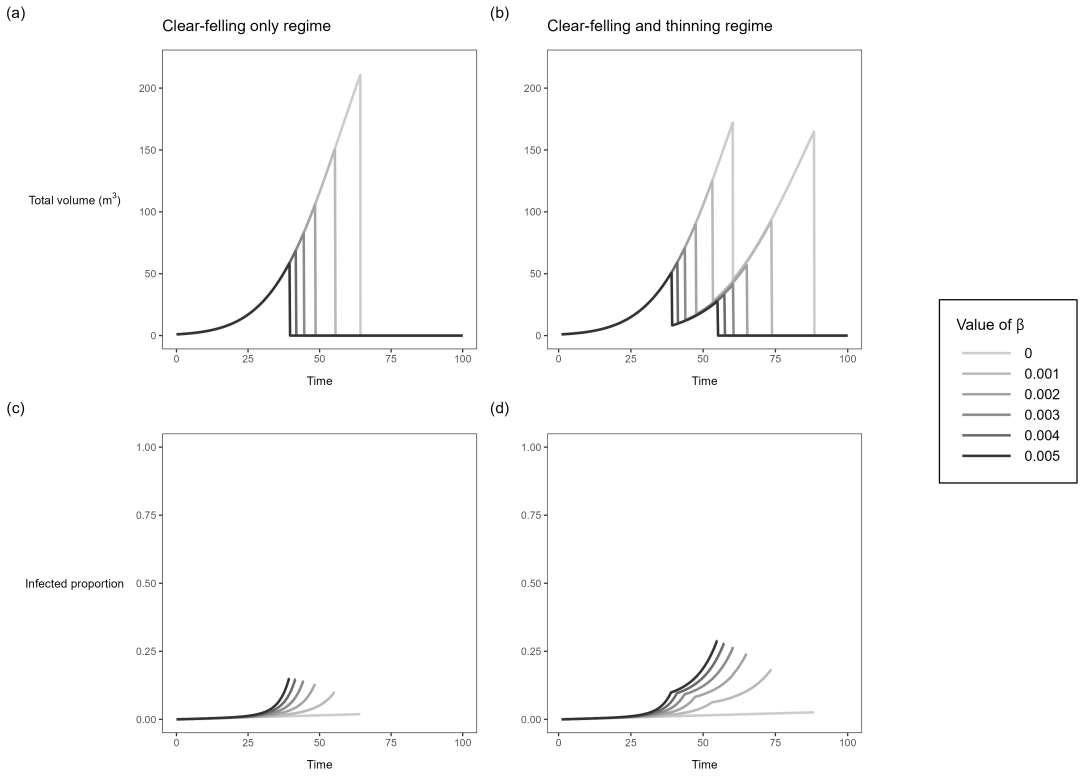

Fig. 2.3: Impact of transmission rate on optimised management strategies when no revenue comes from infected timber ($\rho = 0$). The top row of panels are the total timber volume trajectories ($x(t) + y(t)$) under the optimised regimes, and the bottom shows the corresponding cumulative proportion of timber infected ($\frac{y(t)}{x(t)+y(t)}$). $x(t)$ and $y(t)$ are given by Eq. (2.1) with management variables (γ, T_1 and T_F) from either Eq. (2.8) (rotation only regime) or Eq. (2.7) (thinning and rotation regime). Darker lines within panels indicate higher values of the transmission rate (β). (a) the total timber volume each year under optimised clear-felling only; (b) the total timber volume each year under optimised thinning and clear-felling; (c) the cumulative proportion of timber infected under optimised clear-felling only; (d) the cumulative proportion of timber infected under optimised thinning and clear-felling.

Similarly to the rotation only regime, if the transmission rate is (β) close to zero, it is optimal to follow the "disease-free" strategy with late thinning and long rotation, Figure 2.3(b). If the external infection pressure or spread rate of disease

is higher (increased P or β), this brings forward the optimal rotation (T_F) and thinning time (T_1) and increases the thinned proportion (γ), compared to the "disease-free" strategy, Figure 2.3(b). Salvaging the timber quickly with an early thin and rotation before the infection destroys its value (Figure 2.3(d)) becomes optimal (and is intuitive). We refer to this type of strategy as "salvage quickly". The optimal thinning time, proportion, rotation length, and maximum NPV are more sensitive to changes in the transmission rate (β) than the primary infection rate (P), see Figure 2.A.2 in Section 2.A which shows the optimal strategy in the $\beta - P$ parameter space.

Disease causes a severe reduction in maximum NPV for the thinning and rotation regime, particularly for high transmission rate (β). Figure 2.4(b) highlights this, where we compare the NPV of using the disease-free strategy (thinning and rotating late) to the optimised strategy for a thinning and rotation regime in the presence of disease. Not shortening rotations when the transmission rate is high ($\beta > 0.002$) will result in immense NPV losses ($> £400$).

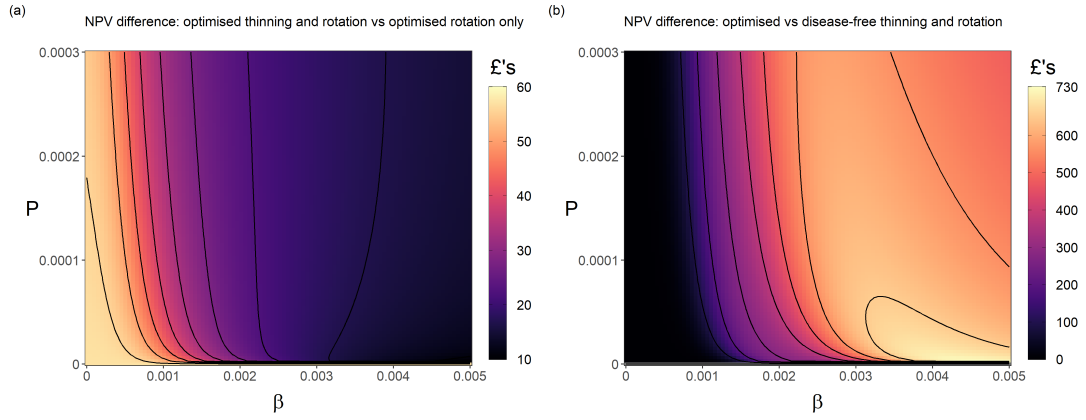

Fig. 2.4: NPV differences between strategies in a $P - \beta$ parameter space. (a) Difference between the maximum NPV of an optimised thinning and clear-felling regime vs an optimised clear-felling only regime. (b) Difference between using the disease-free strategy (thinning and rotating late) and the optimised strategy for a thinning and rotation regime in the presence of disease. The maximum NPV for optimised thinning and clear-felling is given by substitution of Eq. (2.7) in Eq. (2.4), and for clear-felling only by substitution of Eq. (2.8) in Eq. (2.4). The disease-free management strategy for thinning and clear-felling is given by solving Eq. (2.8) for $\beta = P = 0$, keeping all other parameter values as in Table 2.2. P is the primary infection rate and β is the transmission rate. We assume no revenue comes from infected timber ($\rho = 0$). All other parameter values are given in Table 2.2.

For any fixed value of the primary infection and transmission rates (P and β), the optimal rotation lengths are longer in the thinning and rotation regime compared to those in the rotation-only regime. Thinning times in the thinning and rotation regime occur at roughly the same times as rotation times in the rotation-only regime. To see this, compare Figure 2.A.1(a) to Figure 2.A.2(d) in Section 2.A, or compare Figure 2.3(a) to (b) for an example. The extension is likely due to thinning increasing the forest's growth rate, as discussed in Subsection 2.4.1, and thinning reducing secondary infection (Figure 2.2).

Thinning in combination with clear-felling is always optimal compared to a regime of clear-felling only, Figure 2.4(a). Furthermore, thinning is even more profitable when the disease spreads slowly ($\beta < 0.001$ in Figure 2.4(a)). Less timber gets destroyed and can be left to grow for longer. Therefore, there can be more time between thinning and clear-felling, allowing the non-thinned trees to grow and exploit the forest's density-dependent growth. A strategy that cannot be applied with clear-felling only. Furthermore, the benefit to including thinning in the management regime decreases with increased transmission rate (β), Figure 2.4(a)), becoming very small ($< £20$) for $\beta > 0.004$. However, comparing Figure 2.4(b) (where we compare the NPV of using the disease-free strategy, thinning and rotating late, to the optimised strategy for a thinning and rotation regime in the presence of disease) to Figure 2.4(a) in this region ($\beta > 0.004$), we conclude that the largest benefit comes from shortening the rotation, independent of thinning.

Sensitivity to the revenue from a unit of infected timber relative to uninfected timber

In Section 2.4.2 we assumed that $\rho = 0$, a worst-case scenario in which timber revenue comes from uninfected timber only. We now investigate the sensitivity of optimal strategies to the effect of infection on timber revenue (i.e., $\rho > 0$). This scenario implicitly represents a smaller impact of infection on yields (through infected tree growth/deaths) or weakened timber being sold at a reduced price. We again assume that thinning does not affect transmission ($\delta = 0$) and that the growth rate of infected volume is identical to susceptible ($\varepsilon = 1$).

Under different combinations of the transmission rate (β) and the infected timber revenue scaling factor (ρ), the optimal strategy (T_F) in a rotation only regime (Eq. (2.8)) can be categorised into two groups, Figure 2.5(a), (c). There is (i) a long rotation used in the disease-free case, and (ii) shorter rotations to salvage timber before infection lessens its value (highlighted in Figure 2.5(a), (c)). Figure 2.A.3 in Section 2.A gives a breakdown of the optimal strategy in the $\beta - \rho$ parameter space, while Figure 2.5(a) and (c) show the optimal strategy at a transect of the transmission rate ($\beta = 0.004$). Along this transect, the two types of strategy that appear in the whole space can be visualised.

Greater damages from infection (lower ρ) typically decrease the optimal rotation length from the disease-free one. The size of the reduction depends on the transmission rate. When the infection spreads slowly (low β), not enough timber is infected for any reduction in timber value to change the optimal rotation length from the disease-free one, Figure 2.A.3 in Section 2.A. When the impact of infection on timber revenue is high ($\rho \leq 0.5$), the optimal length shortens from the disease-free length as the transmission rate (β) increases, Figure 2.A.3 in Section 2.A. The shorter rotation means a lower volume of timber is salvaged, but it is salvaged before infection destroys its value. However, when the impact of infection on timber revenue is lower ($0.5 < \rho < 1$), there is a tipping point with the transmission rate (β); below this value as the transmission rate increases the optimal rotation length decreases, and above this value the optimal rotation

length switches back to the optimal disease-free rotation length. The higher the impact of infection on timber revenue (lower $\rho \in (0.5, 1)$), the larger the transmission rate must be for the switch, and the less smooth the switch is. At the switch, the NPV of letting the trees grow larger but more get infected overtakes the NPV of felling a lower yield before they become infected.

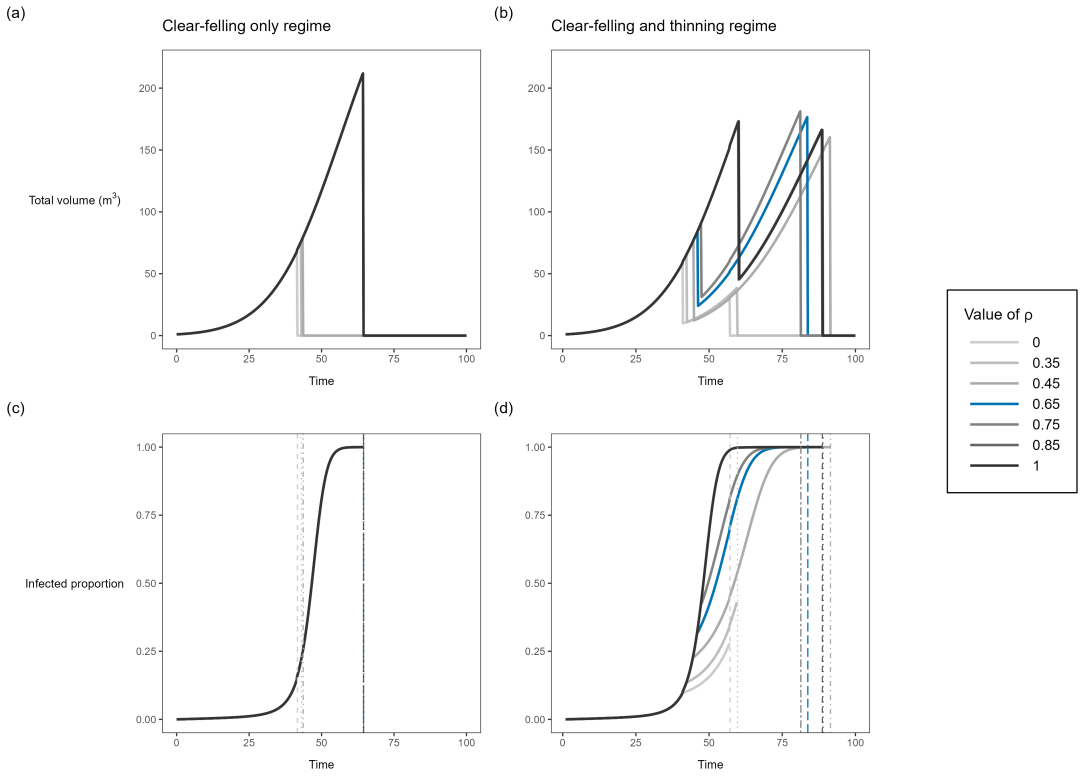

Fig. 2.5: Impact of transmission rate on optimised management strategies when there is revenue from infected timber ($\rho > 0$). The clear-felling only regime, and the thinning and clear-felling regime were optimised for five different values of the revenue from a unit of infected timber relative to susceptible (ρ) by solving Eq. (2.8) and Eq. (2.7). All other parameter values were held at the base case values in Table 2.2. The top row of panels are the total timber volume trajectories ($x(t) + y(t)$) under the optimised strategies, and the bottom shows the corresponding cumulative proportion of timber infected ($\frac{y(t)}{x(t)+y(t)}$). The vertical lines in the bottom row indicate the clear-felling times from the top row. $x(t)$ and $y(t)$ are given by Eq. (2.1) with management variables (γ, T_1 and T_F) from either Eq. (2.8) (rotation only) or Eq. (2.7) (thinning and rotation). Darker lines within panels indicate higher values of the revenue from a unit of infected timber relative to susceptible (ρ). The blue line highlights the shown value of ρ for which the NPV difference between the thinning and clear-felling regime vs the clear-felling regime is largest (Figure 2.6(a)). In panel (a) and (c) the $\rho \geq 0.65$ lines (including blue) are hidden behind the black $\rho = 1$ line. (a) the total timber volume each year under optimised clear-felling only; (b) the total timber volume each year under optimised thinning and clear-felling; (c) the cumulative proportion of timber infected under optimised clear-felling only; (d) the cumulative proportion of timber infected under optimised thinning and clear-felling.

Broadly, three types of strategies emerge when we consider the optimal strategy for the thinning and rotation regime (Eq. (2.7)) under combinations of the transmission rate (β) and the infected timber revenue scaling factor (ρ), Figure 2.5(b) (d). Figure 2.A.4 in Section 2.A gives a breakdown of the optimal thinning and rotation strategy in the $\beta - \rho$ parameter space, while Figure 2.5(b) and (d) show the optimal strategy at a transect of the transmission rate ($\beta = 0.004$). Along this transect, the three types of strategy that appear in the whole space can be visualised. The first is following the "disease-free" strategy discussed previously. Here, the strategy suggests acting as if there is no disease in the forest — long rotation with late thinning (visualised by the $\rho = 0.85, 1$ lines in Figure 2.5(b) (d), and in Figure 2.1). This strategy is optimal when either (i) the transmission rate is very low (low β), or (ii) the transmission rate is very high, but infection does not cause much damage (high ρ). The second strategy is "salvage quickly", introduced in Section 2.4.2, and visualised by the $\rho = 0, 0.35$ lines in Figure 2.5(b) (d). Here, we thin and rotate earlier. "Salvage quickly" is the optimal strategy when the infection spreads quickly (high β), causing significant damage (low ρ). The final strategy is "best of both worlds", which is optimal when the infection spreads reasonably quickly but causes middling damage to timber ($0.5 \leq \rho \leq 0.8$). The strategy suggests thinning to salvage valuable timber before widespread infection, and then leaving the remaining timber to get infected during a long rotation. The $\rho = 0.45, 0.65, 0.75$ lines in Figure 2.5(b) (d) demonstrate this strategy. Note that for any fixed value of the transmission rate and revenue from infected timber (β and ρ), the optimal rotation lengths are

longer in the thinning and rotation regime compared to those in the rotation-only regime. To see this, compare Figure 2.A.1(a) to Figure 2.A.2(d) in Section 2.A, or compare Figure 2.5(a) to (b) for an example. The extension is likely due to thinning increasing the forest's growth rate and reducing secondary infection.

The transitions between these different types of optimal strategy suggested in the $\beta - \rho$ parameter space (Figure 2.A.4 in Section 2.A) depend on what the impact of infection on timber revenue (ρ) and transmission rate (β) are. When the impact of infection on timber revenue is high ($\rho \leq 0.4$), the optimal strategy for the thinning and rotation regime changes smoothly with increases in the transmission rate (β), Figure 2.A.4 in Section 2.A. There is a moderate transition from the general "disease-free" to the "salvage quickly" strategy. Optimal rotation and thinning times decrease at a similar rate, and the optimal thinning proportion increases. When the impact of infection on the timber revenue is lower ($\rho > 0.7$), as the transmission rate (β) increases, the optimal rotation and thinning times initially decrease together, and the optimal thinned proportion increases from the "disease-free" values. Then at a tipping point of β (≈ 0.001), this pattern switches. The optimal rotation and thinning time increase together, and the proportion thinned decreases. Furthermore, as β increases further, the rate that the optimal thinning time increases relative to the optimal rotation length slows down, and the optimal strategy becomes a "best of both worlds" type. The impact of infection on timber value (ρ) determines the sensitivity of this switch and the transmission rate where it occurs. When the impact is minor

($\rho > 0.7$), a gradual switch occurs at smaller values of β. When the impact is more significant ($\rho \approx 0.5$), then a tipping point occurs at larger values of the transmission rate (β). At this tipping point, the optimal strategy changes from a "salvage quickly" one (rotating and thinning early) to "best of both worlds" (rotating late but thinning early), visualised by Figure 2.5(b) (d).

The NPV differences between the optimised rotation-only regime and the thinning and rotation regime are higher in parameter spaces where the optimal rotation lengths for the thinning and rotation regime are longer (low β, or high β and high ρ), Figure 2.6(a). This is because we thin to get an early income, and then exploit the forests density-dependent growth to let the remaining trees grow large. The optimised rotation-only regime cannot offer this option. When the optimal rotation lengths are shorter for both regimes (e.g., high β and low ρ), the density-dependent growth cannot be exploited to the same degree in the optimised thinning and rotation regime (less time between harvests), so the NPV differences are small. The NPV difference between the optimised rotation-only regime and the thinning and rotation regime is highest in the parameter space where the "best of both worlds" thinning and rotation strategy is optimal, see the yellow region in Figure 2.6(a). In the centre of this space the optimal strategy for the rotation-only regime switches between rotating early and rotating late (compare $\rho \leq 0.45$ to $\rho \geq 0.65$ lines in Figure 2.5(a)). The "best of both worlds"

thinning and rotation strategy effectively combines early and late rotation by thinning early and rotating late (blue $\rho = 0.65$ line in Figure 2.5(b)); there is not an equivalent strategy for the rotation only regime.

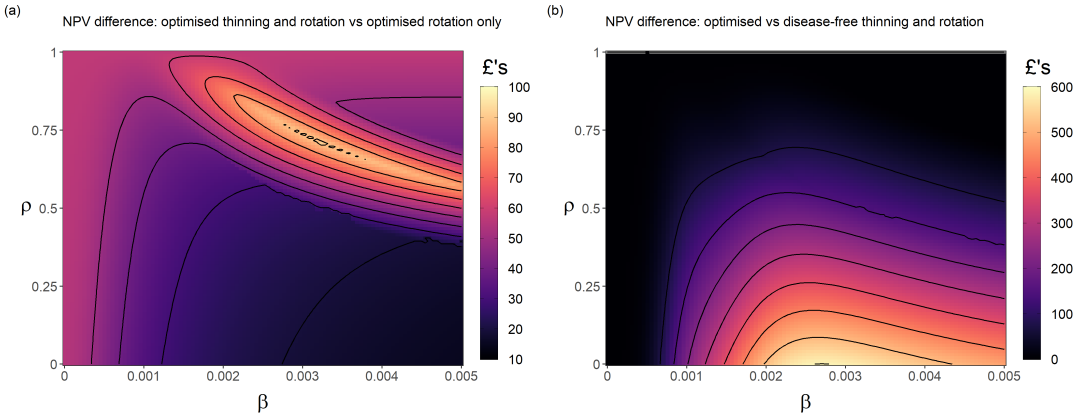

Fig. 2.6: NPV differences between strategies in a $\rho - \beta$ parameter space. (a) Difference between the maximum NPV of an optimised thinning and clear-felling regime vs an optimised clear-felling only regime. (b) Difference between using the disease-free strategy (thinning and rotating late) and the optimised strategy for a thinning and rotation regime in the presence of disease. The maximum NPV for optimised thinning and clear-felling is given by substitution of Eq. (2.7) in Eq. (2.4), and for clear-felling only by substitution of Eq. (2.8) in Eq. (2.4). The disease-free management strategy for the thinning and clear-felling regime is given by solving Eq. (2.8) for $\beta = P = 0$, keeping all other parameter values as in Table 2.2. ρ is the revenue from a unit of infected timber relative to susceptible, and β is the transmission rate. We assume no revenue comes from infected timber ($\rho = 0$). All other parameter values are given in Table 2.2.

The consequences (NPV losses) of not changing the strategy in the thinning and rotation regime from "disease-free" increase with the transmission rate and the impact of infection on timber revenue, Figure 2.6(b). Not shortening the rotation and thinning time when the transmission rate (β) and the impact of infection on timber revenue are high (low ρ) will result in substantial NPV losses

($> £500$), Figure 2.6(b). However, the losses decrease when the transmission rate is exceptionally high as timber is destroyed too quickly compared to the speed at which it grows. When the infection spreads reasonably quickly and causes middling damage to timber ($0.5 \leq \rho \leq 0.8$), it is optimal to use the "best of both worlds" strategy for the thinning and rotation regime. Using this strategy instead of the "disease-free" one in this region provides a slight increase in NPV ($\approx £100$), Figure 2.6(b).

Sensitivity to the impact of thinning on the transmission rate

We now introduce a thinning effect on the transmission rate (setting $\delta > 0$) and investigate the effect on the optimal strategy for the thinning and rotation regime (thinning time, thinning proportion and clear-felling time). We continue to assume that the growth rate of infected volume is identical to susceptible ($\varepsilon = 1$) and no revenue from the infected timber volume ($\rho = 0$).

If the time (T_1) and proportion (γ) of the thinning regime are fixed, increasing the effect of thinning on the transmission rate (δ) will extend the optimal rotation length. This general result follows from the assumption that thinning reduces the transmission rate, $\frac{\partial B(T_F)}{\partial \gamma} < 0$, and can be deduced from Eq. (2.16) by noting that increasing the thinning effect (δ) reduces $B(T_F)$. Using numerical methods we explored this result further to see how the transmission rate (β) and thinning effect (δ) interact with respect to the full optimal thinning and rotation strategy. We solved Eq. (2.7) to find the optimised strategy for different combi-

nations of the transmission rate (β) and the thinning effect (δ), holding all other parameters at the base case (Figure 2.A.7 in Section 2.A).

For low transmission rate values (β), the disease does not spread quickly enough and cause enough damage to warrant changing the strategy from "disease-free", i.e., thinning and rotating late to exploit the forest density-dependent growth, Figure 2.A.7 in Section 2.A. Furthermore, when the thinning effect is low ($0 \leq \delta < 2$), increased transmission rate (β) compresses the thinning and rotation times and slightly increases the proportion thinned, as seen previously in Section 2.4.2. However, when the thinning effect is more substantial ($\delta > 2$) and the transmission rate higher ($\beta > 0.001$), increasing the thinning effect (δ) extends the rotation, brings the optimal thinning time forward, and decreases the optimal proportion thinned. The faster infection spreads (higher β), the sooner we thin. The "disease-free" rotation is an upper bound for the rotation length increase. We conclude that including a strong thinning effect on transmission changes the reason for thinning. Instead of thinning and rotating to get early harvests before infection destroys the value, the optimal strategy suggests thinning early and lightly to protect the final harvest, which can grow larger without being destroyed by the disease. Therefore, the maximum NPV is much higher when there is any thinning effect ($\delta > 0$). The key result of introducing a decline in the transmission rate from thinning bringing the thinning time forward while pushing back the rotation time holds when we change the shape of the decline function. We re-ran the simulation that produced Figure 2.A.7 using the function

$\frac{\beta}{1+\delta\gamma}$ in place of $\beta e^{-\delta\gamma}$ in Eq. (2.3) and found a similar but less prominent result (not shown).

Furthermore, when infected timber is worthless and the infection spreads quickly ($\beta > 0.0025$), without the thinning impacting the transmission rate, the benefit of thinning and clear-felling over clear-felling is small (Figure 2.6(a)). However, comparing Figure 2.A.7(e) ($\rho = 0$) in Section 2.A and Figure 2.6(a), when thinning impacts the transmission rate ($\delta > 0$), the benefit in NPV of thinning and clear-felling over only clear-felling alone becomes massive.

Sensitivity to the infected timber growth rate

We now test the sensitivity of the optimised thinning and rotation regime to the growth rate of infected timber relative to susceptible timber (ε). We assume that thinning does not affect transmission ($\delta = 0$) and that revenue from a unit of infected and a unit of susceptible timber volumes are equal ($\rho = 1$). In this scenario, infection disrupts the growth of timber. Furthermore, due to the forest's density-dependent growth, as infected timber grows at a reduced rate, susceptible timber volume can grow in its place.

In this scenario where infected timber does not grow at all ($\varepsilon = 0$), if we fix the strategy for the thinning and rotation regime to be the "disease-free" one, then the proportion of volume infected follows Figure 2.2(c) and (d). Fixing the epidemiological parameters and comparing (c) and (d) (where $\varepsilon = 0$) to (a) and

(b) in Figure 2.2 (where $\varepsilon = 1$), when the growth rate of infected timber is lower, a larger proportion of the timber volume is susceptible at each harvest.

Under different combinations of the primary (P) and transmission (β) rates, the optimal strategy for the thinning and rotation regime in the $\varepsilon = 0$ scenario (Figure 2.A.5 in Section 2.A) is qualitatively similar to the optimal strategy in the scenario that assumes no revenue from infected timber ($\rho = 0$, Figure 2.A.2 in Section 2.A). However, the NPV's are higher in the $\varepsilon = 0$ scenario. Again, we see that the model is more sensitive to the transmission rate (β) compared to the primary infection rate (P), Figure 2.A.6 in Section 2.A. The optimal rotations shorten and optimal thinned proportion increase from their disease-free values with increases in the transmission rate (β), but at slower rates than when $\rho = 0$. The forest's density-dependent growth is responsible for these slower rates. With $\varepsilon = 0$, susceptible trees/volume can grow larger as the infected trees/volume do not grow. Therefore the forest does not need to be clear-felled sooner to recoup costs before infection destroys the value.

Comparing strategies in the $\beta - \varepsilon$ parameter space (Figure 2.A.6 in Section 2.A) to the $\beta - \rho$ parameter space (Figure 2.A.4 in Section 2.A), we see that ε and ρ have somewhat similar effects on the optimal strategy for the thinning and rotation regime, but with some key differences between the effects of ε and ρ. The three types of strategies ("disease-free", "salvage quickly", and a strategy similar to "best of both worlds") outlined in Section 2.4.2 appear in Figure 2.A.6 in Sec-

tion 2.A. When the infection spreads slowly ($\beta \leq 0.002$), timber is not infected quickly enough for any changes in the infected timber growth rate to alter the optimal strategy from "disease-free" (thinning and rotating later, exploiting density dependence), Figure 2.A.6 in Section 2.A. Similarly, in the $0.8 < \varepsilon \leq 1$ parameter space, it is always optimal to use a "disease-free" type strategy, letting timber become infected under the delayed harvests. The disease does not impact growth enough. If the transmission rate is higher ($\beta > 0.002$), more timber gets infected, and the optimal strategy is much more sensitive to changes in the infected timber growth rate (ε), Figure 2.A.6 in Section 2.A. When the transmission rate (β) is high ($\beta > 0.002$) and the growth rate of infected timber is low ($\varepsilon \leq 0.4$), the optimal rotation length shortens, the optimal thinned proportion increases and thinning occurs slightly later in the rotation, compared to in the "disease-free" type strategy. The strategy becomes a "salvage quickly" type when the growth rate of infected timber is low, and the disease spreads quickly.

Decreasing the infected timber growth rate (ε) has a non-monotonic relationship with optimal rotation length, whereas a reduction in infected timber revenue (ρ) generally always decreases optimal rotation length. When the transmission rate is high ($\beta > 0.003$), reducing the growth rate of infected timber (ε) in the range $0.4 < \varepsilon < 0.8$ gradually decreases the optimal rotation length and thinned proportion, while slightly increasing the timing of the thin relative to the rotation length, Figure 2.A.6 in Section 2.A. Then, at a tipping point of the infected timber growth rate (in $0.4 < \varepsilon < 0.5$), the optimal thinned proportion and rotation

length increase with further reduction in $\varepsilon < 0.4$. The increase is more significant for larger transmission rate values and less prominent for the optimal rotation length. Before this switch, little timber is infected by the thinning time, and far more is infected by rotation time. This is like the "best of both worlds" strategy, seen in the $\beta - \rho$ parameter space (Section 2.4.2 — thinning earlier to salvage before infection, rotating later after infection arrives). The switch highlights how infection spreading slows the overall forest growth and changes the growth dynamics linked to density dependence.

Furthermore, the maximum NPV's are higher in the $\beta - \varepsilon$ parameter space (Figure 2.A.6 in Section 2.A) than in the $\beta - \rho$ parameter space (Figure 2.A.4 in Section 2.A). This is because the forest's density-dependent growth allows susceptible timber volume to grow in place of infected volume and infected timber generates revenue ($\rho = 1$). Also, the transitions between strategies are much smoother in the $\beta - \varepsilon$ parameter space compared to the $\beta - \rho$ parameter space. This is related to the objective function's (Eq. (2.4)) linearity in ρ but not in ε.

2.5 Discussion

In this paper, we developed a bioeconomic model to determine economically optimal harvesting regimes – in terms of thinning and rotation – of an even-aged plantation under the risk of an invading pest. Using a combination of analytic results and sensitivity analysis, we show that the presence of disease effectively

adds to the discount rate in terms of the optimal harvest times (thinning and rotation). However, the complete optimal strategy in the thinning and rotation regime is highly responsive to the anticipated disease characteristics; the transmission rate, the severity of damage caused, the impact on growth and the effect of thinning on disease transmission. The optimal thinning time in our model is when the increase in the thinned timber benefit from an additional year of growth equals the discount rate, plus the loss rate from the spread of infection, minus the discounted change in timber benefit at rotation relative to the thinned timber benefit. Therefore, according to our model, commercial forest managers must decide when to thin to balance harvesting before infection destroys the timber's value, reducing secondary infection and exploiting their forest's density-dependent growth to cultivate target harvests.

Thinning can be used to massively improve the forest's NPV if applied correctly in the presence of disease. We find that adding thinning into the harvesting regime is always optimal, regardless of the disease levels. Timber growth in our model is density-dependent. Thinning, reducing the density and freeing up growing space, exploits this feature to increase the timber benefits produced over the rotation. Lowering the density also has the added benefit of reducing secondary infection. Moreover, when thinning reduces the transmission rate further or presents an opportunity to harvest before a large proportion of the forest is infected, the NPV is improved even further.

Similarly to Macpherson et al. [28], our model suggests it can be optimal to follow the disease-free thinning and rotation times even when disease prevalence is high. When managers expect infection to cause very little damage to timber value or have a small impact on timber growth, they should continue to thin and rotate late. Additionally, we find that the difference in NPV between the harvesting regime with thinning and one without is largest when long rotations are optimal. Long rotations provide more time to exploit the increased forest growth from thinning.

At a unique balance of middling disease transmission rate and severity of damage from infection, it is optimal to thin early to harvest timber before infection arrives, and leave the remaining trees to become infected and harvested at the disease-free rotation time. Furthermore, we find that the NPV of this strategy is significantly higher than in one without thinning — a rotation only regime can only cut early or late. However, this strategy exists within a narrow parameter space, and more work needs to be done to explore the effect of other parameters (e.g., the discount rate, primary infection, price of thinned timber) on it.

When little revenue is salvageable from infected timber (or infection severely impacts timber growth) and thinning has little impact on the transmission rate, shortening the rotation length from the disease-free length is optimal and has a massive benefit to NPV. Managers should thin and rotate early before infection destroys the forest's value, and the quicker the disease spreads, the sooner they

should act. The higher NPV benefit of including thinning in these shorter rotations is small. This finding agrees with an established result in the literature — increased risk of a catastrophic timber loss from natural disasters (storms, fires, severe pest outbreaks) decreases the optimal rotation length [27, 28, 33, 35]. In particular, we confirm that the similar result of Macpherson et al. [28] holds when (i) thinning is added into a rotation only regime, and (ii) the timber production function is density- and age-dependent. Furthermore, the finding agrees with previous studies that integrated thinning and catastrophic natural risks into Faustmann models [33, 35]. Staupendahl and Möhring showed that late risks, ones that increase over time, shorten rotations. In our approach, the rate of disease spread increases with age and timber density (which also increases with age) and could be viewed as a late risk. Furthermore, Halbritter et al. [33] demonstrated that as the expected damage from a catastrophic event increases, optimal rotation lengths decrease.

However, if thinning is expected to reduce the transmission rate significantly, the priority shifts to protect the final harvest. Managers should thin even earlier – slowing the spread of disease – then let the remaining trees grow undisturbed, harvesting them closer to the disease-free rotation time. In this scenario, the NPV of the forest increases significantly, and thinning is the primary driver.

Additionally, including thinning always extends the optimal rotation length compared to that of a rotation-only regime. The optimal rotation length is when the

rate of increase in the forest's clear-felled timber value from an additional year of growth equals the discount rate plus the loss in clear-felled timber value from the spread of infection. Including thinning effectively reduces the discount rate because it increases the growth rate of the forest and reduces secondary infection. This result of thinning extending rotation length agrees with some approaches in the literature that built on the Faustmann Model to integrate disease and thinning [42], but not with others [33, 34]. In Halbritter et al.'s approach [33] thinning does not affect the first-order condition of the optimal rotation length. Petucco and Andrés-Domenech [34] show that including thinning causes a decrease in the optimal rotation length when considering the impact of disease.

In our model, considering disease reduces the optimal rotation length compared to a disease-free scenario. Similarly to Macpherson et al. [28], the presence of disease effectively adds to the discount rate and so the disease-free rotations are the upper bounds. Petucco and Andrés-Domenech [34] found the opposite result: increased prevalence of a defoliator pest that slowed tree growth increased the optimal rotation length past the disease-free one. However, when we investigate the sensitivity of our results to the infected timber growth rate without an effect of thinning on transmission (Section 2.4.2), we do find a similar result to theirs. We showed that decreasing the growth rate of infected timber would initially decrease the optimal rotation length when the disease spreads quickly. Then when the impact on growth was particularly severe, the pattern switched and decreasing the growth rate of infected timber further would increase the optimal

rotation length towards the disease-free length. This finding highlights a delicate relationship in our model: infection spreading slows the overall forest growth and changes the growth dynamics linked to density dependence. It also highlights the sensitivity of our results to the growth functions used. Furthermore, when the disease has a low impact on timber value, it can be optimal to use a longer rotation length when the spread rate is high compared to if it was low. Therefore, slight increases in the severity of disease impact do not always add to the interest rate in our approach.

Trees in our investigation are felled indiscriminately of their infection state during thinning. A reasonable extension would be to increase the heterogeneity/complexity of the thinning regime. One modelling approach is to let forest managers bias thinning towards the infected trees. These eradication strategies exist for destructive pathogens such as *Phytophthora ramorum* in the UK, where all trees within a radius around the detected infected ones are felled [43]. Another approach is to increase the number of thinning operations during the rotation, even to annual thinning. We also do not consider any unfavourable or unintended impacts of thinning on forest diseases. When trees infected with *Heterobasidian Annosum*, a fungal pathogen that rots trees, are cut down, this exposes the stump, releasing spores and exacerbating the spread [44]. In a thinning and rotation regime, as the extent of this exacerbation increases, we could see optimal thinning intensity switch from being nonzero to zero. An understanding of the impact on the entire optimal strategy requires further study.

After thinning, the total timber volume will almost always recover to the carrying capacity (the maximum volume), regardless of the intensity in our model. This may not be a realistic picture, and the link between the volume of individual trees with the volume of the forest is unclear. Our analysis also assumed equal prices for thinned and clear-felled timber, with a fixed price per m^3 for timber. However, as trees grow, their diameters increase alongside their height and volume, and the value of timber grows over time [45]. For example, thinned timber often produces narrow stems sold as wood fuel or firewood. Whereas timber felled later is more mature with a broader set of merchantable applications. We also neglect extraction costs, which could be higher for thinning operations than clear-felling due to economies of scale. As a result, we may have overestimated the net income from thinning, and in turn, the optimal thinning intensities. Having price endogenous to our model, and reflecting individual tree growth, would be an appropriate extension.

We assumed no further planting or harvesting after the single rotation. The main reason behind this is the irreversible nature of tree pests and diseases. A model of multiple rotations would have to incorporate an assumption about what happens to the level of infection between rotations (i.e., if and how the pest/pathogen carries over to the next rotation) [28]. There is significant variation and uncertainty in the ability of pests and diseases to persist after clear-felling [2]. Therefore, any assumption we introduce would be highly context and pest/pathogen-specific,

making it difficult to draw general conclusions on harvesting strategies. Furthermore, forest managers may deploy different planting schemes (patterns or species) for the next rotation to reduce further disease impact [2]. For example, the arrival of Ash Dieback led to a complete ban on the movement and importation of Ash in the UK [46] . If we introduced an assumption where no disease remained after rotation, this could encourage shorter rotation periods, as fresh timber growing without infection is more valuable. However, this would depend on the rates of forest growth and disease progression. Any leftover disease could increase the proportion of timber infected in future rotations and shorten or increase optimal rotation periods. The outcome would depend on the balance between damage caused by infection and income produced over each rotation. After the stand is clear-felled, we assume it lays bare. Changing the land use after rotation to provide a new source of income could be a reasonable strategy for a forest manager. As shown by Macpherson et al. [28], including an annual land rent can implicitly capture this opportunity. They show that when the potential income from felling and receiving annual land rent surpasses additional income from leaving the stand to grow, rotation periods decrease.

A forest owner may wish to consider non-timber benefits such as carbon sequestration or recreation in their management strategy. While we have not included these in our model, the work of Macpherson et al. [29] provides insight into their potential impact on our results. In their paper, non-timber benefits are internalised into a single rotation Faustmann model with disease risk using a green

payment. The green payment counteracts the negative economic effect of disease and incentivises leaving timber unharvested and increasing the optimal rotation length. This effect depends on whether the disease impacts only timber benefits or both timber and non-timber benefits. If non-timber benefits are unaffected, forest owners can be incentivised to never clear-fell their forest. We expect to find similar effects by including non-timber benefits in our model, dependent on how they are generated (e.g., through age, biomass, or forested area). Furthermore, when disease impacts the non-timber benefit and thinning controls disease spread, there may be an additional trade-off with the incentive to thin to protect the non-timber benefits.

2.6 Conclusion

In this paper, we developed a theoretical and generalisable bio-economic model to determine optimal harvesting strategies under a pathogen or pest invasion. To find the optimal strategy, we maximise the return on investment for a commercial forest manager while accounting for the anticipated interactions between thinning, tree growth and disease progression. Furthermore, we analysed various harvesting regimes through a sensitivity analysis of variable disease conditions. The return on investment for the forest manager is highly sensitive to the type of harvesting strategy employed and the disease characteristics. We investigated the

role of thinning in these harvesting strategies and highlighted when its inclusion is vital for forest managers to consider. Our study provides a framework that can help design appropriate forest management strategies in the presence of disease.

Acknowledgements

EM Ph.D. project is funded by the Student Excellence Award, University of Strathclyde. AK contribution was funded by the Scottish Government Strategic Research Programme, Theme 2 project, Disease management options: Insights from comparing forestry and agriculture. EM and AK also acknowledge funding from the NERC project NE/V019988/1, Learning to adapt to an uncertain future: linking genes, trees, people and processes for more resilient treescapes (newLEAF).

Appendix

2.A Heatmaps showing optimal strategies

See the Figures below.

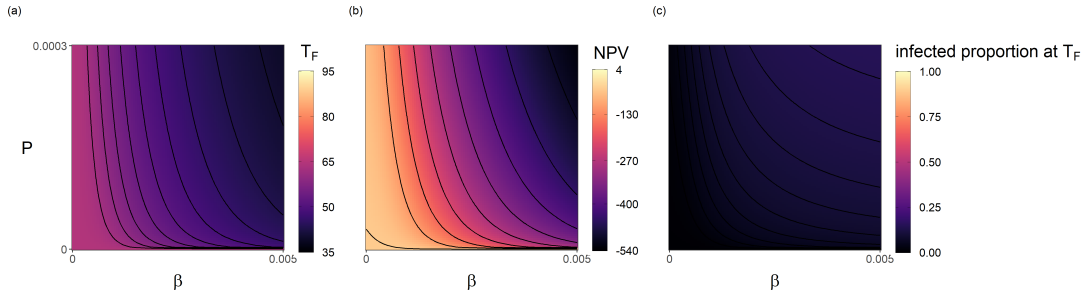

Fig. 2.A.1: Optimal strategy for the rotation only regime in a P-β parameter space when there is no revenue from infected timber ($\rho = 0$). The primary infection rate is P, and the transmission rate is β. The optimal strategy for the rotation only regime is found by solving Eq. (2.8). All other parameter values are given in Table 2.2. (a) Optimal proportion to thin, γ; (b) Optimal time to thin as a fraction of the rotation length, T_1/T_F; (c) fraction of timber that is infected at the optimal thinning time, $\frac{y(T_1)}{x(T_1)+y(T_1)}$; (d) Optimal rotation length, T_F; (e) NPV under the optimised strategy, $\hat{J}$, given by Eq. (2.4); (f) fraction of timber that is infected at the rotation time, $\frac{y(T_F)}{x(T_F)+y(T_F)}$.

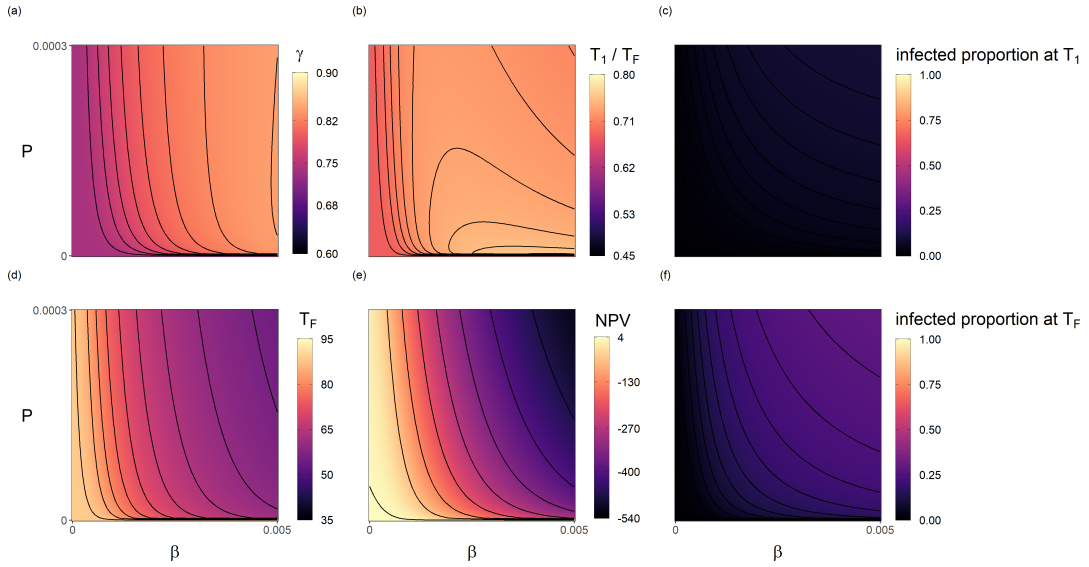

Fig. 2.A.2: Optimal strategy for the thinning and rotation regime in a P-β parameter space when there is no revenue from infected timber ($\rho = 0$). The primary infection rate is P, and the transmission rate is β. The optimal strategy for the thinning and rotation regime is found by solving Eq. (2.7). All other parameter values are given in Table 2.2. (a) Optimal proportion to thin, γ; (b) Optimal time to thin as a fraction of the rotation length, T_1/T_F; (c) fraction of timber that is infected at the optimal thinning time, $\frac{y(T_1)}{x(T_1)+y(T_1)}$; (d) Optimal rotation length, T_F; (e) NPV under the optimised strategy, $\hat{J}$, given by Eq. (2.4); (f) fraction of timber that is infected at the rotation time, $\frac{y(T_F)}{x(T_F)+y(T_F)}$.

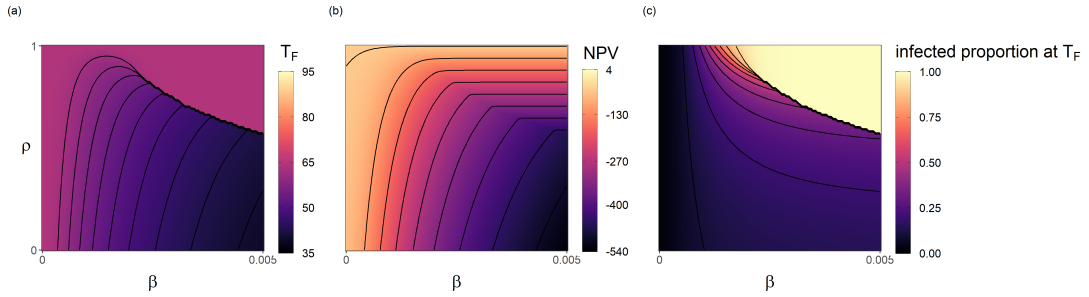

Fig. 2.A.3: Optimal strategy for the rotation only regime in a ρ-β parameter space. The revenue from a unit of infected timber relative to susceptible is ρ, and the transmission rate is β. The optimal strategy for the rotation only regime is found by solving Eq. (2.8). All other parameter values are given in Table 2.2. (a) Optimal rotation length, T_F; (b) NPV under the optimised strategy, $\hat{J}$, given by Eq. (2.4); (c) fraction of timber that is infected at the rotation time, $\frac{y(T_F)}{x(T_F)+y(T_F)}$.

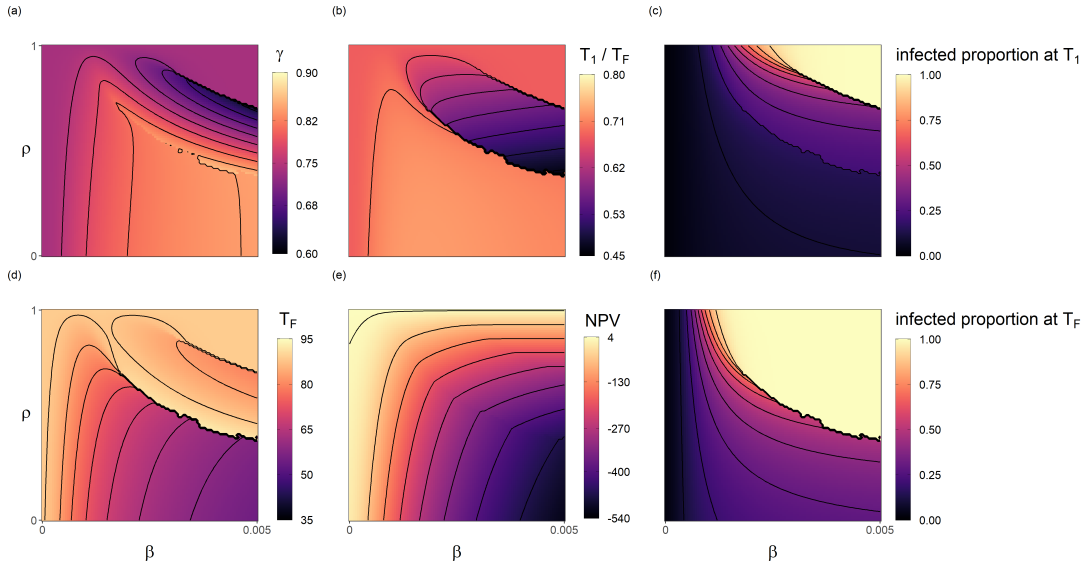

Fig. 2.A.4: Optimal strategy for the thinning and rotation regime in a ρ-β parameter space. The revenue from a unit of infected timber relative to susceptible is ρ, and the transmission rate is β. The optimal strategy for the thinning and rotation regime is found by solving Eq. (2.7). All other parameter values are given in Table 2.2. (a) Optimal proportion to thin, γ; (b) Optimal time to thin as a fraction of the rotation length, T_1/T_F; (c) fraction of timber that is infected at the optimal thinning time, $\frac{y(T_1)}{x(T_1)+y(T_1)}$; (d) Optimal rotation length, T_F; (e) NPV under the optimised strategy, $\hat{J}$, given by Eq. (2.4); (f) fraction of timber that is infected at the rotation time, $\frac{y(T_F)}{x(T_F)+y(T_F)}$.

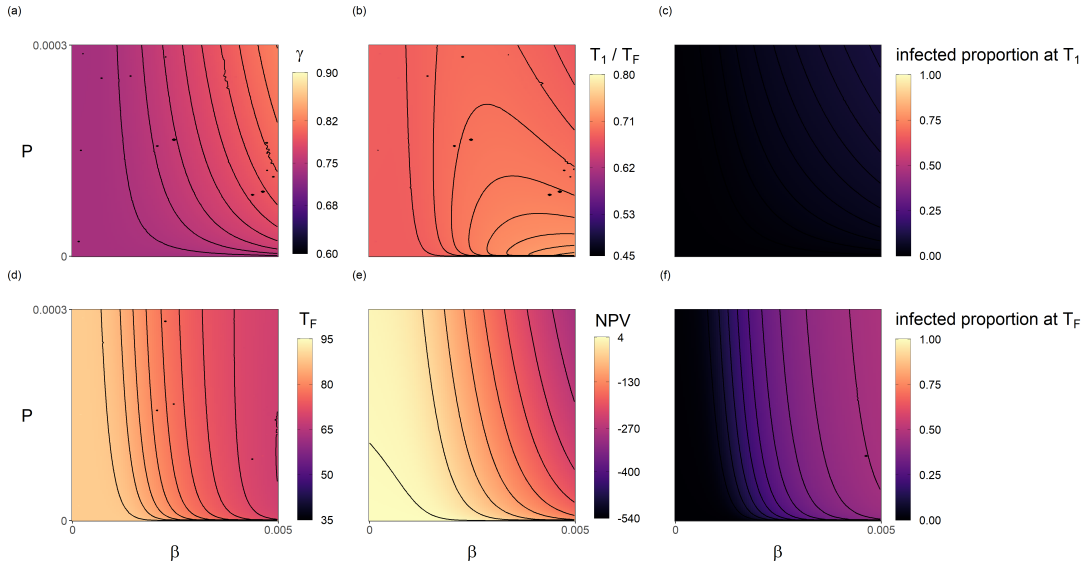

Fig. 2.A.5: Optimal strategy for the thinning and rotation regime in a P-β parameter space when infected timber does not grow ($\varepsilon = 0$). The primary infection rate is P, and the transmission rate is β. The optimal strategy for the thinning and rotation regime is found by solving by Eq. (2.7). Revenue from infected timber is the same value as from susceptible ($\rho = 1$) and all other parameter values are given in Table 2.2. (a) Optimal rotation length, T_F; (b) NPV under the optimised strategy, $\hat{J}$, given by Eq. (2.4); (c) fraction of timber that is infected at the rotation time, $\frac{y(T_F)}{x(T_F)+y(T_F)}$.

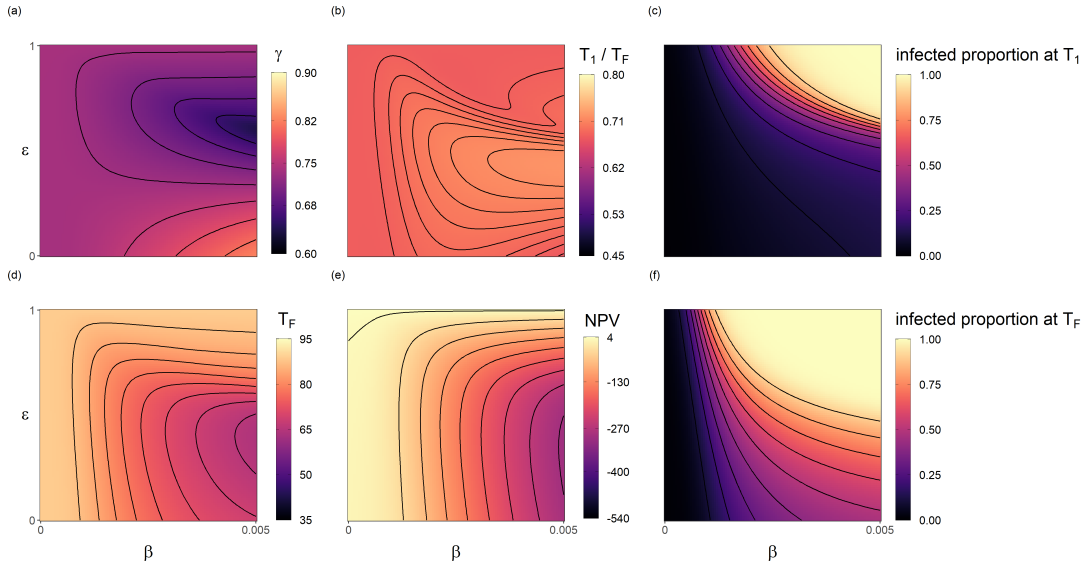

Fig. 2.A.6: Optimal strategy for the thinning and rotation regime in a β-ε parameter space. The transmission rate is β, and growth of infected timber relative to susceptible is ε. The optimal strategy for the thinning and rotation regime is found by solving Eq. (2.7). All other parameter values are given in Table 2.2. (a) Optimal proportion to thin, γ; (b) Optimal time to thin as a fraction of the rotation length, T_1/T_F; (c) fraction of timber that is infected at the optimal thinning time, $\frac{y(T_1)}{x(T_1)+y(T_1)}$; (d) Optimal rotation length, T_F; (e) NPV under the optimised strategy, $\hat{J}$, given by Eq. (2.4); (f) fraction of timber that is infected at the rotation time, $\frac{y(T_F)}{x(T_F)+y(T_F)}$.

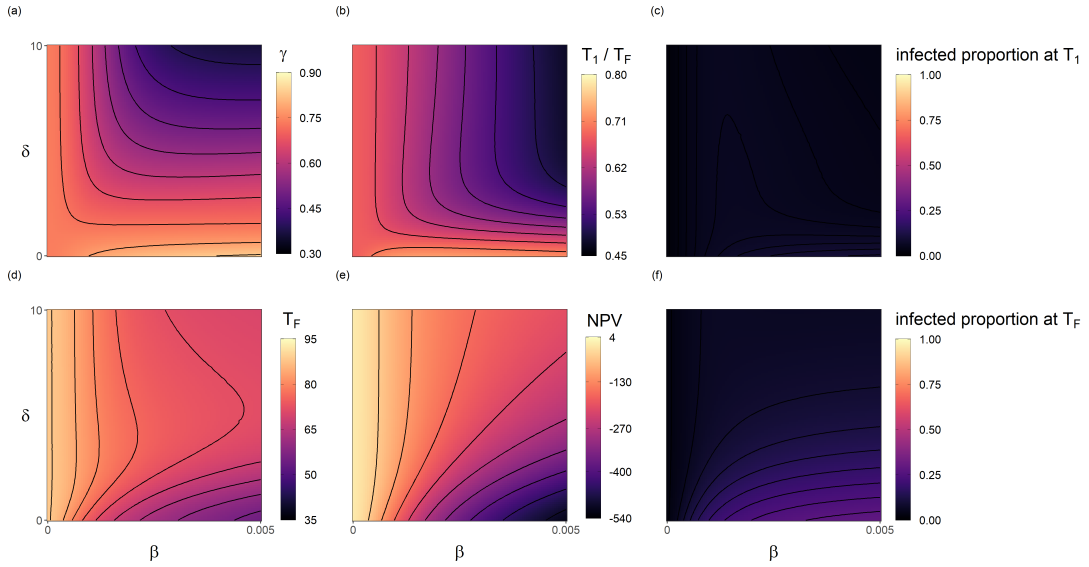

Fig. 2.A.7: Optimal strategy for the thinning and rotation regime in a β-δ parameter space when there is no revenue from infected timber ($\rho = 0$). The transmission rate is β, and δ is the impact of thinning on the transmission rate. The optimal strategy for the thinning and rotation regime is found by solving Eq. (2.7). All other parameter values are given in Table 2.2. (a) Optimal proportion to thin, γ; (b) Optimal time to thin as a fraction of the rotation length, T_1/T_F; (c) fraction of timber that is infected at the optimal thinning time, $\frac{y(T_1)}{x(T_1)+y(T_1)}$; (d) Optimal rotation length, T_F; (e) NPV under the optimised strategy, $\hat{J}$, given by Eq. (2.4); (f) fraction of timber that is infected at the rotation time, $\frac{y(T_F)}{x(T_F)+y(T_F)}$.

2.B Calculations for the optimal rotation length when there is no disease

When there is no disease, forest dynamics are governed by a single equation,

$$x'(t) = g(t)x(1 - \frac{x}{K}) - h(t) \tag{2.17}$$

In this section, we will show that

$$x_{\text{no thin}}(T_F^*) \geq x_{\text{thin}}(T_F^*) \tag{2.18}$$

where T_F^* is the optimal rotation length when there is no disease for a regime without thinning, and $x_{\text{thin}}(t)$ and $x_{\text{no thin}}(t)$ are the timber volumes with and without thinning.

Eq. (2.17) can be solved analytically using the separation of variables method. With no thinning ($h(t) = 0 \ \forall \ t$ as $\gamma = 0$) we have

$$x_{\text{no thin}}(t) = \frac{K}{\left(\frac{K}{x_0} - 1\right) e^{-\int_0^t g(s)ds} + 1} \tag{2.19}$$

and when there is thinning

$$x_{\text{thin}}(t) = \begin{cases} x_{\text{no thin}}(t), & 0 \leq t \leq T_1 \text{ (pre-thinning years)} \\ \dfrac{K}{\left(\frac{K}{(1-\gamma)x_{\text{no thin}}(T_1)} - 1\right) e^{-\int_{T_1}^t g(s)ds} + 1}, & t > T_1 \text{ (post-thinning years)} \end{cases} \tag{2.20}$$

When $T_F^* \leq T_1$, after substitution of Eq. (2.19) and Eq. (2.20) into Eq. (2.18), Eq. (2.18) clearly holds.

When $T_F^* > T_1$, by substitution of $t = T_F^{\text{thin}*}$ into Eq. (2.19) and Eq. (2.20) we have

$$x_{\text{thin}}(T_F^*) = \frac{K}{\left(\frac{K}{(1-\gamma)x_{\text{no thin}}(T_1)} - 1\right) e^{-\int_{T_1}^{T_F^*} g(s)ds} + 1} \tag{2.21}$$

and

$$x_{\text{no thin}}(T_F^*) = \frac{K}{\left(\frac{K}{x_0} - 1\right) e^{-\int_0^{T_F^*} g(s)ds} + 1}$$
$$= \frac{K}{\left(\frac{K}{x_{\text{no thin}}(T_1)} - 1\right) e^{-\int_{T_1}^{T_F^*} g(s)ds} + 1}, \tag{2.22}$$

and as the RHS of Eq. (2.21) is a decreasing function of $\gamma \in (0, 1)$ we have Eq. (2.22) $\geq$ Eq. (2.21), i.e., Eq. (2.18) holds.

Bibliography

[1] E. McTaggart, I. Megiddo, and A. Kleczkowski, "The effect of pests and pathogens on forest harvesting regimes:A bioeconomic model," *Ecological Economics*, vol. 209, no. March, p. 107800, 2023. [Online]. Available: https://doi.org/10.1016/j.ecolecon.2023.107800

[2] M. Roberts, C. A. Gilligan, A. Kleczkowski, N. Hanley, A. E. Whalley, and J. R. Healey, "The effect of forest management options on forest resilience to pathogens," *Front. For. Glob. Chang.*, vol. 3, 2020. [Online]. Available: https://doi.org/10.3389/ffgc.2020.00007

[3] I. L. Boyd, P. H. Freer-Smith, C. A. Gilligan, and H. C. J. Godfray, "The consequence of tree pests and diseases for ecosystem services," *Science*, vol. 342, no. 6160, 2013. [Online]. Available: https://doi.org/10.1126/science.1235773

[4] T. D. Ramsfield, B. J. Bentz, M. Faccoli, H. Jactel, and E. G. Brockerhoff, "Forest health in a changing world: Effects of globalization and climate change on forest insect and pathogen impacts," *Forestry*, vol. 89, no. 3, pp. 245–252, 2016. [Online]. Available: https://doi.org/10.1093/forestry/cpw018

[5] R. N. Sturrock, S. J. Frankel, A. V. Brown, P. E. Hennon, J. T. Kliejunas, K. J. Lewis, J. J. Worrall, and A. J. Woods, "Climate change and forest diseases," *Plant Pathol.*, vol. 60, no. 1, pp. 133–149, 2011. [Online]. Available: https://doi.org/10.1111/j.1365-3059.2010.02406.x

[6] M. Marzano, L. Fuller, and C. P. Quine, "Barriers to management of tree diseases: Framing perspectives of pinewood managers around Dothistroma Needle Blight," *Journal of Environmental Management*, vol. 188, pp. 238–245, 2017. [Online]. Available: https://doi.org/10.1016/J.JENVMAN.2016.12.002

[7] Forestry Commission, "Minimising the impact of the great spruce bark beetle," Forestry Commission, Tech. Rep., 2012. [Online]. Available: https://www.forestresearch.gov.uk/publications/minimising-the-impact-of-the-great-spruce-bark-beetle/

[8] P. H. Freer-Smith and J. F. Webber, "Tree pests and diseases: The threat to biodiversity and the delivery of ecosystem services," *Biodivers. Conserv.*, vol. 26, pp. 3167–3181, 2017. [Online]. Available: https://doi.org/10.1007/s10531-015-1019-0

[9] Office for National Statistics, "Woodland natural capital accounts ecosystem services for England, Scotland, Wales and Northern Ireland, 2020," Office for National Statistics, Tech. Rep., 2020. [Online]. Available: https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/woodlandnaturalcapitalaccountsuk/ecosystemservicesforenglandscotlandwalesandnorthernireland2020#woodland-natural-capital-accounts-ecosystem-services-data

[10] M. J. Klapwijk, A. J. M. Hopkins, L. Eriksson, M. Pettersson, M. Schroeder, Å. Lindelöw, J. Rönnberg, E. C. H. Keskitalo, and M. Kenis, "Reducing the risk of invasive forest pests and pathogens: Combining legislation, targeted management and public awareness," *Ambio*, vol. 45, pp. 223–234, 2016. [Online]. Available: https://doi.org/10.1007/s13280-015-0748-3

[11] C. Potter and J. Urquhart, "Tree disease and pest epidemics in the anthropocene: A review of the drivers, impacts and policy responses in the UK," *Forest Policy Econ.*, vol. 79, pp. 61–68, 2017. [Online]. Available: https://doi.org/10.1016/J.FORPOL.2016.06.024

[12] G. Kerr and J. Haufe, "Thinning practice: A silvicultural guide," Forestry Commission, Tech. Rep., 2011. [Online]. Available: https://www.forestresearch.gov.uk/publications/thinning-practice-a-silvicultural-guide/

[13] R. W. Matthews, T. A. R. Jenkins, E. D. Mackie, and E. C. Dick, *Forest Yield: A Handbook on Forest Growth and Yield Tables for British Forestry.* Forestry Commission, 2016. [Online]. Available: https://www.forestresearch.gov.uk/publications/forest-yield-a-handbook-on-forest-growth-and-yield-tables-for-british-forestry/

[14] Forestry Commission, "Thinning control field guide," Forestry Comission, Tech. Rep., 2015. [Online]. Available: https://www.forestresearch.gov.uk/publications/thinning-control-2015/

[15] B. Zeide, "Thinning and growth: A full turnaround," *Journal of Forestry*, vol. 99, no. 1, pp. 20–25, 2001. [Online]. Available: https://doi.org/10.1093/jof/99.1.20

[16] J. R. Boyle, J. C. Tappeiner, R. H. Waring, and C. Tattersall Smith, *Sustainable Forestry: Ecology and Silviculture for Resilient Forests.* Elsevier, 2016. [Online]. Available: https://doi.org/10.1016/b978-0-12-409548-9.09761-x

[17] L. S. Bulman, R. E. Bradshaw, S. Fraser, J. Martín-García, I. Barnes, D. L. Musolin, N. La Porta, A. J. Woods, J. J. Diez, A. Koltay, R. Drenkhan, R. Ahumada, L. Poljakovic-Pajnik, V. Queloz, B. Piškur, H. T. Doğmuş-Lehtijärvi, D. Chira, V. Tomešová-Haataja, M. Georgieva, L. Jankovský, N. Anselmi, S. Markovskaja, I. Papazova-Anakieva, K. Sotirovski, J. Lazarević, K. Adamčíková, P. Boroń, H. Bragança, A. M. Vettraino, A. V. Selikhovkin, T. S. Bulgakov, and K. Tubby, "A worldwide

perspective on the management and control of dothistroma needle blight," *Forest Pathol.*, vol. 46, no. 5, pp. 472–488, 2016. [Online]. Available: http://doi.wiley.com/10.1111/efp.12305

[18] C. Quine, M. Marzano, L. Fuller, N. Dandy, J. Barnett, G. Brandon, G. Jones, E. Porth, and C. Price, "Social and economic analyses of dothistroma needle blight management," Forest Research, Tech. Rep., 2014. [Online]. Available: https://www.forestresearch.gov.uk/publications/social-and-economic-analyses-of-dothistroma-needle-blight-management/

[19] Forest Research, "Dothistroma needle blight (*Dothistroma septosporum*)," Forest Research, Tech. Rep., 2022. [Online]. Available: https://www.forestresearch.gov.uk/tools-and-resources/fthr/pest-and-disease-resources/dothistroma-needle-blight/

[20] É. Bauce and A. Fuentealba, "Interactions between stand thinning, site quality and host tree species on spruce budworm biological performance and host tree resistance over a 6 year period after thinning," *Forest Ecology and Management*, vol. 304, pp. 212–223, 2013. [Online]. Available: https://doi.org/10.1016/J.FORECO.2013.05.008

[21] T. Hlásny, P. Krokene, A. Liebhold, C. Montagné-Huck, J. Müller, H. Qin, K. Raffa, M. -J. Schelhaas, R. Seidl, and M. Svoboda, *Living with bark beetles: Impacts, outlook and management options.* European Forest Institute, 2019. [Online]. Available: https://doi.org/10.36333/fs08

[22] P. A. Samuelson, "Economics of forestry in an evolving society," *Economic Enquiry*, vol. 14, no. 4, 1976. [Online]. Available: https://doi.org/10.1111/j.1465-7295.1976.tb00437.x

[23] R. Coordes, *Optimal Thinning Within the Faustmann Approach*. Springer, 2014. [Online]. Available: https://doi.org/10.1007/978-3-658-06959-9

[24] C. W. Clark and J. D. De Pree, "Applied mathematics and optimization a simple linear model for the optimal exploitation of renewable resources," *Appl. Math. Optim.*, vol. 5, pp. 181–196, 1979. [Online]. Available: https://link.springer.com/article/10.1007/BF01442553

[25] A. Halbritter and P. Deegen, "A combined economic analysis of optimal planting density, thinning and rotation for an even-aged forest stand," *Forest Policy Econ.*, vol. 51, pp. 38–46, 2015. [Online]. Available: https://doi.org/10.1016/j.forpol.2014.10.006

[26] O. Tahvonen, "Economics of rotation and thinning revisited: The optimality of clearcuts versus continuous cover forestry," *Forest Policy Econ.*, vol. 62, pp. 88–94, 2016. [Online]. Available: https://doi.org/10.1016/j.forpol.2015.08.013

[27] W. J. Reed, "The effects of the risk of fire on the optimal rotation of a forest," *J. Environ. Econ. Manag.*, vol. 11, no. 2, pp. 180–190, 1984. [Online]. Available: https://doi.org/10.1016/0095-0696(84)90016-0

[28] M. F. Macpherson, A. Kleczkowski, J. R. Healey, and N. Hanley, "The effects of disease on optimal forest rotation: A generalisable analytical framework," *Environ. Resour. Econ.*, vol. 70, pp. 565–588, 2016. [Online]. Available: https://doi.org/10.1007/s10640-016-0077-4

[29] ——, "Payment for multiple forest benefits alters the effect of tree disease on optimal forest rotation length," *Ecological Economics*, vol. 134, pp. 82–94, 2017a. [Online]. Available: https://doi.org/10.1016/j.ecolecon.2017.01.008

[30] M. F. Macpherson, A. Kleczkowski, J. R. Healey, C. P. Quine, and N. Hanley, "The effects of invasive pests and pathogens on strategies for forest diversification," *Ecological Modelling*, vol. 350, pp. 87–99, 2017b. [Online]. Available: https://doi.org/10.1016/j.ecolmodel.2017.02.003

[31] H. An, S. Lee, and S. J. Cho, "The effects of climate change on pine wilt disease in south korea: Challenges and prospects," *Forests*, vol. 10, no. 6, p. 486, 2019. [Online]. Available: https://doi.org/10.3390/f10060486

[32] K. M. Cobourn, H. J. Burrack, R. E. Goodhue, J. C. Williams, and F. G. Zalom, "Implications of simultaneity in a physical damage function," *J. Environ. Econ. Manag.*, vol. 62, no. 2, pp. 278–289, 2011. [Online]. Available: http://doi.org/10.1016/j.jeem.2011.02.002

[33] A. Halbritter, P. Deegen, and A. Susaeta, "An economic analysis of thinnings and rotation lengths in the presence of natural risks in even-aged forest stands," *Forest Policy Econ.*, vol. 118, 2020. [Online]. Available: http://doi.org/10.1016/j.forpol.2020.102223

[34] C. Petucco and P. Andrés-Domenech, "Land expectation value and optimal rotation age of maritime pine plantations under multiple risks," *J. Forest Econ.*, vol. 30, no. 1, pp. 58–70, 2018. [Online]. Available: http://doi.org/10.1016/j.jfe.2018.01.001

[35] K. Staupendahl and B. Möhring, "Integrating natural risks into silvicultural decision models: A survival function approach," *Forest Policy Econ.*, vol. 13, no. 6, pp. 496–502, 2011. [Online]. Available: http://doi.org/10.1016/j.forpol.2011.05.007

[36] D. Braby, "Timber price indices," Forest Research, Tech. Rep., 2022. [Online]. Available: https://www.forestresearch.gov.uk/tools-and-resources/statistics/statistics-by-topic/timber-statistics/timber-price-indices/

[37] M. S. Mullett, K. V. Tubby, J. F. Webber, and A. V. Brown, "A reconsideration of natural dispersal distances of the pine pathogen *dothistroma septosporum*," *Plant Pathology*, vol. 65, no. 9, pp. 1462–1472, 2016. [Online]. Available: https://doi.org/10.1111/ppa.12522

[38] Forest Research, "Pathogens and hosts of dothistroma needle blight," Forest Research, Tech. Rep., 2022b. [Online]. Available: https://www.forestresearch.gov.uk/tools-and-resources/fthr/pest-and-disease-resources/dothistroma-needle-blight/pathogens-and-hosts-of-dothistroma-needle-blight/

[39] H. Mäkinen and A. Isomäki, "Thinning intensity and growth of norway spruce stands in finland," *Forestry*, vol. 77, no. 4, pp. 349–364, 2004. [Online]. Available: http://doi.org/10.1093/FORESTRY/77.4.349

[40] Scottish Forestry, "Dothistroma needle blight in scotland," Scottish Forestry, Tech. Rep., 2023. [Online]. Available: https://forestry.gov.scot/sustainable-forestry/tree-health/tree-pests-and-diseases/dothistroma-needle-blight

[41] A. Kleczkowski, A. Hoyle, and P. McMenemy, "One model to rule them all? Modelling approaches across OneHealth for human, animal and plant epidemics," *Philosophical Transactions of the Royal Society*, vol. 374, no. 1755, 2019. [Online]. Available: http://doi.org/10.1098/rstb.2018.0255

[42] P. Loisel, "Faustmann rotation and population dynamics in the presence of a risk of destructive events," *J. Forest Econ.*, vol. 17, no. 3, pp. 235–247, 2011. [Online]. Available: http://doi.org/10.1016/j.jfe.2011.02.001

[43] R. O'Hanlon, J. Choiseul, J. M. Brennan, and H. Grogan, "Assessment of the eradication measures applied to *Phytophthora ramorum* in irish *Larix kaempferi* forests," *Forest Pathol.*, vol. 48, no. 1, 2018. [Online]. Available: http://doi.org/10.1111/efp.12389

[44] M. Garbelotto and P. Gonthier, "Biology, epidemiology, and control of *Heterobasidion* species worldwide," *Annu. Rev. Phytopathol.*, vol. 51, pp. 39–59, 2013. [Online]. Available: https://doi.org/10.1146/annurev-phyto-082712-102225

[45] P. W. West, *Growing Plantation Forests*, 2nd ed. Springer, 2014. [Online]. Available: https://doi.org/10.1007/978-3-319-01827-0

[46] J. Clark and J. Webber, "The ash resource and the response to ash dieback in great britain," in *Dieback of European Ash (Fraxinus spp.): Consequences and Guidelines for Sustainable Management*, R. Vasaitis and R. Enderle, Eds., 01 2017, pp. 228–237. [Online]. Available: https://livingashproject.org.uk/useful-information/

Chapter 3

COVID-19 in Scottish care homes: A metapopulation model of spread among residents and staff

This chapter is a manuscript that was produced in conjoint with Matthew Baister (MB), Paul McMenemy (PM), Itamar Megiddo (IM) and Adam Kleczkowski (AK). The work took place between May 2020 and July 2021, and we developed a mathematical model to investigate the spread of COVID-19 within care homes in Scotland, focusing on residents and staff members. Plos Computational Biology and Plos One rejected this paper; therefore, it has recently been rewritten and resubmitted to the Epidemics journal (Elsevier). Each author's contributions are outlined briefly below and shown in Table 3.1.

AK, IM, and PM played supervisory roles and provided feedback on the manuscript.

PM wrote a foundational source code to simulate the model, which the PhD author (EM) and MB further developed. The PhD author (EM) worked in tandem with MB on; writing and editing the manuscript, developing the model equations, and on formal model analysis and investigation that further developed the code (parameterisation, sensitivity analysis, and data visualisation). The PhD author also conducted the literature review and constructed the model schematics. MB collected data sources for parameter estimates.

Table 3.1: Description of authors' contributions to the manuscript.

Contributor Role	**Role Definition**	**Name**
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	AK
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	MB, EM
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	MB, EM
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	AK
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	MB, EM
Methodology	Development or design of methodology; creation of models	AK, PM, MB, EM, IM
Project Administration	Management and coordination responsibility for the research activity planning and execution.	AK, PM, IM, MB, EM
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	AK, PM, MB, EM
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	AK, PM, MB, EM
Supervision	Oversight and leadership responsibility for the research activity planning and, execution including mentorship external to the core team.	AK, PM, IM
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.	MB, EM
Visualisation	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	MB, EM
Writing- Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	MB, EM
Writing- Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.	AK, PM, IM, MB, EM

3.1 Abstract

Care homes in the UK were disproportionately affected by the first wave of the COVID-19 pandemic, accounting for almost half of COVID-19 deaths during the period of 6^{th} March – 15^{th} June 2020. Understanding how infectious diseases establish themselves throughout vulnerable communities is crucial for minimising deaths and lowering the total stress on the National Health Service (NHS Scotland). We model the spread of COVID-19 in the health board of NHS Lothian, Scotland over the course of the first wave of the pandemic with a compartmental Susceptible - Exposed - Infected reported - Infected unreported - Recovered - Dead (**SEIARD**), metapopulation model. Care home residents, care home workers and the rest of the population are modelled as subpopulations, interacting on a network describing their mixing habits. We explicitly model the outbreak's reproduction rate and care home visitation level over time for each subpopulation, and execute a data fit and sensitivity analysis, focusing on parameters responsible for inter-subpopulation mixing: staff-sharing, staff shift patterns and visitation. The results from our sensitivity analysis show the main driver of infection in care homes is from staff, with visitation (before cancellation) and staff-sharing less significant in affecting outbreak size. Our findings indicate that protecting care home staff from

the disease, coupled with reductions in staff-sharing across care homes and expedient cancellations of visitations, can significantly reduce the outbreak size in care homes.

3.2 Introduction

The outbreak of the SARS-CoV-2 induced disease (COVID-19) pandemic has had a profound impact, causing 3.7 million deaths by early June 2021 and global economic shocks [1]. In the UK, the care home population suffered a disproportionate amount of COVID-19 related deaths. From the week ending 13$^{\text{th}}$ March 2020 to the week ending 26$^{\text{th}}$ June 2020 (the "first wave"), 54,510 deaths were associated with COVID-19 in the UK, 40% of which were among care home residents [2]. The COVID-19 pandemic has highlighted the vulnerability of care homes to epidemics, as their resident population is elderly and often suffers from several co-morbidities [3], their systems have not been developed with infection prevention and control (IPC) in mind, and their IPC guidelines have been borrowed from hospitals - a completely different setting [4].

Networks of care homes are ecosystems connected by staff working across facilities, and these connections can increase the risk of COVID-19 ingress into care homes, and to protect their vulnerable community, we need to understand the ecosystem dynamics. We find it natural to describe this using a heterogeneous patch size metapopulation model framework.

Very few models explore COVID-19 transmission at a community level and explicitly include the unique dynamics in care homes. For example, in [5, 6] agent-based models (ABMs) of single homes are used to investigate the impact of testing strategies on the disease burden. A report by Nguyen et al. [7] uses an ABM to investigate the impact on care home residents of various vaccine coverage, and reducing the weekly testing of staff. However, the models in [6, 7] do have an external force of infection (FOI) from the community, based on prevalence data, representing staff interaction with the community and visitors. These models [5–7] do not assess the relative impact of the different COVID-19 pathways into care homes. Nguyen et al. [8] extend [6, 7], using a hybrid ABM-System Dynamics model, to explore the conditions under which visitation, heterogeneous care homes sizes, and the cohorting of residents impacts COVID-19 outbreak severity.

Rosello et al. [9] model an individual care home with a stochastic compartmental model, using multiple forces of infection to capture COVID-19 pathways, including visitors, hospital discharges, staff working at other homes, and staff infections from the community. They find that importations of infections by staff from the community are the main driver of outbreaks, and importation by visitors or from hospitals is rare, but do not explicitly model disease spread throughout a network of care homes. In [10] individual care homes and the general public are independent, deterministic SEIR models, with a stochastic external FOI connecting the general public to each home. This FOI depends on the prevalence of COVID-19 in the general public, and the size of each home. Transmission rates in homes and in the general public do not vary over time. In [11], two

weakly-coupled SEIR sub-models with time-dependent transmission rates define the dynamics; one sub-model describes the general public and one describes all care home residents in Stockholm as a single homogeneous group. Again, a single FOI acts on the residents to capture infections from staff and visitors. These models [10, 11] do not differentiate between, and therefore allow comparison of, the COVID-19 pathways into care homes. Bunnik et al. [12] use a compartmental metapopulation model to explore the trade-offs between increasing protection for a "vulnerable" population and relaxing restrictions for the "non-vulnerable" after the first lockdown in Scotland. They use time-dependent transmission rates with three metapopulation groups; vulnerable, shielders and general public. We extend and apply the methodology of [12] in our model, investigating protection to a vulnerable group (care home residents) in ways other than shielding.

We construct a **SEIARD** compartmental metapopulation model to describe the first wave of COVID-19 in a regional health board in Scotland. The population is divided into groups of care home residents, staff, and general public. Our care home resident group are not a single homogeneous unit as in [11, 12] but are separate units, creating a refined spatial/geographic structure. These units are not independent as in [10] but are linked by a staff-sharing network which, to our knowledge, is unique. We calibrate this model to 2020 data from the NHS Lothian Health Board and explore the sensitivity of the results to changes in key parameters. We investigate the importation of infections by staff from the community, visitation, staff-sharing, and additionally, we shed light on the exposure of care homes at the beginning of the first wave, e.g., via hospital discharges [13].

Our aim is to assess the impact of inter-subpopulation mixing on the spread of COVID-19 into and throughout the susceptible care home community in order to identify potential mitigation strategies to minimise the impact of future outbreaks. Our model enables this investigation by coupling the general public and individual care homes with the explicit movement of staff and visitors between the two populations.

3.3 Materials and methods

3.3.1 Mathematical model

Care homes and their residents are enclosed societies, isolated to some extent from the general population. Their connection to broader society primarily consists of interaction with staff and visits from the general population. Care home staff can potentially played a vital role in COVID-19 introduction and spread throughout the care home population. Firstly, staff exposure to infection from the general population can establish an outbreak in a home. Secondly, some staff work across multiple homes - a concept we refer to as staff-sharing. Staff acting as a bridge between care homes and the general population and staff-sharing induces a network, connecting care homes in a given community via their workers. This creates the potential for COVID-19 to spread from one home to another; hence, investigation of this pathway is important.

We develop a deterministic **SEIARD** compartmental metapopulation model with heterogeneous subpopulation sizes. Each subpopulation consists of a host human population, categorised further into six compartments of COVID-19 infection status: Susceptible (**S**), i.e., everyone who is not infected; Exposed (**E**), those exposed to the virus (and infected) but not yet infectious; Infectious and reported (**I**), infectious individuals that have been identified with a positive test; Unreported infectious (**A**), infectious individuals that have not been identified with a positive test; Recovered (**R**), those who had COVID-19 and recovered; and Dead (**D**), those who died from their illness. Symptomatic and asymptomatic individuals are not modelled explicitly; instead, asymptomatic infections contribute towards a reduction in the reporting rates. This model is illustrated in Figure 3.1 (a).

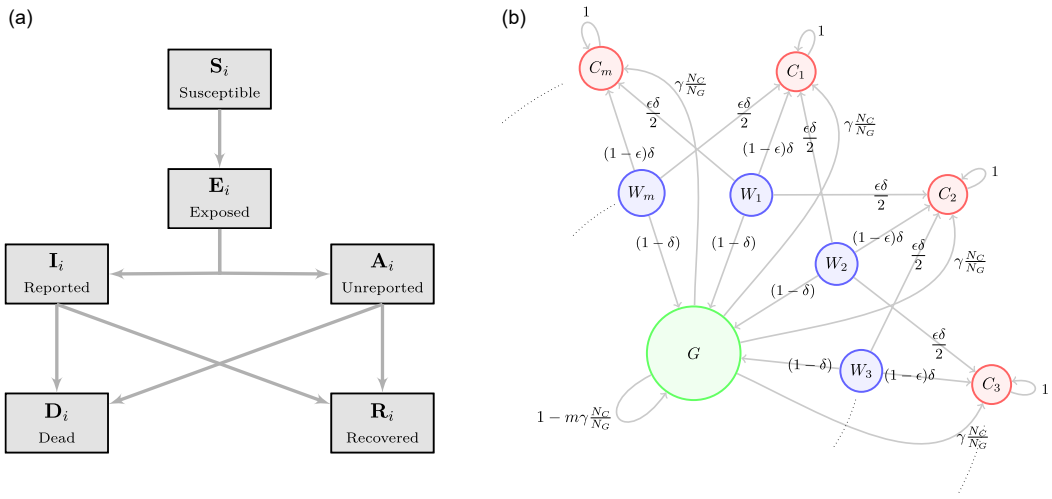

Fig. 3.1: Schematics for the compartmental and metapopulation structure. **(a)**: **SEIARD** compartmental structure of the model; **(b)**: Time-share network of interaction amongst subpopulations. Directed edge weights are p_{ik}, the proportion of people from subpopulation i who travel to mix at effective population k.

The metapopulation structure represents the population of the NHS Lothian health board in Scotland, but could be applied to other appropriate scenarios. We distinguish between care home residents, care home workers and the general population, modelling the $m = 109$ care homes for older adults in NHS Lothian [14]. The j^{th} home has a resident subpopulation, C_j, with a corresponding care home worker subpopulation, W_j. The general population is encapsulated by the subpopulation G (Figure 3.1b). Each care home includes the same number of residents, a simplifying assumption made due to lack of publicly available data on care home sizes in Lothian. We also assume the worker and resident subpopulations are the same size [15].

Each node of the network, $i \in X := \{C_1, C_2, ..., C_m, W_1, W_2, ..., W_m, G\}$ with $|X| = n$, is described in terms of the **SEIARD** compartmental model with equations:

$$\begin{aligned}
\frac{d\mathbf{S}_i}{dt} &= -\mathbf{S}_i\Lambda_i \\
\frac{d\mathbf{E}_i}{dt} &= \mathbf{S}_i\Lambda_i - \frac{\mathbf{E}_i}{\lambda} \\
\frac{d\mathbf{I}_i}{dt} &= r_i\frac{\mathbf{E}_i}{\lambda} - \frac{\mathbf{I}_i}{\tau} \\
\frac{d\mathbf{A}_i}{dt} &= (1-r_i)\frac{\mathbf{E}_i}{\lambda} - \frac{\mathbf{A}_i}{\tau} \\
\frac{d\mathbf{R}_i}{dt} &= (1-\mu_i)\frac{(\mathbf{I}_i + \mathbf{A}_i)}{\tau} \\
\frac{d\mathbf{D}_i}{dt} &= \mu_i\frac{(\mathbf{I}_i + \mathbf{A}_i)}{\tau}
\end{aligned} \tag{3.1}$$

Susceptibles in subpopulation i ($\mathbf{S}_i$), are introduced to a FOI Λ_i, and moved to the exposed class ($\mathbf{E}_i$). After a non-infectious latent period of λ days, they become infectious, testing positive at a reporting rate of r_i. These identified infections move to the class $\mathbf{I}_i$. Hence, any unidentified infections, $\mathbf{A}_i$, occur at rate 1 - r_i. After τ infectious days, a proportion μ_i of the infected population ($\mathbf{I}_i + \mathbf{A}_i$) die and the rest ($1 - \mu_i$) recover. For simplicity, and considering the model describes a short period of approximately 3 months, non-COVID-related deaths are not considered. For similar reasons, we do not include a birth rate or admission of new residents to care homes from the general population.

We assume a constant reporting rate for care home residents ($r_{i\in\{C_1,\ldots,C_m\}} = r_C$), workers ($r_{i\in\{W_1,\ldots,W_m\}} = r_W$), and the general public (r_G). The parameters τ and λ describe the infectious period and latency period, respectively, and are assumed to be the same across all subpopulations. Mortality rates, μ_i, vary by subpopulation, reflecting the positive association of serious outcomes of COVID-19 with age [16]. As we are modelling over a period of 4 months (approx. first wave), and immunity after COVID-19 infection lasts as long as 5 months [17] [18], we do not consider a transition from Recovered to Susceptible.

We model visitation to each care home by multiplying the proportion of the population, N_C/N_G, that visit the care home, and the duration of the visit, $\gamma(t)$, measured as a proportion of a day. The proportion remains constant over time while $\gamma(t)$ varies over time. In the model, each resident has one visitor per day,

up until 13th March 2020 [6]. Then $\gamma(t)$ drops to 0, reflecting the policy change to essential visitation only [14, 19]. $\gamma(t)$ is described by the function f(Ω_x), with $\Omega_x = \{t, \omega^x_{end}, \omega^x_{rate}, \omega^x_{low}, \omega^x_{high}\}$, defined below:

$$f(\Omega_x) = \frac{(\omega^x_{high} - \omega^x_{low})}{(1 + \exp(\omega^x_{rate}(t - \omega^x_{end})))(1 + \exp(-\omega^x_{rate}(t + 82)))} + \omega^x_{low}, \tag{3.2}$$

with the shape of a sigmoidal logistic function. The value of 82 is used in the function so that when $t = 0$, $f(\Omega_x) = \omega^x_{high}$. The function drops from ω^x_{high} to ω^x_{low} at a time controlled by ω^x_{end}, such that when $t = \omega^x_{end}$, $f(\Omega_x) = (\omega^x_{high} + \omega^x_{low})/2$. The ω_{rate} parameter changes the gradient of the descent at $t = \omega_{end}$.

Thus, visitation rate is described by $\frac{N_C(0)}{N_G(0)}\gamma(t) = \frac{N_C(0)}{N_G(0)}f(\Omega_\gamma)$. Given that visitation drops to 0 in the first 2 weeks of the simulation, the changes in population size over that time is negligible, hence we can keep the proportion of the population constant and control visitation by solely changing $\gamma(t)$.

Workers spend a constant proportion of their time, δ, at care homes. With $\delta = 0.5$, a worker compartment, W_i, spends half of their time at care homes, C_i, over the course of a day. This is equivalent to care homes splitting staff into two 12 hour shifts. Workers thus spend the rest of their time, $1-\delta$, mixing in the general population, G. Care homes operate with staff under differing working schedules and require a minimum number of staff to maintain adequate levels of care. This

places constraints on feasible values of δ. We assume this minimum value to be $\delta = 0.2$. This value equates to staff being split into five shifts throughout the day. Other possible shifts include care homes having a day and night shift ($\delta = 0.5$) or three 8-hour shifts ($\delta = 0.33$).

During the first wave of COVID-19 in Scotland, there was both intra-organizational staff-sharing between homes (i.e., staff who work at multiple homes belonging to the same care provider), as well as inter-organizational staff-sharing (use of bank or agency staff) [20, 21]. Therefore, a constant proportion of each homes' assigned workers, ε, were exchanged between homes every day. We refer to this as *staff-sharing*. We have made the simplifying assumption that the staff-sharing network has a topology of a circle, whereby the shared staff for home j are split evenly between homes $j-1$ and $j+1$ (Figure 3.1b). We also assume care home residents do not leave their homes.

Interaction across subpopulations is heterogeneous and is described in terms of time-sharing, determining proportions of subpopulations mixing in groups with each other. In the i^{th} subpopulation there are $N_i(t) = \mathbf{S}_i(t) + \mathbf{E}_i(t) + \mathbf{I}_i(t) + \mathbf{A}_i(t) + \mathbf{R}_i(t)$ active individuals who can mix with others. The proportion from subpopulation i who travel to, and mix with, subpopulation k is p_{ik}. The effective population size of subpopulation k, given that others have travelled to it and some people from k have left, is $\hat{N}_k(t) = \sum_{j \in X} p_{jk} N_j(t)$. We assume these effective populations, $\hat{N}_k(t)$, are well mixed, so people who travel to each population can meet all others there. There are two types of effective populations; the care

homes and the general population. Care home j, comprises $\hat{N}_{C_j}(t)$ people: its residents, its working staff, staff from other care homes, and visitors. The general population consists of $\hat{N}_G(t)$ people; this includes all the staff not at work and the non-visiting general population.

Our specific time-share assumptions are represented visually as a directed, weighted network in Figure 3.1 (b). The corresponding weighted adjacency matrix, the travel/time-share matrix, is $\boldsymbol{T} \in \mathbb{R}^{n\times n}$, whose $[i,j]^{th}$ element is p_{ij} and each matrix row sums to 1. The rows and columns of $\boldsymbol{T}$ are in the order of $\{C_1,C_2,...,C_m,W_1,W_2,...,W_m,G\}$. $\boldsymbol{T}$ consists of the partitions $\{T_{CC},T_{CW}$, T_{WC},T_{WW},T_{CG}, $T_{WG},T_{GC},T_{GW}\}$. For example, the submatrix T_{WC} defines the proportion of time that each worker subpopulation spends mixing in each care home. To clarify notation: matrix $\boldsymbol{I}_m$ indicates the identity matrix of dimension m, matrix $[a]_{m\times m}$ indicates a matrix of dimension $m\times m$ with all entries a. Hence $\boldsymbol{T}$ and the subsequent sub-matrices are as follows:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T_{CC}} & \boldsymbol{T_{CW}} & \boldsymbol{T_{CG}} \\ \boldsymbol{T_{WC}} & \boldsymbol{T_{WW}} & \boldsymbol{T_{WG}} \\ \boldsymbol{T_{GC}} & \boldsymbol{T_{GW}} & \boldsymbol{T_{GG}} \end{bmatrix}_{n\times n}, \tag{3.3}$$

$$\boldsymbol{T_{CC}} = \boldsymbol{I}_m, \quad \boldsymbol{T_{CW}} = \boldsymbol{T_{WW}} = \Big[\, 0 \,\Big]_{m\times m}, \quad \boldsymbol{T_{CG}} = \Big[\, 0 \,\Big]_{m\times 1}, \quad \boldsymbol{T_{GW}} = \Big[\, 0 \,\Big]_{1\times m},$$

$$\boldsymbol{T_{WG}} = \left[1-\delta\right]_{m\times 1}, \quad \boldsymbol{T_{GC}} = \left[\frac{N_C(0)}{N_G(0)}\gamma(t)\right]_{1\times m}, \quad \boldsymbol{T_{GG}} = \left[1-m\frac{N_C(0)}{N_G(0)}\gamma(t)\right]_{1\times 1},$$

$$\boldsymbol{T_{WC}} = \begin{bmatrix} (1-\varepsilon)\delta & \frac{\varepsilon\delta}{2} & 0 & \dots & 0 & \frac{\varepsilon\delta}{2} \\ \frac{\varepsilon\delta}{2} & (1-\varepsilon)\delta & \frac{\varepsilon\delta}{2} & \dots & 0 & 0 \\ 0 & \frac{\varepsilon\delta}{2} & (1-\varepsilon)\delta & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & (1-\varepsilon)\delta & \frac{\varepsilon\delta}{2} \\ \frac{\varepsilon\delta}{2} & 0 & 0 & \dots & \frac{\varepsilon\delta}{2} & (1-\varepsilon)\delta \end{bmatrix}_{m\times m}. \quad (3.4)$$

Disease transmission in the model is assumed to be frequency-dependent. The FOI integrates which infections occur to whom, from whom and where the infection takes place, as in [22, 23]. The FOI acting on subpopulation i, Λ_i (see Eq. (3.5)), accounts for the mixing that subpopulation i does in a day with all other subpopulations. It is most easily understood by considering $\Lambda_i \mathbf{S}_i$:

$$\Lambda_i \mathbf{S}_i = \sum_{k\in L_i} \frac{p_{ik}\mathbf{S}_i}{\hat{N}_k(t)} \sum_{j\in X} \beta_{ji}(t) p_{jk}(\mathbf{I}_j + \mathbf{A}_j) \quad (3.5)$$

The set of effective populations that subpopulation i travels to is L_i, consistent with the non-zero elements in the i_{th} row of the travel matrix $\boldsymbol{T}$. At effective population k, there is $p_{ik}\mathbf{S}_i$ susceptible individuals from i. At k there will also be $p_{jk}(\mathbf{I}_j + \mathbf{A}_j)$ infectious people from j who have travelled to k. The transmission rate between subpopulation j and i is $\beta_{ji}(t)$, therefore

$$\frac{p_{ik}\mathbf{S}_i}{\hat{N}_k}\beta_{ji}(t)p_{jk}(\mathbf{I}_j + \mathbf{A}_j)$$

is the number of new daily infections in i caused by people from j at the effective population k.

The transmission rates $\beta_{ji}(t)$ allow us to represent heterogeneous interaction patterns of individuals between and within different subpopulations. They incorporate the transmission dynamics of COVID-19 changing over time and location, for example, through lockdowns or other changes in behaviour [22]. We write $\beta_{ji}(t) = \frac{R(t)_{ji}}{\tau}$, describing the transmission rate $\beta_{ji}(t)$ between subpopulations j and i, with the reproduction rate, $R(t)_{ji}$, divided by the infectious period, τ. The contact rate and infection probability between subpopulations i and j is captured by $R(t)_{ji}$. We assume only the transmission rates between and within the subpopulation types (residents C, workers W, general public G) differ. Therefore, the transmission rates are arranged in a symmetric partitioned matrix $\boldsymbol{\beta} \in \mathbb{R}^{n\times n}$ whose $[j,i]^{th}$ element is $\beta_{ji}(t)$. The rows and columns of $\boldsymbol{\beta}$ are in the order of

$\{C_1, C_2, ..., C_m, W_1, W_2, ..., W_m, G\}$.The matrix $\boldsymbol{\beta}$ contains block sub-matrices;

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta_{CC}} & \boldsymbol{\beta_{WC}^T} & \boldsymbol{\beta_{GC}^T} \\ \boldsymbol{\beta_{WC}} & \boldsymbol{\beta_{WW}} & \boldsymbol{\beta_{GW}^T} \\ \boldsymbol{\beta_{GC}} & \boldsymbol{\beta_{GW}} & \boldsymbol{\beta_{GG}} \end{bmatrix}, \tag{3.6}$$

$$\boldsymbol{\beta_{CC}} = \Big[\beta_C(t)\Big]_{m\times m},\ \boldsymbol{\beta_{WC}} = \Big[\beta_C(t)\Big]_{m\times m},\ \boldsymbol{\beta_{WW}} = \Big[\beta_W(t)\Big]_{m\times m},$$

$$\boldsymbol{\beta_{GC}} = \Big[\beta_C(t)\Big]_{1\times m},\ \boldsymbol{\beta_{GW}} = \Big[\beta_G(t)\Big]_{1\times m},\ \boldsymbol{\beta_{GG}} = \Big[\beta_G(t)\Big]_{1\times 1}. \tag{3.7}$$

The matrix notation above is the same as for the travel matrix $\boldsymbol{T}$, and $\boldsymbol{\beta_{WC}^T}$ is the transpose of matrix $\boldsymbol{\beta_{WC}}$. For simplicity, we have assumed that the resident-resident, worker-resident, and general population-resident transmission rates are equal. Similarly, we assume the general population-worker and general population-general population transmission rates are the same. The transmission rates are described by:

$$\beta_C(t) = \frac{f(t, \omega^c_{end}, \omega^c_{rate}, \omega^c_{low}, \omega^c_{high})}{\tau},$$
$$\beta_W(t) = \frac{f(t, \omega^c_{end}, \omega^c_{rate}, \omega^c_{low}, (\omega^c_{high} + \omega^G_{high})/2)}{\tau},$$
$$\beta_G(t) = \frac{f(t, \omega^G_{end}, \omega^G_{rate}, \omega^G_{low}, \omega^G_{low}, \omega^G_{high})}{\tau} \tag{3.8}$$

where the function $\{f\}$ (Eq. (3.2)) models the reproduction rate. To simplify and to reduce the number of parameters, we relate the reproduction rate for workers in terms of the residents and general population. As care home workers balance their time between care homes and the general population, we assume the workers pre-lockown maximum reproduction rate is the average of the care homes and general populations, $\omega^W_{high} = (\omega^C_{high} + \omega^G_{high})/2$. We assume the reproduction rate for workers and residents drops at the same time, and to the same value.

3.3.2 Model calibration process

We used data from the network of care homes in NHS Lothian [14] complemented by Public Health Scotland Open Data, breaking down COVID-19 cases and deaths per health board [24, 25], to inform and calibrate our model. Parameters were found using a mixture of methods, as indicated in Table 3.2, including literature search, sensitivity of results, and rigorous fit based on minimising the sum of squares of residuals.

Table 3.2: Parameter definitions, alongside their base case values and source.

Parameter	Description	Value	Source
ε	Staff-sharing	0.4	Data fit
δ	Proportion of time workers spend at care homes	0.5	Assumption[a]
$r_{i\in\{C_1,\ldots,C_m\}} = r_C$	Reporting rate for residents	0.53	Estimated [14, 26–28]
$r_{i\in\{W_1,\ldots,W_m\}} = r_W$	Reporting rate for workers	0.52	Estimated [14, 24, 28–30]
r_G	Reporting rate for general public	0.077	Estimated [24, 29, 30]
$\mu_{i\in\{C_1,\ldots,C_m\}} = \mu_C$	Death rate for residents	0.25	Estimated [14, 26–28]
$\mu_{i\in\{G,W_1,\ldots,W_m\}} = \mu_G$	Death rate for general public (and workers)	0.017	Estimated [24, 25, 29, 30]
τ	Infectious period	7 days	[31]
λ	Latent period	5.8 days	[32]
m	Number of care homes	109	[14]
$N(0) = \sum_i N_i(0)$	Total initial population	907,580	[30]
$N_{C_i}(0)$	Initial resident subpopulation size	48	[14]
$N_{W_i}(0)$	Initial worker subpopulation size	48	[15]
$N_G(0)$	Initial general public subpopulation size	897,116	Estimated [14, 15, 30]
$\omega^{\mathbf{C}}_{\mathbf{end}}$	Timing of R_t descent for residents and workers	42 days	Manually fitted [b]
$\omega^{\mathbf{C}}_{\mathbf{rate}}$	Rate of descent of R_t for residents and workers	0.5	Assumption[c]
$\omega^{\mathbf{C}}_{\mathbf{low}}$	Post-descent R_t for residents and workers	0.6	Assumption[d]
$\omega^{\mathbf{C}}_{\mathbf{high}}$	Pre-descent R_t for residents	4.7	Data fit
$\omega^{\mathbf{G}}_{\mathbf{end}}$	Timing of R_t descent for general population	22 days	Manually fitted [b]
$\omega^{\mathbf{G}}_{\mathbf{rate}}$	Rate of descent of R_t for general population	0.5	Assumption[c]
$\omega^{\mathbf{G}}_{\mathbf{low}}$	Post-descent R_t for general population	0.6	[33]
$\omega^{\mathbf{G}}_{\mathbf{high}}$	Pre-descent R_t for the general population	4.1	Data fit
$\omega^{\gamma}_{\mathbf{end}}$	Timing of descent for visitation	10 days	[14, 19]
$\omega^{\gamma}_{\mathbf{rate}}$	Rate of descent of visitation	3	[14, 19]
$\omega^{\gamma}_{\mathbf{low}}$	Post-descent value for visitation	0	Assumption[e]
$\omega^{\gamma}_{\mathbf{high}}$	Pre-descent R_t for visitation	0.083	Data fit
H_{seeded}	Number of homes seeded	4	Data fit
$E_G(0)$	Initial general population infections	120	Data fit

[a] We assume workers spend half day at work, other half mixing in general population. Alternatively, workers do 12hr shifts. Units are given where appropriate in the Value column.

[b] Manually set by matching the model output to the infection peak dates in the NHS Lothian data [14, 24].

[c] Initial model exploration indicated that higher rates (steeper drops in R_t) resulted in infection peaks (in general public and residents) falling too quickly compared to the NHS Lothian data [14, 24]. The value of 0.5 corresponds to a descent of ~ 2 weeks.

[d] We assume the reproductive rate for every sub-population drops to the Scottish government's [33] estimated R_t after lockdown (so $\omega^C_{low} = \omega^G_{low}$).

[e] Equals 0 to reflect the policy change to essential visitation only [14, 19], and to avoid the complication of modelling end-of-life visitation.

Data

NHS Lothian is the second-largest health board in Scotland [34], providing public health services to an estimated 907,580 people (2019 mid-year population estimate [30]). The daily confirmed positive tests of COVID-19 cases reported across the entire health board were taken from the Public Health Scotland Open Data [24]. This data does not delineate which cases occurred in care homes, and thus, we retrieved the subset of cases in care homes from Burton et al. [14], which reports a 7-day average of confirmed cases in care home residents. Weekly COVID-19 deaths at the NHS Lothian health board level come from National Records Scotland [25]. Care home resident deaths are a subset of these and are published in [14]. Both death data are weekly counts of registered deaths where COVID-19 is mentioned on the death certificate (either as the underlying cause or as a contributory factor) [25].

Parameters set from evidence and assumptions

In this section, we describe our assumptions on some of the parameters responsible for inter-subpopulation mixing ($\omega^{\gamma}_{rate}, \omega^{\gamma}_{low}, \delta$) and seeding infection in the model.

For the rate at which visitation levels fell, ω^{γ}_{rate}, we chose the value of 3 to follow the rapid visitation policy changes in care homes [14, 19]. We have made the simplifying assumption that the post-lockdown visitation level $\omega^{\gamma}_{low} = 0$, to

avoid the complications of modelling end-of-life visitation in care homes. For simplicity, we assume that all homes operate under two 12-hour shifts per day, i.e., $\delta = 0.5$. Other shifts are explored in the sensitivity analysis.

We make a number of assumptions about the population initially infected, the first being that workers were not initially infected in the model. In the general population, we assume an equal amount of exposed and infected individuals (with and without symptoms), i.e., $E_G(0) = I_G(0) + A_G(0)$. In our model, care homes were seeded with infections via the parameter $H_{seeded} = |\{C_j \in \{C_1, \ldots, C_m\} : E_{C_j}(0) > 0\}|$, representing introductions such as hospital discharges. To account for the delay in infections at the start of the pandemic in care homes compared to the general population, as seen in the data Figure 3.2, we assume for all $j \in \{1, \ldots, m\}$, $I_{C_j}(0) = A_{C_j}(0) = 0$. Initially infected homes were seeded equally spaced on the circle sharing structure (see Figure 3.1 (b)). If a home is seeded then we assume $E_{C_j}(0) = 1$, and if not, $E_{C_j}(0) = 0$.

We use available data and assumptions to inform our reporting rates (r_G, r_C, r_S), death rates (μ_G, μ_C) and reproductive rate parameters $(\omega^C_{end}, \omega^G_{end}, \omega^C_{rate}, \omega^G_{rate}, \omega^C_{high}, \omega^G_{high})$. The details can be found in Appendix $S1$.

Data fit

While some model parameter values can be found based on the external data and literature, as shown in the previous section, other parameters were estimated using a formal fit to the cases and deaths data for the Lothian NHS health board (Table 3.2). These parameters were varied subject to constraints based on a combination of assumptions and information from the literature. We used the aggregated sum of squared errors (SSE) method of model output against the four data sets for NHS Lothian cases and deaths and choosing the parameter set which minimised this error. The data for NHS Lothian population cases and care home cases were in the form of daily and seven day averages respectively. The death data for both the NHS Lothian population and care homes were in weekly counts. To make the fitting consistent, we transformed the daily and seven day average data into weekly data for conformity (Figure 3.2). The constraints on the parameters in our model left 6 free parameters for formal fitting. Their ranges used for the data fit are shown in Table 3.3.

Table 3.3: Parameters used for the data fit and the sets of values simulated over.

Parameter	Values considered
ω^C_{high}	$\{3.3, 3.4, \ldots, 5\}$
ω^G_{high}	$\{3.3, 3.4, \ldots, 4.5\}$
ε	$\{0.1, 0.2, \ldots, 0.5\}$
$E_G(0)$	$\{100, 110, \ldots, 180\}$
ω^γ_{high}	$\{0.042, 0.083, 0.17\}$
H_{seeded}	$\{1, 2, \ldots 10\}$

To investigate the question of how many care homes were exposed at the start of the pandemic, we ran the fitting separately for H_{seeded} fixed at 1 through 10. We simulated the model over 21,060 combinations of the remaining parameters to calculate the minimum SSE, for each value of H_{seeded}. We investigate the distribution of the parameters in Table 3.3 in the top ten best fitting scenarios, for each value of H_{seeded}.

3.3.3 Sensitivity analysis

For our sensitivity analysis, we measured the change in total care home resident deaths when changing pairs of the time-share parameters (ω^γ_{high}, δ, ε), keeping all other parameters at the base case. The results were stored in a 50×50 grid and visualised using heat-maps to compare the key movements that spread COVID-

19 into and throughout care homes. We also performed a sensitivity analysis for other parameters that fall outside the focus of this research, found in Appendix S3.

3.4 Results

In this section, we first show how the model captures the NHS Lothian data for cases and deaths in the period from March to June 2020, and then show how sensitive the results are to changes in key parameters.

3.4.1 Data fit

The model captures the key features of the COVID-19 related cases and deaths in both care home and general populations, Figure 3.2. The minimum SSE was 33,042, with our model predicting 3,165 total cases and 817 total deaths compared to the total 3,123 cases and 709 deaths in the data (Figure 3.2c). The average difference between data and predictions was 3.5 cases/deaths per week. For care homes, the best fit model predicted 871 cases and 411 deaths compared to 903 cases and 423 COVID-19 related deaths in the data. Our model does not predict the initial jump in deaths in care homes due to our assumption that infection reporting is constant. Further, our model overestimates the number of deaths for all populations despite a good fit for the cases, as the calculation of death rates is tied to the reporting rates.

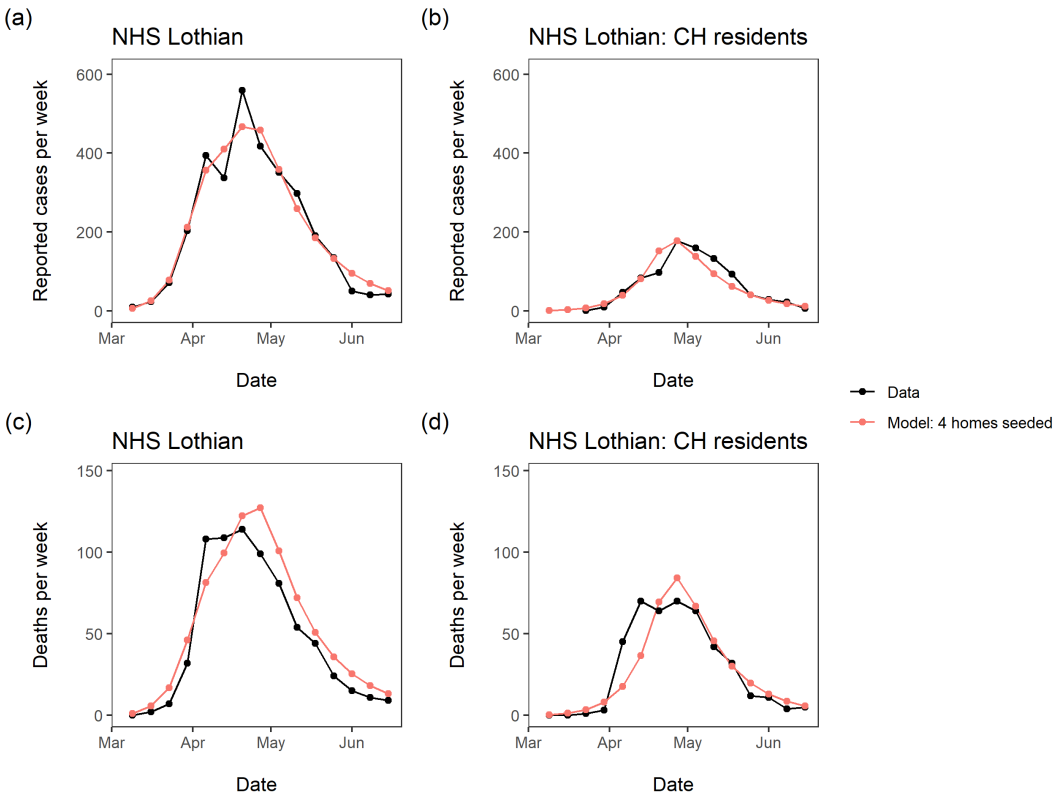

Fig. 3.2: Surveillance data and fitted model. Data used for fitting are black points, and model solution with parameter values in Table 3.2 are red points. (a) reported cases per week for all NHS Lothian inhabitants (care home residents, workers and the general population); (b) reported cases per week in NHS Lothian care home residents; (c) deaths per week for all NHS Lothian inhabitants (care home residents, workers and the general population); (d) deaths per week in NHS Lothian care home residents.

To assess the initial level of care home exposure to the virus, we consider the quality of fit as a function of H_{seeded}. The minimum sum of squares of residuals takes the shape of a parabola, with a minimum at $H_{seeded} = 4$, see Figure 3.B.2 in Appendix S2. This suggests that a relatively small number of homes were exposed to COVID-19 at the start of the first wave.

3.4.2 Sensitivity analysis

In our model, staff catching infections from the community is controlled by δ; staff spreading infections between homes through staff-sharing by ε; and visitors bringing infections into homes from the outside community by ω^{γ}_{high}. We determine which time-share parameters have the greatest impact on resident deaths from Figure 3.3. This shows the combined impact of varying pairs of these parameters on the total resident deaths.

Changing staff-sharing, ε, and staff shift patterns, δ, the final number of predicted resident deaths do not change significantly unless both are at their extreme values ($\delta = 1,\ \varepsilon = 0$) (Figure 3.3a). With the time staff spend at care homes at baseline ($\delta = 0.5$), a change from full staff-sharing ($\varepsilon = 0.5$) to removing sharing completely ($\varepsilon = 0$) we see a reduction of $\sim 25\%$ resident deaths. Similarly, with staff-sharing at baseline, and increasing δ from 0.2 to 1, we see a reduction in resident deaths of $\sim 30\%$. However, when considering $\delta = 1$, we should also restrict $\varepsilon = 0$, as staff living in care homes would not be shared across them. This parameter combination results in a reduction in resident deaths of $\sim 65\%$, compared to our baseline value.

Varying staff-sharing (ε) and pre-lockdown visitation (ω^{γ}_{high}) does not significantly affect predicted deaths while holding all other parameters constant at their respective values in Table 3.2, Figure 3.3 (b). However, from the isoclines, we see that an increase in staff-sharing (ε) leads to a greater number of deaths com-

pared to raising pre-lockdown visitation levels (ω^{γ}_{high}). Reducing pre-lockdown visitation from 2 visiting hours per resident to 0 hours per resident, would reduce our predicted first-wave deaths by about 10% of the death count in the first wave in NHS Lothian. Unsuprisingly, increasing δ to 1 and reducing staff-sharing and visitation parameters to 0 has the largest reduction in resident deaths ($\sim$ 90% from baseline), Figure 3.3 (c).

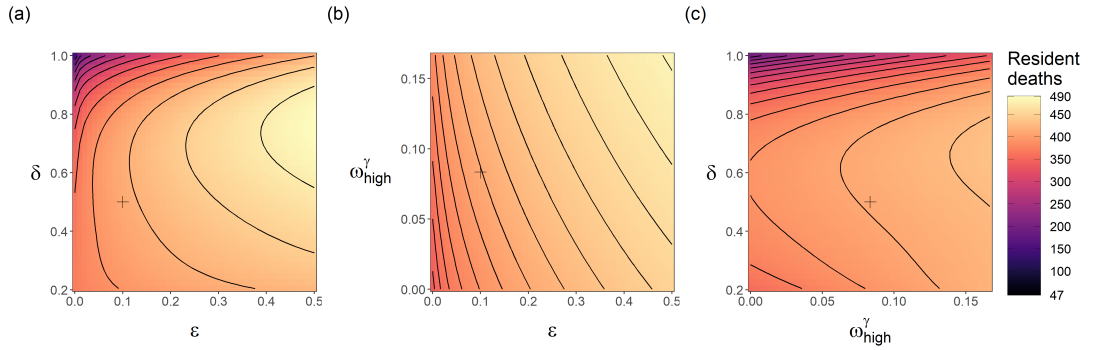

Fig. 3.3: Sensitivity of the final resident deaths to the time-share/mixing parameters ($\delta, \varepsilon, \omega^{\gamma}_{high}$). Proportion of CH staff time at work is δ, proportion of staff shared between homes is ε, and pre-lockdown visitation is ω^{γ}_{high}. Each panel shows the combined impact of varying two of the time-share/mixing parameters, with all other model parameters fixed as the base case (Table 3.3). The black lines in each panel are isoclines. The cross in each panel indicates the base case value for each parameter.

3.5 Discussion

To assess the impact of inter-subpopulation mixing in our care home model, we deployed a combination of modelling, data fit and simulations. Our model suggests that care homes are more at risk to outbreaks through staff infections from

the general population, compared to visitation or staff-sharing. These findings complement the results from Rosello et al., who used a stochastic compartmental model on single care homes in England [9].

Our model has a number of limitations. Changing worker shift patterns in care homes (δ) only weakly affects the model outcome for most "reasonable" values e.g. a 2-shift ($\delta = 0.5$) or 3-shift pattern ($\delta = 0.33$). It is only in the extremes where substantial differences are seen. For example, when δ is close to 1 and pre-lockdown visitation is low, this greatly reduces the final outbreak size in care homes. Thus, our results point to a strategy of staff living-in the care home, in conjunction with timely lowering of visitation, as an effective pandemic response. This was implemented in France, where outbreaks in care homes were reduced significantly in care homes where staff self-confined [35]. If living within the care home is not possible, this result of very high levels of δ may imply that the strategy of segmenting the staff away from both care home residents and the general population whilst they are not at work would be effective, i.e., organisation of accommodation for care home workers [3]. From our model, we observe that the most effective solution to keeping care homes safe from infection is to focus on the pathway from general population to workers to residents.

Eliminating staff-sharing did not eliminate outbreaks in our model simulations, suggesting that focusing on staff-sharing alone would be an inadequate control strategy. Reducing staff-sharing does reduce the outbreak severity, however, in our model, this impact is low.This conclusion is limited due to our assumption

of the circular contact structure, which in turn reflects limited data availability regarding the contact structure of the care home industry in Lothian (most likely due to commercial sensitivity). A different contact structure could result in staff-sharing leading to changes in exportation of infection from care homes with outbreaks. A more thorough examination on the contact-structure of this system and how that impacts disease spread dynamics would be an important contribution to current literature.

One way to achieve this would be to consider an addition of highly-connected hubs [36], which we expect would increase the effect of staff-sharing. In our simulations, staff-sharing has an effect when there exists a non-uniform distribution of infections in worker sub-populations. Worker sub-populations acting as hubs would acquire disease quicker and skew the distribution of infection amongst worker sub-populations. However, the general population strongly connects all nodes in the network, and dominates the impact of the staff-sharing network on disease spread, as we assume a single general population with full mixing. At the geographical scale modelled here (a health board) this is an appropriate assumption, although it would not hold for larger scales, e.g., the national level (Scotland). Furthermore, our results from considering no staff-sharing ($\varepsilon = 0$) will be the same for any contact structure, and we have shown that a disease control measure of solely no staff-sharing is insufficient at controlling spread. Thus, it is important to note that, for staff living in care homes and no staff-sharing, results are independent of contact structure.

A reduction in visitation reduces predicted resident deaths, as speculated in [3], and our model supports findings that visitation was not the driving cause of infection in care homes [37]. Since visitation was banned, the evidence for visitation causing outbreaks is limited. Investigation into continuing visitation during lockdown or similar would be necessary to see how the outcome would be different if visitation did not change at all; this was not the focus of our investigation.

From the data fit, there were a low number of care homes infected at the beginning of the first wave ($H_{seeded} = 4$). These initial infections could represent hospital discharges or other pathways of transmission. This finding aligns with the claims that resident hospital discharges might not have been the primary cause of care home outbreaks [13, 38]. However, it is important to acknowledge that hospital discharges continued during the first wave of COVID-19 [14]. Therefore, obtaining more detailed data and incorporating continuous discharges into our model is needed to fully assess the role of hospital discharges in care home outbreaks.

One of the model's limitations is that it does not explicitly account for the variation in susceptibility with age [39], and is only implicitly addressed by considering different values of β within and outwith of care homes, while keeping the staff and general population homogeneous. Due to the unavailability of data regarding care home worker infections, we expressed worker transmission rates in terms of transmission rates for care home residents and the general population.

We assumed the resident-resident and resident-worker transmission rates were equal. However, contacts between care-giving staff and residents are likely more frequent and closer than between residents. On the other hand, there may be more adherence or better knowledge of how to use PPE among staff. Also note that contact between residents could be reduced more easily during the pandemic [3].

The data fit was achieved by minimising the SSE for each of the four time series data. This method requires the errors to be independent, follow a normal distribution, and for the variance to be constant. With the data, we could not estimate the variance over time. To mitigate the effect of the differing variance of the data sets, we shifted the four time series to the same scale, this being weekly cases/deaths.

In our model, we assume a uniform home size in order to keep the model generic. As a result (and since the model is deterministic) the risk of staff and visitors bringing in infections is the same for all care homes, which may result in an underestimation of the initial rate of spread. An obvious extension of the paper would be to consider various sources of heterogeneity, including size. The size of individual care homes is believed to be the main factor that influences the likelihood of a care home outbreak [14, 20, 40]. However, larger homes typically have more staff and therefore a higher chance of experiencing an outbreak before the smaller ones. In general, we expect larger care homes to receive an increased force of infection from all sources, proportional to its increased size, and therefore

an increased outbreak risk. This in turn could increase risk for smaller homes directly connected to the larger ones through staff-sharing and visitations, and the overall outbreak risk. However, this effect could be balanced by a lowered risk associated with small care homes, with the total population size kept constant.

The National Records Scotland death data used were the dates of death registration, not the actual date of death. This is limiting, as we are an average of three days behind in the prediction of deaths [25]. The data for care home resident deaths includes deaths in hospitals, including nosocomial infections, which we do not take into account in our model. We expect this not to limit the interpretation of our results, as hospital deaths of care home residents were approximately only 5% of the total care home resident deaths [25].

We do not distinguish explicitly between symptomatic and asymptomatic individuals; however, asymptomatic infections implicitly affect this model's reporting rates. We do not explicitly model self-isolation or any behavioural change after infection, nor delays or changes in reporting. For simplicity, the model assumes that infections are immediately reported. Since we are not explicitly modelling behaviour changes once individuals are infected, we do not expect incorporating reporting delays to significantly affect our results on the infection pathways. Reporting differed over time, especially in the early weeks of the pandemic when testing was scarce; in care homes, the national policy was to test only the first few symptomatic residents [14]. With a constant reporting rate, we overestimate

the number of positive cases prior to the policy change, and underestimate the cases afterwards. A time-dependent reporting rate would impact our death rate and reproductive rate parameters, and therefore requires further study.

We focused our analysis on the mixing patterns of patient-facing care home workers and have not considered the impact of non-care staff. From a study on care homes in Norfolk during April and May 2020, the number of non-care staff in homes was found to be the most statistically significant predictor of COVID-19 entry into homes [41]. We do note, however, that this study did not consider the role of staff-sharing. A valuable extension of our model would include non-care worker subpopulations with different contact structures to other homes compared to patient-facing staff.

Data regarding care home outbreaks were limited due to the commercial nature of care home organisations in Scotland. Making this data available would allow for additional modelling approaches. Adding further heterogeneity into the system by including a distribution of home sizes and types would further improve the modelling approach. Including a stochastic component to this model could lead to more insight into "super-spreader" events in care homes [42] and their effect on epidemic response.

3.6 Conclusion

For this study, we developed a compartmental metapopulation model which allowed for the parameterisation of mixing across subpopulations, specifying where subpopulations spend their time at each location as a proportion of their day. This model has been used for simulating the spread of COVID-19 throughout care homes in the Scottish health board of NHS Lothian. However, this methodology can be applied in many other contexts, for example, modelling the mixing of individuals across different locations with levels of disease transmission differing per location, such as workplaces or prisons.

In our study, we investigated the inter-subpopulation mixing between care home residents, care home staff and the general population, in relation to the spread of COVID-19 during the first wave of the pandemic. With this view, we find that a complete restriction on staff-sharing or visitation alone is not wholly sufficient as a disease control measure. However, the modelled strategy of staff living in care homes with no staff-sharing is effective at reducing COVID-related deaths of care home residents by 65% during the first wave of the pandemic. Importantly, this result holds for any contact structure between staff and care homes. Unsurprisingly, a restriction in the movements of all populations (no visitation, no staff-sharing, staff living in care homes) effectively controls disease spread, reducing cases in residents by approximately 90%.

Results from our model indicate that protecting care home staff from the disease, coupled with reductions in staff-sharing across care homes and expedient cancellations of visitations, can lead to an effective reduction in outbreak size in care homes. Our findings highlight the need for more planning and support for care homes and their staff in organising quick and effective responses to emerging pandemics.

3.7 Acknowledgements

The authors thank Public Health Scotland for their support, and Prof. Bruce Guthrie (University of Edinburgh) for sharing the Lothian care home data. EMT has been partially supported by the University of Strathclyde Student Excellence Award and MB by Defra. Results were obtained using the ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk) based at the University of Strathclyde. Results were also obtained using the Pin cluster at the University of Stirling.

Appendix

3.A Parameter assumptions and estimates

In this section, we describe our estimates and assumptions on parameters for our reporting rates (r_G, r_C, r_S), death rates (μ_G, μ_C) and reproductive rate parameters ($\omega^C_{end}, \omega^G_{end}, \omega^C_{rate}, \omega^G_{rate}, \omega^C_{high}, \omega^G_{high}$).

A Scottish population study between 10^{th} April to 15^{th} June [29] estimated a combined adjusted seroprevalence across their study period (first wave = 10^{th} April to 15^{th} June) of 4.3% (95% CI 4.2%-4.5%). As of the week beginning 15^{th} June 2020, there had been 18,077 positive tests [24], which as a percentage of Scotland's population (2019 census [30]) is $\sim 0.33\%$. We use this information to assume a constant reporting rate in the first wave for the general public of $r_G = 0.33/4.3 \sim 0.077$.

In Scotland, the policy from the start of March to 16$^{\text{th}}$ April 2020 was to test only the first few symptomatic care home residents, and afterwards, was to test all symptomatic residents [14]. Assuming when there is an outbreak in a home, 40% of the residents end up infected (40% incidence) [26, 27]. Given 48 residents per care home, until 16$^{\text{th}}$ of April we assume a reporting rate of (a few tested)/(total infected) = $3/(0.4 \times 48) = 5/32$. After 16$^{\text{th}}$ April, we assume all the symptomatic cases are reported, giving a reporting rate of 4/5 (an estimated symptomatic proportion of COVID-19 cases in long term aged care is 80% [28]). Between the start of our simulation (6$^{\text{th}}$ March 2020) and 16$^{\text{th}}$ April 2020 is a time difference of 42 days, and between 17$^{\text{th}}$ April 2020 and the end of our simulation period (15$^{\text{th}}$ June 2020) is a time difference of 60 days. Therefore, for 42 days, we assume a reporting rate of 5/32, and for 60 days, it is 4/5. The weighted average and constant CH reporting rate over the simulation period is $r_C = (5/32)(42/102) + (4/5)(60/102) \sim 0.53$.

Until the 17$^{\text{th}}$ of April, we assume the staff reporting rate was the same as the general public (0.077). From then on, we assume the care home testing policy change (on the 17$^{\text{th}}$ of April) extended to their staff [14], and the reported percentage of cases was 83% (the symptomatic proportion [28]). Our weighted average and constant staff reporting rate over the simulation period is $r_S = 0.077 \times (42/102) + 0.8360/102 \sim 0.52$.

There are two constant death rates in our model: a resident death rate (μ_C) and a general population death rate (μ_G). We assume care home staff have the same death rate as the general population. There were $\sim$ 899 positive tests and 423 deaths in NHS Lothian care home residents over the study period. Using our resident reporting rate, we estimate there were $899.1/0.53 \sim 1697$ total residents infected with COVID-19 over the study period. Therefore, we estimate a resident death rate of $\mu_C = 423/1697 \sim 0.25$. Similarly, there were 3123 total positive tests and 709 deaths over the study period in NHS Lothian overall. Using our general reporting rate, r_G, we estimate a general population death rate of $\mu_G = 709/(3123/0.077) \sim 0.017$.

Under our parameterisation, the timing of the drop in reproductive rates for care home residents (ω^C_{end}) and for the general population (ω^G_{end}) control the timing of peak infections in each respective population, independent of all other parameter values. This is linked to the reproductive rate function (f) at an inflection point at $t = \omega^x_{end}$ (where $f(\Omega_x)$ takes the value of $(\omega^x_{high} + \omega^x_{low})/2$). Therefore, we manually set these parameters (ω^C_{end}, ω^G_{end}) by matching the model output to the infection peak dates in the NHS Lothian data [14, 24].

The ω^G_{rate} and ω^C_{rate} parameters were assumed to be 0.5. Initial model exploration indicated that higher rates (steeper drops in R_t) resulted in infection peaks (in general public and residents) falling too quickly compared to the NHS Lothian data [14, 24]. From sensitivity analysis, we found that changing the val-

ues of these parameters does not affect the disease dynamics (in terms of total infections/deaths). The ω_{rate} parameter controls the steepness of the descent from ω_{high} to ω_{low}, however the timing of the start of the descent changes to almost cancel out the effect of changing the steepness of this drop. The value of $\omega^G_{rate} = \omega^C_{rate} = 0.5$, corresponds to a descent of ~ 2 weeks.

The ω^G_{low} and ω^C_{low} values were set to 0.6, the estimated R_t after the first wave in Scotland [33]. Due to the uncertainty in the timing of the drop and the R_t peak value, we did not use this source for the ω^G_{high} and ω^G_{end} parameters.

Closed environments are conducive to COVID-19 transmission and super-spreading events [42], therefore we assume pre-lockdown transmission rates within care homes are not less than the general populations, $\omega^C_{high} \geqslant \omega^G_{high}$.

3.B Supplementary data fitting figures

In this section, we show results from our data fitting process for the other parameters in our model.

The reproduction rates change rapidly over the period of April - May 2020, Figure 3.B.1 reflecting the delayed effect of the lockdown. The care home resident population's fall in reproduction rate is delayed by ~ 3 weeks compared to the

fall in the general population. This delay is informed by the data, due to the ω_{end} parameters controlling the timing of the peaks in Figure 3.2 - we must be careful when attempting to interpret this delay.

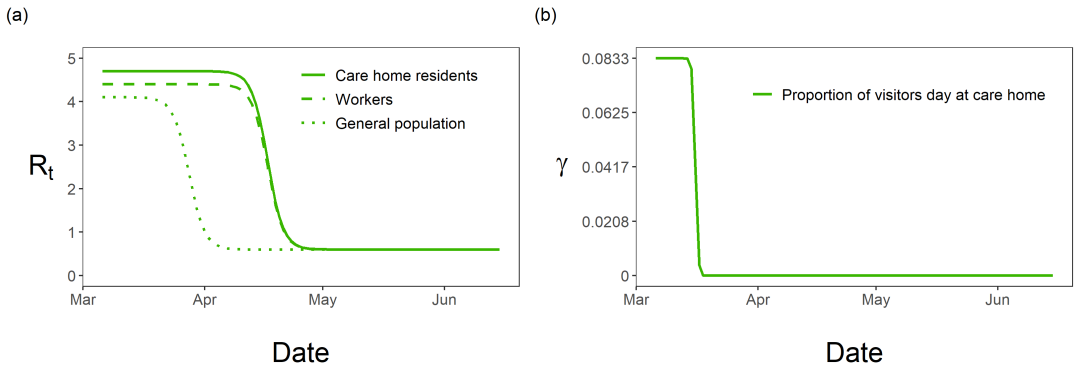

Fig. 3.B.1: Fitted time-dependent parameters. (a) Fitted reproductive numbers over time for care home residents, $R_C(t)$, workers, $R_W(t)$, and general population, $R_G(t)$; (b) fitted visitation, γ, over time with drop highlighting the change in policy.

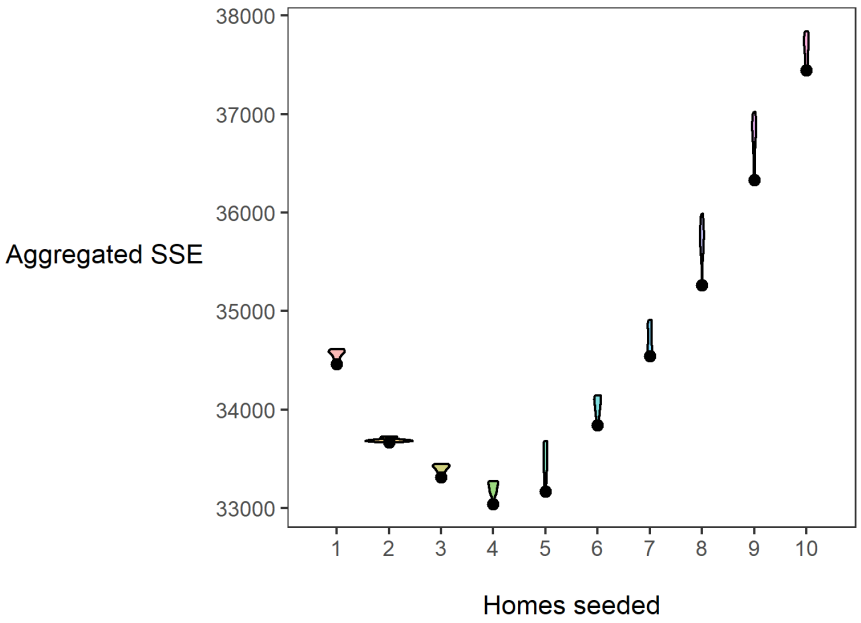

Fig. 3.B.2: Quality of fit as a function of homes seeded. Each violin is the distribution of aggregated sum of squared errors (SSE) in the top ten best-fitting parameter sets, for a number of homes seeded. Black dots indicate the minimum aggregated SSE achieved for each home seeded.

The optimal choice (in terms of minimum SSE) for the parameters as used in the data fit (Table 3.3) is relatively stable with respect to changes in H_{seeded}, Figure 3.B.3. The pre-lockdown reproduction rate in care homes, ω^C_{high}, appears stable in the range of 4.5 to 4.7, changing for 10 homes seeded with the optimal value lowering to 3.9. This highlights the clear link between the reproduction rate and the exposure of care homes at the beginning of the pandemic. For the pre-lockdown reproduction rate in the general population, ω^G_{high}, we see a stable optimal value in the range of 3.9 to 4.2. We see a lot of uncertainty

in the proportion of staff shared ε, for H_{seeded} outside of the range of 4 to 8. The distribution for ε changes for $H_{seeded} = 10$, with the optimal value going back up to 0.5. This coincides with the rise in $E_G(0)$ and the substantial fall in ω^C_{high}. This points to the correlation between these 3 parameters. Similarly, there is a lot of uncertainty in the value of pre-lockdown visitation, ω^γ_{high}, with an optimal choice for every value considered in our fitting as we vary H_{seeded}. This uncertainty highlights that the parameters in our model are highly correlated. As seen in Figure 3.C.1 and Figure 3.3, Figure 3.B.3 can also be seen to hint at how effective these parameters are at affecting the outcome of the model. For example, the variability in the chosen value of ε and ω^γ_{high} can also be attributed to the relatively small affect they have on the model outcome.

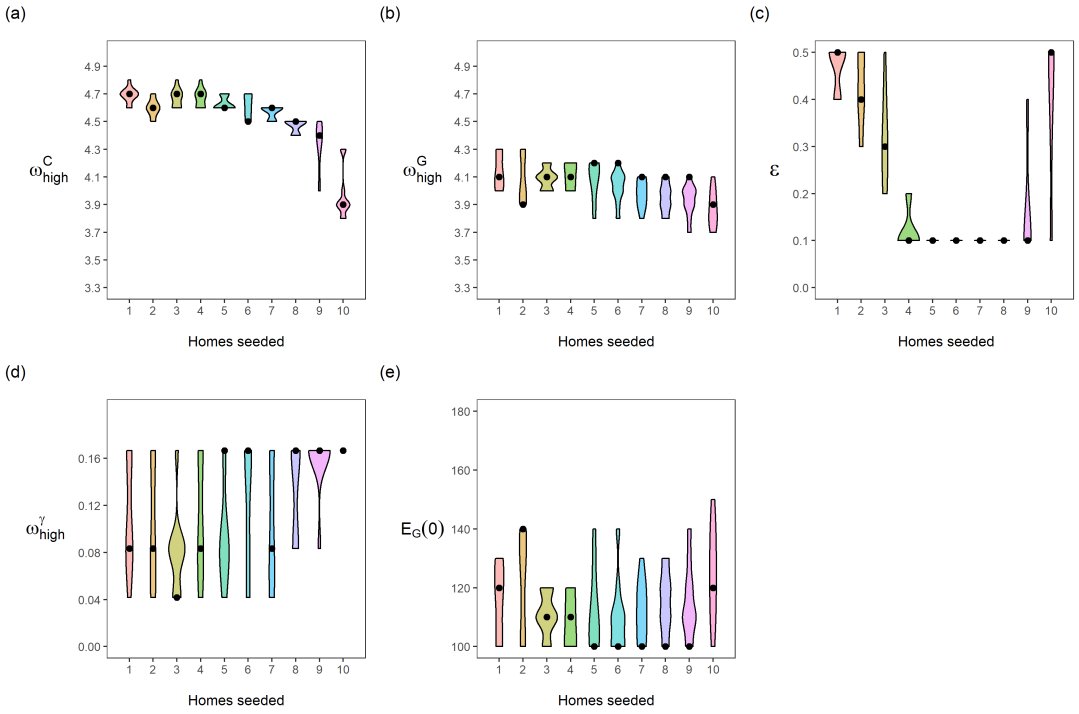

Fig. 3.B.3: Distribution of fitted parameters as a function of homes seeded. Each panel is a different calibrated parameter. Each violin in a panel is the distribution of individual parameters in the top ten best fitting parameter sets, for a number of homes seeded. (a) pre-lockdown care home resident R_t, ω^C_{high}; (b) pre-lockdown general public R_t, ω^G_{high}; (c) staff-sharing, ε; (d) visitation pre-lockdown, ω^γ_{high}; (e) general public seeded cases, $E_G(0) = I_G(0) + A_G(0)$. Black dots indicate the parameter value giving the lowest aggregated sum of squared errors (SSE), for each number of homes seeded.

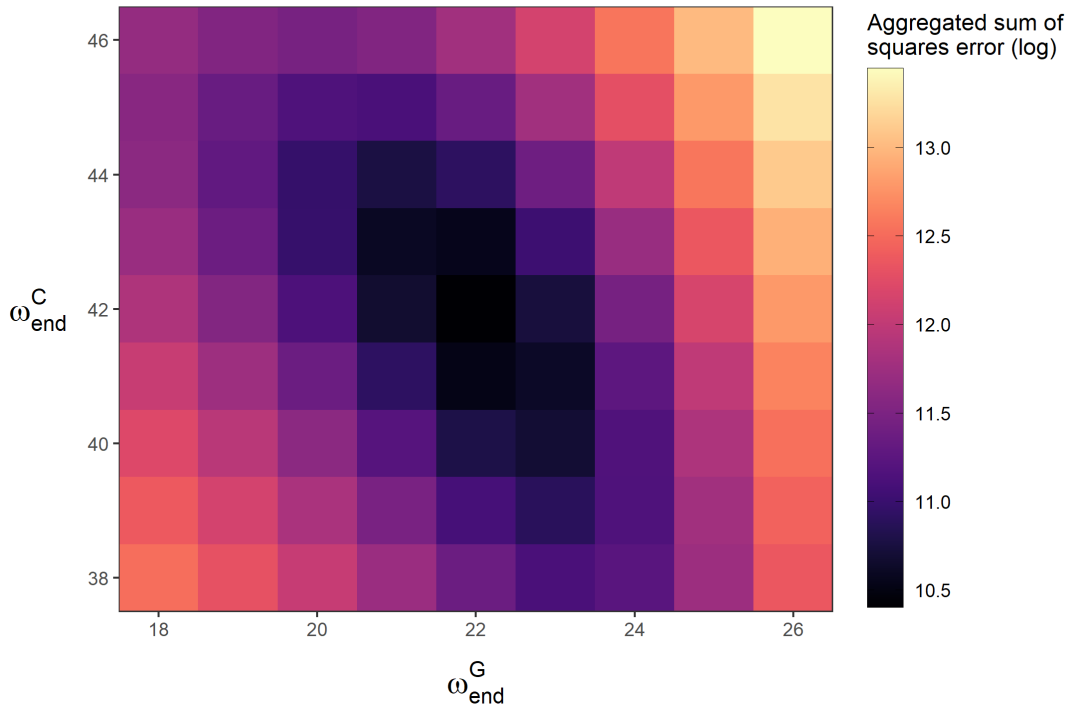

Fig. 3.B.4: Natural log of aggregated sum of squared error (SSE) in a $\omega_{end}^{C} - \omega_{end}^{G}$ parameter space. ω_{end}^{C} controls the timing of the drop in resident R_t, and ω_{end}^{G} controls the timing of the drop in general population R_t. These parameters were fitted manually, achieving a minimum for the values shown in Table 3.2. In the plot, all other parameters are held at the base case (Table 3.2).

3.C Supplementary sensitivity analysis figures

In this section, we include figures from our sensitivity analysis for the non-movement parameters.

After identifying the parameter set that minimises the SSE, the base case (Table 3.3), we performed a sensitivity analysis. We measured the change in each population's deaths when shifting individual parameters in Table 3.C.1 from the base case. This allowed us to assess the relative impact of individual parameters on each population.

Table 3.C.1: Parameters involved in the sensitivity analysis. Sensitivity shift is the unit of change used for each parameter from its base case. These values were chosen to measure the change in each population's deaths to small perturbations of individual parameters from it's base case.

Parameter	**Sensitivity shift**
ω_{end}^{C}	42 ± 1
ω_{high}^{C}	4.7 ± 0.1
ω_{low}^{C}	0.6 ± 0.1
ω_{rate}^{C}	0.5 ± 0.1
ω_{end}^{G}	22 ± 1
ω_{high}^{G}	4.1 ± 0.1
ω_{low}^{G}	0.6 ± 0.1
ω_{rate}^{G}	0.5 ± 0.1
δ	0.5 ± 0.1
ε	0.4 ± 0.05
ω_{high}^{γ}	0.083 ± 0.0167
$E_G(0) = I_G(0) + A_G(0)$	120 ± 10
λ	5.8 ± 0.3
τ	7 ± 0.4

Figure 3.C.1 indicates the sensitivity of the predicted deaths to the parameters in Table 3.C.1 for care home residents, workers and general population. Predictions are most sensitive to the infectious period, τ, and latency period, λ. The parameters $\{\omega^C_{end}, \omega^C_{high}, \omega^C_{low}\}$ significantly influence care home and worker deaths, without affecting predicted deaths in the general population. Interestingly, a change in one day from when the care home reproduction rate drops, results in almost 10% change in predicted resident and worker deaths. The parameters, $\{\omega^G_{end}, \omega^G_{high}, \omega^G_{low}\}$, controlling the timing of the reproduction rate and it's value before and after lockdown for the general population, significantly affect predicted deaths in the general population. Interestingly, changing the value of δ by 0.1 (20%) results in very little effect to the residents ($< 5\%$) and general population ($< 1\%$) but a 15% change in predicted worker deaths.

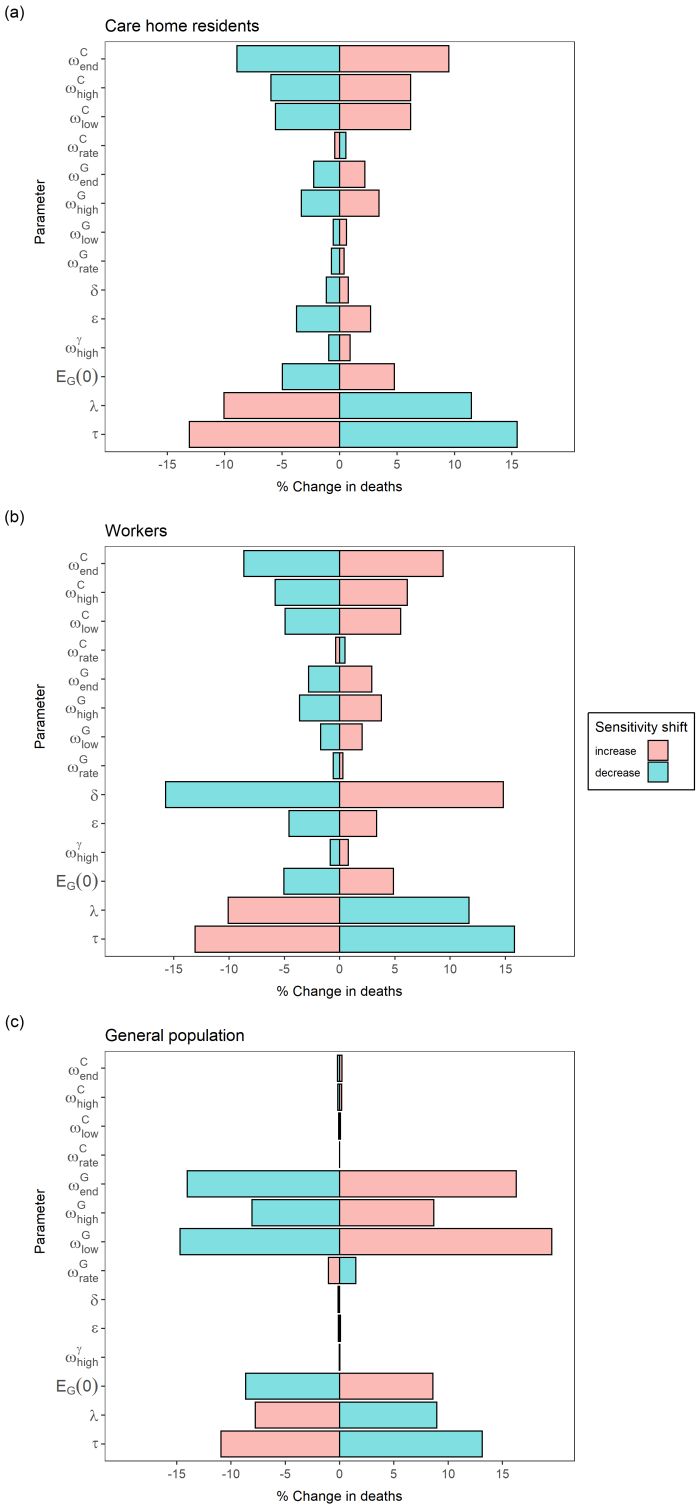

Fig. 3.C.1: Sensitivity of the final deaths in each population to perturbations in model parameters. Each bar shows the % change in final deaths in a population caused by shifting an individual parameter from the base case, keeping all other parameters fixed at the base case (Table 3.2). Each parameter is increased or decreased from its base case value by the corresponding 'sensitivity shift' value in Table 3.C.1.

Bibliography

[1] World Health Organisation, "WHO Coronavirus (COVID-19) Dashboard With Vaccination Data," 2021. [Online]. Available: https://covid19.who.int/

[2] D. Bell, A. Comas-Herrera, D. Henderson, S. Jones, E. Lemmon, M. Moro, S. Murphy, D. O'Reilly, and P. Patrignani, "COVID-19 mortality and long-term care: a UK comparison," LTCcovid, Tech. Rep. August, 2020. [Online]. Available: https://ltccovid.org/wp-content/uploads/2020/08/COVID-19-mortality-in-long-term-care-final-Sat-29-1.pdf

[3] UK Government Social Care Working Group, "Commission : What are the appropriate layers of mitigation to deploy for care homes in the context of post vaccination risk landscape?" Tech. Rep., 2021. [Online]. Available: https://www.gov.uk/government/publications/scwg-what-are-the-appropriate-mitigations-to-deploy-in-care-homes-in-the-context-of-the-post-vaccination-risk-landscape-26-may-2021

[4] L. K. Nguyen, I. Megiddo, and S. Howick, "Challenges of infection prevention and control in Scottish long-term care facilities," *Infection control and hospital epidemiology*, vol. 41, no. 8, pp. 943–945, 8 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32317037/

[5] D. R. M. Smith, A. Duval, K. B. Pouwels, D. Guillemot, J. Fernandes, B.-T. Huynh, L. Temime, and L. Opatowski, "Optimizing COVID-19 surveillance in long-term care facilities: a modelling study," *BMC Medicine 2020 18:1*, vol. 18, no. 1, pp. 1–16, 12 2020. [Online]. Available: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-01866-6

[6] L. K. N. Nguyen, S. Howick, D. McLafferty, G. H. Anderson, S. J. Pravinkumar, R. Van Der Meer, and I. Megiddo, "Evaluating intervention strategies in controlling COVID-19 spread in care homes: An agent-based model," *Infection Control and Hospital Epidemiology*, vol. 42, no. 9, pp. 1–11, 2020. [Online]. Available: https://doi.org/10.1017/ice.2020.1369

[7] L. K. N. Nguyen, I. Megiddo, and P. S. Howick, "REPORT 3 : IMPACT OF VARIOUS VACCINATION COVERAGES ON THE SPREAD OF COVID-19 AND DEATHS IN CARE HOMES," Department of Management Science, University of Strathclyde, Tech. Rep. January, 2021. [Online]. Available: https://ltccovid.org/wp-content/uploads/2020/12/Report-3_Care-homes_Vaccination_Strathclyde.pdf

[8] L. K. N. Nguyen, S. Howick, D. McLafferty, G. H. Anderson, S. J. Pravinkumar, R. Van Der Meer, and I. Megiddo, "Impact of visitation and cohorting policies to shield residents from covid-19 spread in care homes: an agent-based mode," *AJIC: American Journal of Infection Control*, vol. 49, 2021. [Online]. Available: https://doi.org/10.1016/j.ajic.2021.07.001

[9] A. Roselló, R. C. Barnard, D. R. M Smith, S. Evans, F. Grimm, N. G. Davies, S. R. Deeny, G. M. Knight, W. John Edmunds, and A. Affiliations, "Impact of non-pharmaceutical interventions on SARS-CoV-2 outbreaks in English care homes: a modelling study," *BMC Infectious Diseases*, vol. 22, no. 324, 2022. [Online]. Available: https://doi.org/10.1186/s12879-022-07268-8

[10] C. E. Overton, H. B. Stage, S. Ahmad, J. Curran-Sebastian, P. Dark, R. Das, E. Fearon, T. Felton, M. Fyles, N. Gent, I. Hall, T. House, H. Lewkowicz, X. Pang, L. Pellis, R. Sawko, A. Ustianowski, B. Vekaria, and L. Webb, "Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example," *Infectious Disease Modelling*, vol. 5, pp. 409–441, 2020. [Online]. Available: https://doi.org/10.1016/j.idm.2020.06.008

[11] Joachim Oberhammer, "Social-distancing effectiveness tracking of the COVID-19 hotspot Stockholm," 2020, medRxiv preprint. [Online]. Available: https://doi.org/10.1101/2020.06.30.20143487.

[12] B. A. D. v. Bunnik, A. L. K. Morgan, P. R. Bessell, G. Calder-Gerver, F. Zhang, S. Haynes, J. Ashworth, S. Zhao, R. N. R. Cave, M. R. Perry, H. C. Lepper, L. Lu, P. Kellam, A. Sheikh, G. F. Medley, and M. E. J. Woolhouse, "Segmentation and shielding of the most vulnerable members of the population as elements of an

exit strategy from COVID-19 lockdown," *Philosophical Transactions of the Royal Society B*, vol. 376, no. 1829, 2021. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2020.0275

[13] Public Health Scotland, "Discharges from NHS Scotland Hospitals to Care Homes between 1 March and 31 May," Public Health Scotland, Tech. Rep., 2020. [Online]. Available: https://publichealthscotland.scot/publications/discharges-from-nhsscotland-hospitals-to-care-homes/discharges-from-nhsscotland-hospitals-to-care-homes-between-1-march-and-31-may-2020/#:~:text=Between%201%20March%20and%2031%20May%202020%2C%20there%20were%205%2C204,discharges%20during%20the%20same%20period.

[14] J. K. Burton, G. Bayne, C. Evans, F. Garbe, D. Gorman, N. Honhold, D. McCormick, R. Othieno, J. E. Stevenson, S. Swietlik, K. E. Templeton, M. Tranter, L. Willocks, and B. Guthrie, "Evolution and effects of COVID-19 outbreaks in care homes: a population analysis in 189 care homes in one geographical region of the UK," *The Lancet Healthy Longevity*, vol. 1, no. 1, pp. e21–e31, 2020. [Online]. Available: http://dx.doi.org/10.1016/S2666-7568(20)30012-X

[15] E. S. Knock, L. K. Whittles, J. A. Lees, P. N. Perez-Guzman, R. Verity, R. G. FitzJohn, K. A. Gaythorpe, N. Imai, and et al, "Report 41 - The 2020 SARS-CoV-2 epidemic in England: key epidemiological drivers and impact of interventions | Faculty of Medicine

| Imperial College London," Imperial College London, Tech. Rep. December, 2020. [Online]. Available: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-41-rtm/

[16] P. M. McKeigue and H. M. Colhoun, "Evaluation of "stratify and shield" as a policy option for ending the COVID-19 lockdown in the UK," 2020, medRxiv preprint. [Online]. Available: https://doi.org/10.1101/2020.04.25.20079913.

[17] T. J. Ripperger, J. L. Uhrlaub, M. Watanabe, R. Wong, Y. Castaneda, H. A. Pizzato, M. R. Thompson, C. Bradshaw, C. C. Weinkauf, C. Bime, H. L. Erickson, K. Knox, B. Bixby, S. Parthasarathy, S. Chaudhary, B. Natt, E. Cristan, T. El Aini, F. Rischard, J. Campion, M. Chopra, M. Insel, A. Sam, J. L. Knepler, A. P. Capaldi, C. M. Spier, M. D. Dake, T. Edwards, M. E. Kaplan, S. J. Scott, C. Hypes, J. Mosier, D. T. Harris, B. J. LaFleur, R. Sprissler, J. Nikolich-Žugich, and D. Bhattacharya, "Orthogonal SARS-CoV-2 Serological Assays Enable Surveillance of Low-Prevalence Communities and Reveal Durable Humoral Immunity," *Immunity*, vol. 53, no. 5, pp. 925–933, 2020. [Online]. Available: https://doi.org/10.1016/j.immuni.2020.10.004

[18] J. M. Dan, J. Mateus, Y. Kato, K. M. Hastie, E. D. Yu, C. E. Faliti, A. Grifoni, S. I. Ramirez, S. Haupt, A. Frazier, C. Nakao, V. Rayaprolu, S. A. Rawlings, B. Peters, F. Krammer, V. Simon, E. O. Saphire, D. M. Smith, D. Weiskopf, A. Sette, and S. Crotty,

"Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection," *Science*, vol. 371, no. 6529, pp. 1–23, 2021. [Online]. Available: https://doi.org/10.1126/science.abf4063

[19] The Scottish Government, "Coronavirus (COVID-19) social care response - 13 March 2020," 2020. [Online]. Available: https://www.careinspectorate.com/images/COVID-19_-_Letter_from_Cabinet_Secretary_for_Health_and_Sport_-_Social_care_guidance_-_13_March_2020.pdf

[20] J. Reilly, D. Crawford, and D. O. Boyle, "CARE HOME REVIEW: A rapid review of factors relevant to the management of COVID-19 in the care home environment in Scotland," Scottish Government: Cabinet Secretary for Health and Sport, Edinburgh, Tech. Rep., 2020. [Online]. Available: https://www.gov.scot/publications/root-cause-analysis-care-home-outbreaks/

[21] Office For National Statistics, "Impact of coronavirus in care homes in England: 26 May to 19 June 2020," Office For National Statistics, Tech. Rep. July, 2020. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/impactofcoronavirusincarehomesinenglandvivaldi/26mayto19june2020

[22] A. Rădulescu, C. Williams, and K. Cavanagh, "Management strategies in a SEIR-type model of COVID 19 community spread," *Scientific Reports*, vol. 10, no. 1, pp. 1–16, 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-77628-4

[23] D. Calvetti, A. P. Hoover, J. Rose, and E. Somersalo, "Metapopulation Network Models for Understanding, Predicting, and Managing the Coronavirus Disease COVID-19," *Frontiers in Physics*, vol. 8, no. June, pp. 1–16, 2020. [Online]. Available: https://doi.org/10.3389/fphy.2020.00261

[24] Public Health Scotland, "Daily COVID-19 Cases in Scotland - Daily Case Trends By Health Board - Scottish Health and Social Care Open Data," 2020. [Online]. Available: https://www.opendata.nhs.scot/dataset/covid-19-in-scotland/resource/2dd8534b-0a6f-4744-9253-9565d62f96c2

[25] National Records of Scotland, "Deaths involving coronavirus (COVID-19) in Scotland - National Records of Scotland," 2020. [Online]. Available: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/general-publications/weekly-and-monthly-data-on-births-and-deaths/deaths-involving-coronavirus-covid-19-in-scotland

[26] S. N. Ladhani, J. Y. Chow, R. Janarthanan, J. Fok, E. Crawley-Boevey, A. Vusirikala, E. Fernandez, M. S. Perez, S. Tang, K. Dun-Campbell, E. W. Evans, A. Bell, B. Patel, Z. Amin-Chowdhury, F. Aiano, K. Paranthaman, T. Ma, M. Saavedra-Campos, R. Myers, J. Ellis, A. Lackenby, R. Gopal, M. Patel, C. Brown, M. Chand, K. Brown, M. E. Ramsay, S. Hopkins,

N. Shetty, and M. Zambon, "Investigation of SARS-CoV-2 outbreaks in six care homes in London, April 2020," *EClinicalMedicine*, vol. 26, p. 100533, 9 2020. [Online]. Available: https://doi.org/10.1016/j.eclinm.2020.100533

[27] N. S. Graham, C. Junghans, R. Downes, C. Sendall, H. Lai, A. McKirdy, P. Elliott, R. Howard, D. Wingfield, M. Priestman, M. Ciechonska, L. Cameron, M. Storch, M. A. Crone, P. S. Freemont, P. Randell, R. McLaren, N. Lang, S. Ladhani, F. Sanderson, and D. J. Sharp, "SARS-CoV-2 infection, clinical features and outcome of COVID-19 in United Kingdom nursing homes," *Journal of Infection*, vol. 81, no. 3, 2020. [Online]. Available: https://doi.org/10.1016/j.jinf.2020.05.073

[28] O. Byambasuren, M. Cardona, K. Bell, J. Clark, M. L. McLaws, and P. Glasziou, "Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis," *Journal of the Association of Medical Microbiology and Infectious Disease Canada*, vol. 5, no. 4, pp. 223–234, 2020. [Online]. Available: https://doi.org/10.3138/jammi-2020-0030

[29] E. Dickson, N. E. Palmateer, J. Murray, C. Robertson, C. Waugh, L. A. Wallace, L. Mathie, K. Heatlie, S. Mavin, P. Gousias, B. Von Wissman, D. J. Goldberg, and A. McAuley, "Enhanced Surveillance of COVID-19 in Scotland: population-based seroprevalence surveillance for SARS-CoV-2 during the first wave of the epidemic," *Public Health*, vol. 190, 2020. [Online]. Available: https://doi.org/10.1016/j.puhe.2020.11.014

[30] National Records of Scotland, "Population Estimates Time Series Data," 2019. [Online]. Available: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/population-estimates-time-series-data

[31] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, and G. M. Leung, "Temporal dynamics in viral shedding and transmissibility of COVID-19," *Nature Medicine*, vol. 26, no. 5, pp. 672–675, 2020. [Online]. Available: https://doi.org/10.1038/s41591-020-0869-5

[32] C. McAloon, . Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, P. Wall, M. Green, L. O'Grady, and S. J. More, "Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research," *BMJ Open*, vol. 10, no. 8, pp. 1–9, 2020. [Online]. Available: http://dx.doi.org/10.1136/bmjopen-2020-039652

[33] The Scottish Government, "Coronavirus (COVID-19): modelling the epidemic in Scotland (Issue No. 25)," Scottish Government, Tech. Rep. 29, 2020. [Online]. Available: https://www.gov.scot/publications/coronavirus-covid-19-modelling-epidemic-issue-no-29/

[34] NHS Lothian, "NHS Lothian," 2021. [Online]. Available: https://www.nhslothian.scot/Pages/default.aspx

[35] J. Belmin, N. Um-Din, C. Donadio, M. Magri, Q. D. Nghiem, B. Oquendo, S. Pariel, and C. Lafuente-Lafuente, "Coronavirus Disease 2019 Outcomes in French Nursing Homes That Implemented Staff Confinement with Residents," *JAMA Network Open*, vol. 3, no. 8, pp. 1–9, 2020. [Online]. Available: https://doi.org/10.1001/jamanetworkopen.2020.17533

[36] A.-L. Barabási and E. Bonabeau, "Scale-Free Networks," *Scientific American*, vol. 288, no. 5, pp. 60–69, 2003. [Online]. Available: https://www.scientificamerican.com/article/scale-free-networks/

[37] A. Comas-herrera, "Rapid review of the evidence on impacts of visiting policies in care homes during the COVID-19 pandemic," *Internal long-term care policy network*. [Online]. Available: https://ltccovid.org/wp-content/uploads/2020/11/Rapid-review-of-evidence-on-impacts-of-visiting-policies-in-care-homes-during-the-COVID-pandemic-LSE068110.pdf

[38] P. H. England, "A data linkage approach to assessing the contribution of hospital-associated SARS-CoV-2 infection to care home outbreaks in England, 30 January to 12 October 2020," Tech. Rep. October, 2021. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/983349/Data_linkage_approach_to_assessing_the_contribution_of_hospital-associated_SARS-CoV-2_infection_to_care_home_outbreaks_in_England.pdf

[39] N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, C. A. Pearson, B. J. Quilty, A. J. Kucharski, H. Gibbs, S. Clifford, A. Gimma, K. van Zandvoort, J. D. Munday, C. Diamond, W. J. Edmunds, R. M. Houben, J. Hellewell, T. W. Russell, S. Abbott, S. Funk, N. I. Bosse, Y. F. Sun, S. Flasche, A. Rosello, C. I. Jarvis, and R. M. Eggo, "Age-dependent effects in the transmission and control of COVID-19 epidemics," *Nature Medicine*, vol. 26, no. 8, pp. 1205–1211, 2020. [Online]. Available: https://doi.org/10.1038/s41591-020-0962-9

[40] J. K. Burton, M. McMinn, J. E. Vaughan, J. Fleuriot, and B. Guthrie, "Care-home outbreaks of COVID-19 in Scotland March to May 2020: National linked data cohort analysis," *Age and Ageing*, vol. 50, no. M5, pp. 1–11, 2021. [Online]. Available: https://doi.org/10.1093/ageing/afab099

[41] J. Brainard, S. Rushton, T. Winters, and P. R. Hunter, "Introduction to and spread of COVID-19-like illness in care homes in Norfolk, UK," *Journal of Public Health*, vol. 43, no. 2, pp. 228–235, 2020. [Online]. Available: https://doi.org/10.1093/pubmed/fdaa218

[42] D. Majra, J. Benson, J. Pitts, and J. Stebbing, "SARS-CoV-2 (COVID-19) superspreader events," *Journal of Infection*, vol. 82, no. 1, pp. 36–40, 2021. [Online]. Available: https://doi.org/10.1016/j.jinf.2020.11.021

Chapter 4

Sickness Absence Rates in NHS England Staff during the COVID-19 Pandemic

This chapter is a manuscript produced conjointly with Dr Itamar Megiddo (IM) and Adam Kleczkowski (AK). We used multivariate regression and time series models to investigate sickness absence rates in NHS England staff during the COVID-19 pandemic. This paper will be submitted to BMC Public Health shortly. Each author's contributions are outlined briefly below and shown in Table 4.1.

AK and IM played supervisory roles and provided feedback on the manuscript. The PhD author (EM) worked on; the literature review, writing and editing the manuscript, developing the source code, and on formal analysis and investigation.

Table 4.1: Description of authors' contributions to the manuscript.

Contributor Role	Role Definition	Name
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	AK
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	EM
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	EM
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	AK
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	EM
Methodology	Development or design of methodology; creation of models	AK, EM, IM
Project Administration	Management and coordination responsibility for the research activity planning and execution.	AK, IM, EM
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	AK, EM
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	EM
Supervision	Oversight and leadership responsibility for the research activity planning and, execution including mentorship external to the core team.	AK, IM
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.	EM
Visualisation	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	EM
Writing- Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	EM
Writing- Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.	AK, IM, EM

4.1 Abstract

The COVID-19 pandemic placed immense strain on healthcare systems worldwide, with NHS England facing substantial challenges in managing staff illness-related absences amid surging treatment de-

mands. Understanding the impact of the pandemic on sickness absence rates among NHS England staff is crucial to developing effective workforce management strategies and ensuring the continued delivery of healthcare. In this study, we use publicly available data to investigate the impact of the COVID-19 pandemic on sickness absence rates among NHS England staff between June 2020 and 2022. Our analysis highlights significant increases in sickness absence rates coinciding with the arrival of COVID-19 in England, which continued to rise throughout the pandemic. High periods of COVID-19 activity strongly correlated with staff absence, and the main categories driving the dynamics were COVID-19-related or mental health absences. We demonstrate that sickness absences in these two categories can be estimated accurately using multivariate linear regression and Seasonal ARIMA time series models, respectively. Moreover, we show that indicators of COVID-19 activity (positive tests, hospitalisations, ONS incidence) contain useful information about staff infection pathways. Our findings can inform targeted interventions and policies to reduce sickness absence, improve workforce health, and enhance productivity within NHS England. Additionally, this study contributes to a deeper understanding of the pandemic's impact on healthcare systems, informing strategic workforce management management and public health preparedness.

4.2 Introduction

Sickness absence is a huge social and economic burden to the National Health Service (NHS) England, costing an estimated £1.65 billion each year [1]. NHS England is the largest employer in the UK, with 1.2m full-time equivalent staff as of April 2022 [2], but their monthly average sickness absence rate far exceeds the UK public sector (4.2% in NHS staff [1] vs 2.9% [3], between 2009-2019). Furthermore, sickness absence hampers the provision and management of patient care and is detrimental to the working conditions of the remaining staff who can work. The NHS also incurs additional costs through sick pay and the need to hire agency staff to cover shifts [4].

In many epidemics, healthcare workers (HCWs) have consistently faced an increased risk of occupational infection [5] and the COVID-19 pandemic was no exception. Notably, patient-facing healthcare workers in NHS Scotland were three times more likely to be hospitalised with COVID-19 compared to the general population [6]. A study in NHS England produced a similar finding [5]. Data analysis from Appelby et al. [7] shows a large excess sickness absence in NHS England staff during March-May 2020 compared to the previous ten-year average for each month, coinciding with the first wave of COVID-19 in England. During waves of COVID-19, NHS England faced a twofold crisis: a surge in demand for treatment and an increase in absence due to staff illness. Reports surfaced of insufficient staff to provide adequate care, placing immense pressure on the system [8]. Furthermore, COVID-19 outbreaks recurred in waves not confined to winter

like the NHS was used to with seasonal influenza [9]. These waves exhibited multiple peaks and high troughs, resulting in constant strain on the system with few lulls. Although the introduction of vaccines in early 2021 reduced the likelihood of severe illness [10], the risk of staff contracting disease remained, rendering them unable to work [11].

Emerging evidence suggests the burden on NHS England staff led to absences due to reasons that were not directly COVID-19, such as other respiratory diseases and mental health. In a comparison between 2019 and 2020, Edge et al. [12] found increases (at least initially) in asthma, chest and respiratory disease, infectious diseases and mental illness, but decreases in other categories such as musculoskeletal disorders, injury and fracture, gastrointestinal disease, genitourinary and gynaecological disease and, most notably, cancer. Van der Plaat et al. [13] explored absences in NHS England due to mental health over the same period and found a spike in absences in March-April 2020, which declined to typical levels by May and June. They also found regional correlation between the percentage change in new mental health absences and absences attributed to COVID-19, suggesting an interaction between the two.

Prior to the COVID-19 pandemic, models were developed to predict sickness absence rates while including the effects of seasonal and pandemic influenza. Asghar et al. [14] predicted absence rates (for all reasons) at the workforce level for NHS England. They showed that Autoregressive Integrated Moving Average

(ARIMA) and Seasonal ARIMA (SARIMA) time series models could accurately make 6-month predictions of the overall sickness absence rate in NHS ambulance staff. In particular, using multivariate regression models to estimate sickness absence attributed to influenza (seasonal and pandemic) from proxy variables for influenza activity (e.g., hospitalisations or tests) is an established approach in the literature [15–17]. This paper uses this methodology to estimate COVID-19 sickness absence in HCWs from proxy variables for COVID-19 activity. Ip et al. [15] used a multivariable linear regression model to estimate the excess all-cause and acute respiratory infection-related related sickness absenteeism rates for HCW's in Hong Kong during influenza epidemics between 2004-2009. They highlight how periods of heightened influenza activity are linked to increased sickness absence. Similarly, Schanzer et al. [16] accurately estimated absenteeism from 1998 to 2009 in Canada using multivariate regression models of trend, seasonality and proxy variables for influenza activity. Linear regression models, stratified by pay status, were developed by Challener et al. [17] for unscheduled HCW absences as a function of ILI prevalence in Minnesota.

Another notable approach in the literature was predicting the risk and duration of future sickness absence for individuals based on absence history and other employee characteristics/traits (particularly from health checkups). For example, Gémes et al. [18] developed piecewise hazard regression (survival analysis) models to healthcare professionals predict the duration of sickness absence due to stress-related disorders. Similarly, Laaksonen et al. [19] used proportional

hazards models and data from government employees in Helsinki to show that previous absence episodes increase the risk of future sickness absence, and that the risk of recurrring sickness absences is higher for longer sickness-absence spells.

There is a need for further investigation to understand the long-term effects of the pandemic on sickness absence rates. Building upon the research by Appleby et al. [7], Edge et al.[12], and Plaat et al. [13], we aim to investigate the impact of COVID-19 on sickness absence rates in NHS England from late 2020 into 2022. In this paper, we develop multivariate regression and time series predictive models using publicly available data on sickness absence rates in NHS England. Our research addresses several key questions. Firstly, what were the key trends and sources of variability in NHS England Sickness absence rates during this period? Secondly, can we use this information to develop models to explain the sickness absence rates? Lastly, what is the relationship between indicators of COVID-19 activity (e.g., PCR tests) and these sickness absence rates? Understanding sickness absence trends and future workforce availability is critical for effective staff and resource planning in healthcare. Furthermore, understanding the relationship between Covid-19 and healthcare staff's availability to work has multiple benefits such as informing strategies for future waves and exploring the effects of different policy interventions [20].

This paper is structured as follows. In Section 4.3 we describe the data and methods deployed to analyse it. The corresponding results are given in Section 4.4 and discussed in Section 4.5 where we give our concluding remarks.

4.3 Methods

We obtained data on absences in NHS England staff from the NHS England's Sickness Absence Rates publication. This contains monthly observations of sickness absence rate broken down by reason from January 2015 until April 2022, taken directly from the electronic staff record. The sickness absence rate is defined as the ratio between the "full-time equivalent (FTE) number of days lost" and "FTE number of days available" [2]. We define COVID-19-related sickness absence as sickness absence in any of three diagnostic categories/reasons; S13 cold/cough/flu, S15 chest and respiratory problems, and S27 infectious disease. This is similar to how an earlier NHS report defined COVID-19 sickness absence - sickness absence in any of five diagnostic categories (cough/flu, chest/respiratory, infectious diseases, other, unknown) with COVID-19 recorded as a related reason [21].

We use COVID-19 community incidence and positivity estimates from the Office for National Statistics Coronavirus (COVID-19) Infection Survey [22]. The incidence rate estimates the number of new PCR-positive infections per day, and the positivity rate estimates the number of people who would test positive for

COVID-19 each day. They reflect COVID-19 infections of people living in private households (general population households and households of NHS staff) but not patients in hospitals or care homes. The daily confirmed PCR-positive COVID-19 tests reported for England were taken from the UK Health Security Agency (UKHSA) Coronavirus Dashboard [23]. The daily number of new patients admitted to hospitals with COVID-19 in England is from the same source. This data counts people admitted to hospitals who tested positive for COVID-19 14 days before admission or during their stay in the hospital. Inpatients diagnosed with COVID-19 after admission are reported as being admitted the day before their diagnosis [24]. We converted all surveillance data to monthly observations for comparability to the sickness absence data. The bi-weekly ONS incidence estimates were first interpolated to get daily estimates, then summed for a monthly one. Similarly, we estimated the average number of people testing positive each day in a given month by interpolating the bi-weekly estimates of positivity and taking an average. July 2020 is the first month with an observation from all three sources of COVID-19 surveillance data, Figure 4.1.

To investigate the relationship between indicators of COVID-19 activity and COVID-19-related sickness absence rates, we developed univariate and multivariate regression models from combinations of COVID-19 surveillance data through an exhaustive selection process. Statistical significance was determined by considering a type I error probability below 5% ($\alpha < 0.05$), and we compare the fits of these models and their predictive power using adjusted R^2 and Akaike

information criterion (AIC). We included the following covariates: the number of PCR-positive COVID-19 tests, the number of new COVID-19 hospitalisations, and the community incidence of COVID-19. We expected these indicators of COVID-19 severity to correlate positively with COVID-19-related sickness absence. We train the regression coefficients for the models using data from July 2020 until December 2021. Using the trained models we estimate the COVID-19 sickness absence trend between January 2022 and July 2022 and compare to the observed data.

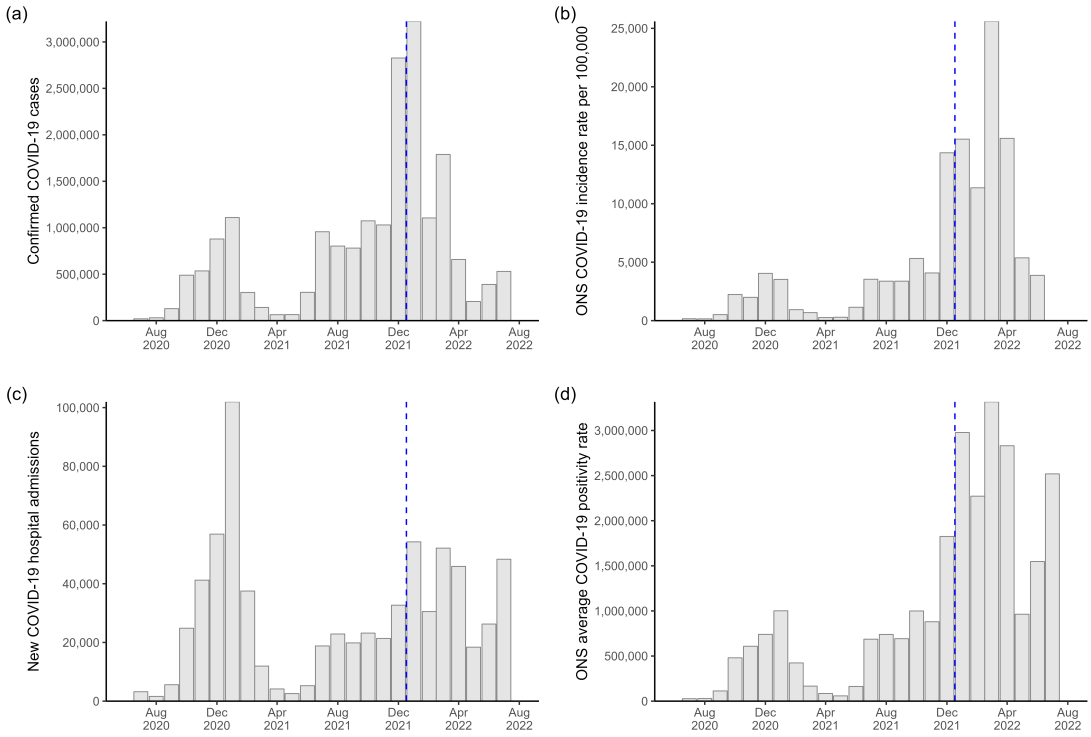

Fig. 4.1: COVID-19 surveillance data for England. Monthly observations where months before the dashed blue vertical line were used for developing the regression models. (a) confirmed PCR-positive COVID-19 tests; (b) ONS estimated new COVID-19 infections per 100,000; (c) new COVID-19 hospitalisations; (d) ONS average testing positive for COVID-19 each day. Note that ONS data were not published before July 2020 and the ONS stopped publishing incidence estimates in June 2022 [22].

Furthermore, we investigate the impact of COVID-19 on mental health-related sickness absence (recorded as S10 anxiety/stress/depression/other psychiatric illnesses). We developed a deterministic time series model trained on pre-COVID-19 (January 2015-March 2019) data and used it to predict the post-COVID-19 trend. We fit a seasonal autoregressive integrated moving average (SARIMA) model to the historic sickness absence time series using the Auto.arima function

from R's "timetk" library [25]. We also fit a time series model including some post-COVID-19 months (January 2015-december 2021). We then forecast absence rates for the first six months of 2022 and compare these estimates to the first three months of observed data.

All analyses were performed in Rstudio version 4.2.1 [26].

4.4 Results

We begin with a high-level exploration of the sickness absence trends between 2015 and 2022, before investigating the reasons for absence in further detail. We then assess the relationship between COVID-19 and sickness absence rates using regression models built from COVID-19 surveillance data. Furthermore, we analyse mental health-related absence using Seasonal Autoregressive Integrated Moving Average (SARIMA) time series models, similar to [14], to highlight the impact of COVID-19.

4.4.1 Workforce level absence trends

During 2015-2019 the overall sickness absence rate for NHS England followed a sinusoidal trend, with peaks in winter (January/December) and troughs in summer (May/June), Figure 4.2. The absence rate ranged between 3.67 and 5.01%, with 4.19% of the workforce absent each month on average. Therefore, in

February 2020, we expect the absence rate to be decreasing. Instead, there was a massive surge in the overall sickness absence rate. This surge peaked at 6.19% in April, the highest sickness absence rate recorded in a month since 2015, before decreasing to typical levels by July. The sickness absence rate rose towards the end of 2020 and peaked in January 2021 at 5.74%: an all time high. It decreased towards a typical 3.97% in March 2021 but since then has continuously climbed upwards. From July 2021 until March 2022, the sickness absence rate for each month was consistently at least 1% higher than the previous year. A new record high of 6.68% was reached in the winter peak of January 2022. Aside from some months in the summer of 2020 (May, June, July, August) and recently in 2022 (August, September), sickness absence rates from February 2020 onwards have typically risen year on year. The average sickness absence rate for each year between 2020 and 2022 increased by roughly 0.45%.

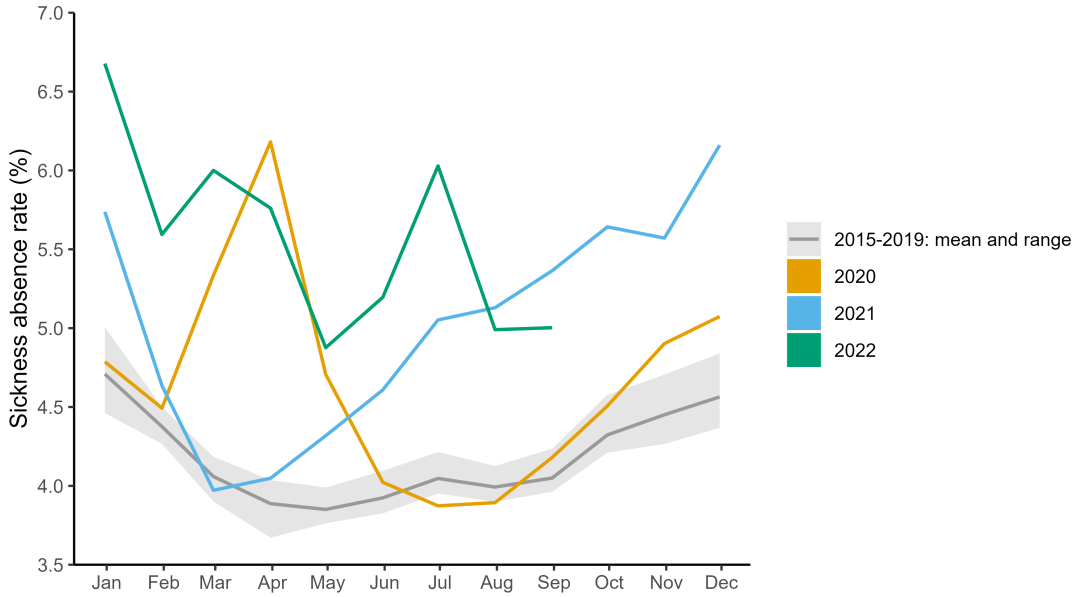

Fig. 4.2: Sickness absence rates for NHS England staff by month. The yellow, blue and green lines indicate the overall sickness absence rates by month for 2020, 2021, and 2022 respectively. The dark grey line shows the mean sickness absence rate for a fixed month in 2015-2019, and the light grey region highlights the minimum and maximum rate during this period.

The strong seasonality in sickness absence rates (peaks in winter, troughs in summer) are driven by four categories of absence: cold/cough/flu, chest and respiratory problems, gastrointestinal problems, and anxiety/depression/other psychiatric illness, Figure 4.3.

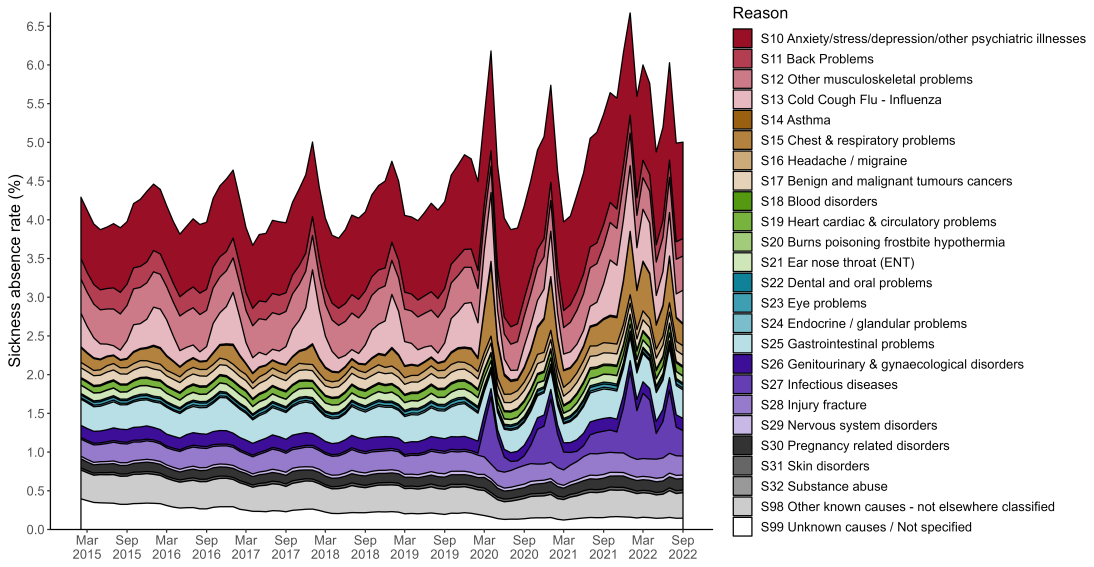

Fig. 4.3: Sickness absence rates for NHS England staff by month and reason. The main reason behind a member of staff's sickness absence was recorded in the electronic staff record. This timeseries shows the monthly observations of sickness absence rates from January 2015 until the end of March 2022. Each colour highlights the proportion of the overall monthly sickness absence rate attributed to each main reason for absence.

Our expectation that the indicators of COVID-19 severity, the COVID-19 surveillance data, will be positively correlated with the absence rate is hinted at by the surge in absences in March 2020 in Figure 4.2. Figure 4.4 shows that reasons for absence in the COVID-19 category (cold/cough/flu, chest/respiratory, and infectious disease) were responsible for the surge in sickness absence rates in February 2020. Absence rates for these reasons then returned to their previous sinusoidal cycle but were amplified in magnitude, explaining the winter 2020 peak seen in Figure 4.2. The low of summer troughs and the duration and peak of winter waves of sickness absence rates were significantly higher, Fig-

ure 4.4 and Figure 4.5. In particular, the average sickness absence rate for these categories (cold/cough/flu, chest/respiratory, and infectious disease) more than doubles when we compare 2015-2019 to 2020-2022 (0.49% vs 1.24%).

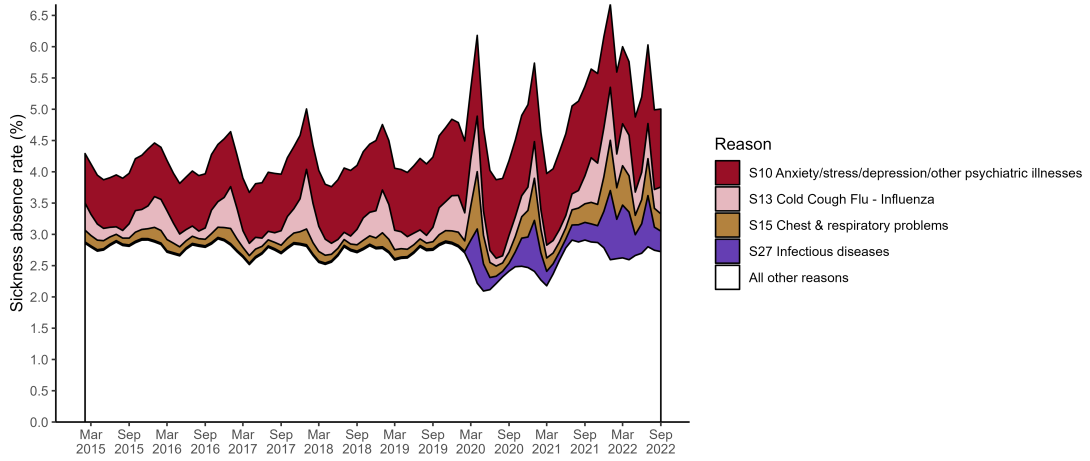

Fig. 4.4: Sickness absence rates for NHS England staff by month and reason: grouped. The main reason behind a member of staff's sickness absence was recorded in the electronic staff record. This timeseries shows the monthly observations of sickness absence rates from January 2015 until the end of March 2022. Each colour highlights the proportion of the overall monthly sickness absence rate attributed to either cold/cough/flu (pink), chest and respiratory problems (dark yellow), infectious disease (purple), anxiety/stress/depression/other psychiatric illnesses (red), or any other reason (white).

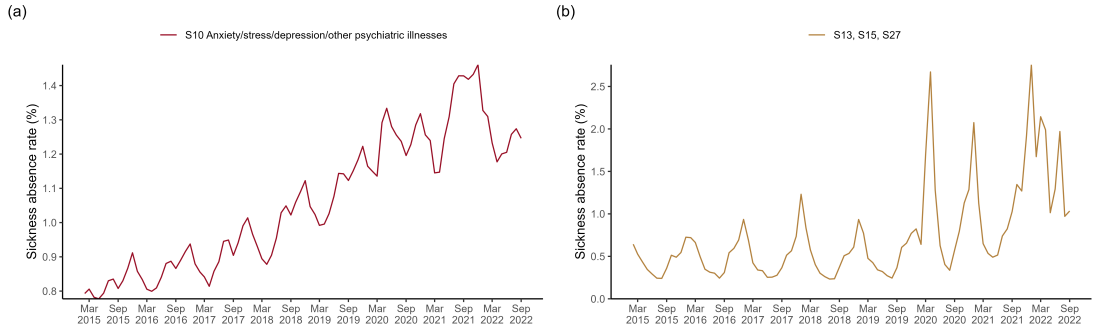

Fig. 4.5: Sickness absence rates for NHS England staff by month and specific reasons. (a) mental health-related absence; (b) absence for cold/cough/flu, chest and respiratory problems, or infectious disease.

S10 anxiety/depression/other psychiatric illness (mental health) was consistently the main reason for staff absence from January 2015 to March 2022. It accounted for 20% of the staff absences on average each month, Figure 4.4. Furthermore, sickness absence for mental health reasons has increased yearly since 2016, with bi-annual peaks in July and December, Figure 4.5 (a). This trend appears to be slightly disrupted after March 2020, with the July peaks of 2020 and 2021 being high relative to the December ones, and a deep trough in April 2021. Furthermore, Figure 4.5 (a) and (b) demonstrate clearly that the huge rising rise in absences in mid-2021 (Figure 4.2) was driven by a combination of COVID-19 related and mental health-related sickness absence.

4.4.2 COVID-19 related absence

In this section, we describe the relationships between the indicators of COVID-19 activity and the COVID-19-related sickness absence data according to the regression models. We investigate the performance of the regression models over the training period (July 2020 until December 2021), Table 4.2, and then the estimated COVID-19 sickness absence trend between January 2022 and July 2022 according to trained models.

Table 4.2: Regression models of the COVID-19 related sickness absence rate. Coefficients were estimated using data between July 2020 and December 2021. Each numbered column in the table indicates a different model. The rows in the first section contain the regression coefficients (top) and their corresponding standard error (bottom), with the significance of the coefficient indicated by the number of asterisks.

	Dependent variable:						
	Total sickness absence rate in the S13, S15 and S27 categories						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
New hospital admissions	1.68e-05***				1.19e-05***	9.89e-06***	9.68e-06***
	(2.78e-06)				(1.51e-06)	(1.45e-06)	(1.79e-06)
PCR positive tests		5.93e-07***		−7.4e-07*	4e-07***		−5.04e-08
		(1.09e-07)		(3.31e-07)	(5.54e-08)		(2.33e-07)
Average positivity - ONS			9.33e-07***	1.96e-06***		6.41e-07***	7.17e-07.
			(1.24e-07)	(4.74e-07)		(7.63e-08)	(3.63e-07)
Constant	0.541***	0.566***	0.443***	0.360***	0.401***	0.361***	0.357***
	(0.095)	(0.100)	(0.087)	(0.086)	(0.050)	(0.046)	(0.051)
Observations	18	18	18	18	18	18	18
Adjusted R^2	0.676	0.626	0.767	0.814	0.923	0.939	0.935
Akaike Inf. Crit.	9.523	12.090	3.539	0.354	−15.452	−19.820	−17.880
F Statistic	36.404***	29.441***	57.070***	38.187***	102.437***	132.634***	82.818***
	(df = 1; 16)	(df = 1; 16)	(df = 1; 16)	(df = 2; 15)	(df = 2; 15)	(df = 2; 15)	(df = 3; 14)

Note: . $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Observed COVID-19 sickness absence was low in July 2020 (0.5%) and then rose steadily towards a peak of 2.2% in January 2021, before dropping back to 0.5% by May 2021, Figure 4.6 (b). The trend between June 2021 and December 2021 was a staggered rise. Absences started to increase in June 2021, reaching

1.5% by October, then increasing from around 1.5% in November to 2% by December. The rise peaked in January 2022 at 3%. They then dropped to 1.8% in March, increasing slightly in April 2022 (2%), and continued to fall to 1% by July 2022.

We fit a univariate regression model to estimate the staff sickness absence rate based on the number of positive PCR COVID-19 tests, Figure 4.6. The number of PCR positive tests was a significant predictor of the NHS staff sickness absence rate ($F(1,16)=29.4$, $p< .0001$), with an adjusted $R^2 = 62.6\%$ and AIC = 12.1. There was a positive relationship between the estimated number of positive PCR COVID-19 tests and the absence rate, Figure 4.6 (a). Absences started to increase in June 2021, reaching 1.5% by October, then increasing from around 1.5% in November to 2% by December. The univariate model captures the general trend in absences between July 2020 and December 2021 reasonably well, Figure 4.6 (b). However, it does a poorer job of estimating the magnitude of the waves. It slightly overestimates the low absences (a trough) in July 2020. The model underestimates the magnitude of the peak in absences in February 2021 and estimated a peak, drop and steady rise from July 2021, when there was a smoother resurgence. The model matches this rising trend towards December 2021 well, Figure 4.6 (b). The observed rising COVID-19 sickness absence trend peaks in January 2022 at 3%. It then falls to 1.5% in March, slightly increases in April, and then drops towards 1% by June 2022. The trained univariate model

explains the peak in January 2022 well; however, it estimates a larger drop in absences than observed in March 2022. This causes a larger underestimation of absences between March and July 2022.

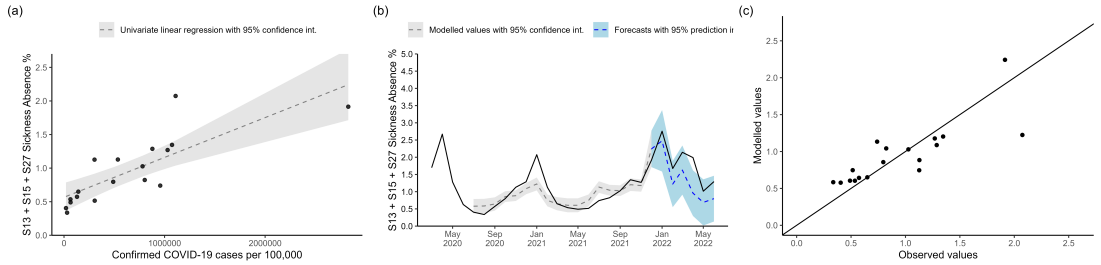

Fig. 4.6: COVID-19 related absence as univariate model of confirmed PCR-positive COVID-19 tests. (a) Scatterplot of tests against the absence rate, including the fitted regression line. (b) Timeseries of modelled values (dashed grey) with 95% confidence interval (grey, shaded), and observed sickness absence trend (solid black). (c) Scatter plot of modelled values against the observed values. The black line is the theoretical line of equality (modelled=observed).

We fit a univariate regression model to estimate the staff sickness absence rate based on the estimated number of new COVID-19 infections according to the ONS, Figure 4.7. The estimated number of new COVID-19 infections, according to the ONS, was a significant predictor of the NHS staff sickness absence rate ($F(1,16)=29.4$, $p<.001$), with an adjusted $R^2 = 51.4\%$ and AIC = 16.8. There is a positive relationship between the estimated number of new COVID-19 infections according to the ONS and the absence rate, Figure 4.7 (a). Similarly to the univariate model of tests, this model captures the general trends in absences well, Figure 4.7 (b). Furthermore, it also does a poorer job of estimating the magnitude of the waves. The model overestimates the low absences (a trough) in July 2021. It also underestimates the magnitude of the peak in absences in

February 2021 before overestimating and then underestimating the trough into a steady rise between June and october 2021. The model reasonably matches the rising trend towards the end of 2021, Figure 4.7 (b). This rising COVID-19 sickness absence trend peaks in January 2022 at 3%. It then falls to 1.5% in March, slightly increases in April, and then drops towards 1% by June 2022. Using the trained model of ONS estimated incidence results in a slight overestimation of the January 2022 peak and a large overestimation of the March/April 2022 rise (3.5% vs 2%).

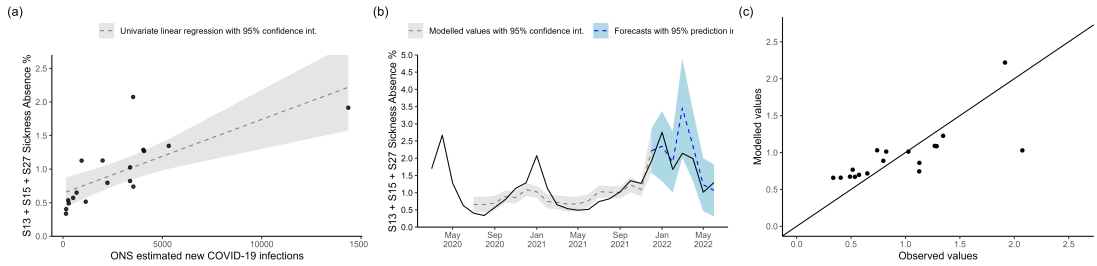

Fig. 4.7: COVID-19 related absence as univariate model of the ONS estimated COVID-19 incidence. (a) Scatterplot of tests against the absence rate, including the fitted regression line. (b) Timeseries of modelled values (dashed grey) with 95% confidence interval (grey, shaded), and observed sickness absence trend (solid black). (c) Scatter plot of modelled values against the observed values. Black line is theoretical line of equality (modelled=observed).

We fit a univariate regression model to estimate the staff sickness absence rate based on the number of new COVID-19 hospitalisations, Figure 4.8. The number of new COVID-19 hospitalisations was a significant predictor of the NHS staff sickness absence rate ($F(1,16)=36.4$, $p<.001$), with an adjusted $R^2 = 67.6\%$ and AIC = 9.5. There is a positive relationship between the number of hospitalisations and the absence rate, Figure 4.8 (a). The univariate model explained the wave

of absences in December 2020 - February 2021 very well, both in the timing of waves and magnitude of the peak and troughs, Figure 4.8 (b). However, the model estimates only a slight rise and stagnation of absences between September 2021 - December 2021 (late 2021) when there was a steady resurgence, Figure 4.8 (b). This results in a large underestimation of December 2021 (1.7% vs 1%). Figure 4.8 (c) highlights those few months in late 2021 where the model underestimates the absences; otherwise, the model performs well. The univariate model trained using data between June 2020 and December 2021 significantly underestimates the peaks and troughs in the sickness absence trend between January and July 2022. However, it picks up when the gradient of the observed trend changes. For example, the observed COVID-19 sickness absence rising trend in late 2021 peaks in January 2022 at 3%. It then falls to 1.5% in March, slightly increases to 2% April, and then drops towards 1% by June 2022. However, the trained univariate model estimated a rise to only 1.4% by January 2022, then fluctuations between $1 - 1.5\%$ that meet the observed 1% by July 2022.

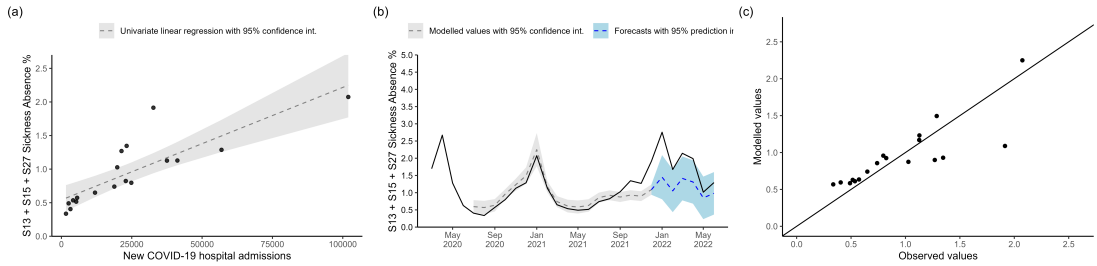

Fig. 4.8: COVID-19 related absence as univariate model of new COVID-19 hospitalisations. (a) Scatterplot of tests against the absence rate, including the fitted regression line. (b) Timeseries of modelled values (dashed grey) with 95% confidence interval (grey, shaded), and observed sickness absence trend (solid black). (c) Scatter plot of modelled values against the observed values. Black line is the theoretical line of equality (modelled=observed).

There are two outliers Figure 4.8. One where the hospitalisation rate is >100,000 and the absence rate is 2%, and another where hospitalisation is closer to 35,000, and the absence rate is 2%. The outliers suggest a problem with scale and could be skewing the linear relationship, causing the poorer fit to absences in late 2021. Alternatively, it may suggest a threshold in hospitalisations where further increase does not impact the absence rate. Furthermore, some evidence of heteroscedasticity in Figure 4.8 (a), suggests right skewness of the surveillance data. However, log transformations of the surveillance data did not improve the model fits (results not shown).

We fit a univariate regression model to estimate the staff sickness absence rate based on the estimated COVID-19 posivity rate according to the ONS, Figure 4.9. The estimated number of new COVID-19 infections, according to the ONS, was a significant predictor of the NHS staff sickness absence rate ($F(1,16)=57.7$, $p<.001$), with an adjusted $R^2 = 76.7\%$ and AIC = 3.5. A positive relationship exists between the estimated COVID-19 posivity rate according to the ONS and the absence rate, Figure 4.9 (a). The model matches the trend in absences well, Figure 4.9 (b). The univariate model slightly underestimates the magnitude of the peak in absences in January 2021 (1.5% vs 2.2%) and estimates a small rise when the absences stayed low (a trough) in July 2021 (1.4% vs 1%). However, it picks up the rest of the behaviour almost perfectly, including the rising trend between September and December 2021, Figure 4.9 (b).

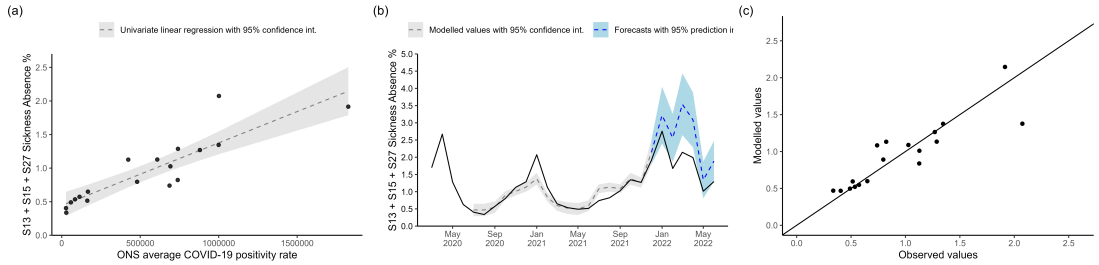

Fig. 4.9: COVID-19 related absence as univariate model of the ONS estimated average COVID-19 positivity rate. (a) Scatterplot of tests against the absence rate, including the fitted regression line. (b) Timeseries of modelled values (dashed grey) with 95% confidence interval (grey, shaded), and observed sickness absence trend (solid black). (c) Scatter plot of modelled values against the observed values. The black line is the theoretical line of equality (modelled=observed).

Across the univariate models, the COVID-19 positivity rate according to the ONS was the strongest predictor of absence (AIC = 3.5, adj. R^2 = 76.7%), Table 4.2. The number of new COVID-19 hospital admissions outperformed the number of positive PCR COVID-19 tests (AIC = 9.5, adj. R^2 = 67.6% vs AIC = 12.1, adj. R^2=62.3%). The univariate model of hospitalisations explained the wave of absences in December 2020 - February 2021 well in magnitude and timing, whereas the other two predictors (comparing Figure 4.8 (b) to Figure 4.7 (b) and Figure 4.6 (b)) underestimated this wave. However, the magnitude of the underestimation was much smaller for the ONS predictor. These other two predictors (estimated number of new infections according to the ONS, PCR positive tests) better estimate the wave of absences in late 2021 (post-September). The univariate model of hospitalisations underestimates absences during this time. The univariate models of tests and estimated ONS estimate a similar absence trend, Figure 4.7 (b) and Figure 4.6 (b), which suggests these data sources con-

tain similar information about absences.

The multivariate models combining hospitalisations with either tests (F(2, 15) = 102.44, p<.001) (AIC = -15.5, adj R^2 =92.3%) or ONS positivity (F(2, 15) = 132.63, p<.001) (AIC = -19.8, adj R^2 =93.9%) generate a significant improvement in prediction error and adjusted R^2, compared to the best univariate model (ONS positivity; AIC = 3.5, adj. R^2 = 76.7%), Table 4.2. Whereas the multivariate model of tests and ONS estimated positivity (F(2, 15) = 38.187, p<.001) provides only a slight improvement in predictive power (reduction in prediction error) (AIC = 0.354, adj R^2 = 81.4%), Table 4.2. These also suggest that the testing and ONS data stream contain similar additional information related to paths of staff infection (and therefore absence) not captured by the hospitalisation data stream alone.

Furthermore, the multivariate model containing all three predictors (F(3, 14) = 82.82, p<.001) has a slightly higher AIC (poorer prediction error) and lower adjusted R^2 (AIC = -17.847, adj R^2 = 93.5%) compared to the simpler multivariate model that includes hospitalisations and the ONS estimated positivity. This, in combination with I) the models combining hospitalisations with either tests or ONS performing similarly and II) the model of ONS and tests performing worse, suggests that the testing and ONS data streams contain similar information about absences.

We investigated whether there was multicollinearity in the multivariate regression models using the variance inflation factor (VIF). The variance inflation factor was low in the models combining hospitalisations with ONS or positive tests (roughly 1.2 for each predictor). In contrast, it was high for the ONS and positive test predictors (>30) when they were included in the same model, while the VIF for the hospitalisations predictor was moderately low when it was included (2.4). This suggests little correlation between hospitalisations and positive tests or ONS but a high correlation between ONS and positive tests.

The multivariate model of ONS and hospitalisations was the best performing model, slightly outperforming the model of hospitalisations and tests (AIC = -19.82, adj R^2 = 93.9% vs AIC = -15.45, adj R^2 = 92.3%) and the multivariate model with all three covariates (AIC = -17.88, adj R^2 = 93.5%). It captures the trend in absences (June 2020-December 2021) very well overall, Figure 4.10 (a), particularly in the timing and magnitude of peaks and troughs. The winter 2020 wave fits well. However, the model estimated that during June-December 2021, there would be a rise, constant level, then a rise again in absence rates when instead there was a steady rise in absences. As a result, the rise in absence rates was slightly overestimated in June/July 2021 and underestimated in November 2021. This is evident in figure Figure 4.10 (b), where some observed absence rate values between 0.6% and 1.5% are over- and under-estimated.

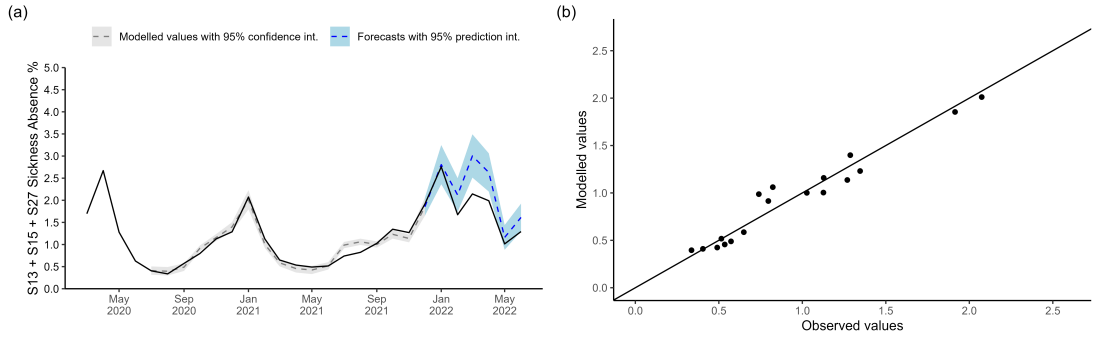

Fig. 4.10: COVID-19 related absence as a multivariate regression model of new hospitalisations and ONS estimated COVID-19 incidence (a) Timeseries of modelled values for 2020-2022 (dashed grey) with 95% confidence interval (grey, shaded), predictions for 2022 (dashed blue) with 95% prediction interval (blue, shaded), and observed sickness absence trend (solid black). (c) Scatter plot of modelled values against the observed values. Black line is theoretical line of equality (modelled=observed).

We used the multivariate model of ONS and hospitalisations to estimate COVID-19 sickness absence rates between January and July of 2022, Figure 4.10. The model matches the peak of absence in January 2022 (3%) and that February would drop, however it severely overestimated the rise in March. The model estimated that COVID-19-related absence rates would be at an all-time high in March, when instead, they were 1% lower - and lower than in January 2022. The model captures a fall from April into July where the model and observed rate coincide again at 1%.

4.4.3 Mental health-related absence

A seasonal ARIMA(0, 1, 3)(0, 1, 0)[12] model fit the mental health-related absence rates from January 2015 until March 2020 best (minimizing AIC), Figure 4.11 (a). The AIC = −238.46 and model equations are shown below

$$(1 - B)(X_t - X_{t-12}) = \left(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^2\right) \varepsilon_t.$$

Where $\theta_1 = -0.378$, $\theta_2 = 0.489$, $\theta_3 = -0.773$ and B is the backshift operator, such that $BX_t = X_{t-1}$. This model has a moving average (MA) part of order 3 with a 1st-order difference and a 12-month period. The next month's absence rate is a linear combination of the previous month, the same month in the previous year, and the month prior from the previous year, plus a new white noise term and the last three months' noise terms.

We used this model to forecast mental health-related absence from March 2020 until September 2022. In April and May 2020, there was a spike in mental health-related absences following the first wave of COVID-19 in England. The seasonal ARIMA model trained on pre-COVID-19 data did not pick up this spike, instead estimating a slower rise in absences that would peak in July 2020, typical of previous years. Excluding these two early months, the model explains the increasing trend with seasonal peaks in from July 2020 until November 2021 well, with estimates falling within 95% intervals, Figure 4.11 (b). However, the model overestimates the January 2022 peak and estimates a typical rise in absence rates

in March/April 2022 when instead there was a deeper drop than usual after the winter peak, Figure 4.11 (a). The model continues to significantly overestimate absence rates through the summer of 2022 until the last data point in September.

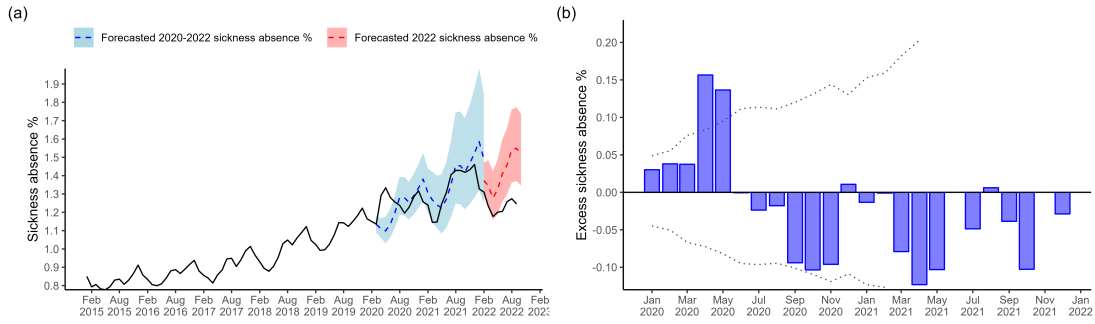

Fig. 4.11: Seasonal ARIMA models of mental health-related sickness absence. (a) Time series showing the observed trend (solid black line) and predictions. We used an ARIMA model trained with 2015-2019 to predict the 2020-2021 trend (dashed blue line, with the light blue shaded region indicating prediction interval). Another model was trained with 2015-2021 data to predict the 2022 rates (dashed red line, with the light red shaded region indicating prediction interval). (b) Barplot showing the difference between the observed trend and the predictions from the ARIMA model trained on 2015-2019 data (shaded blue bars). Bars above the horizontal zero line (solid black) are months where the observed trend was higher than the model predictions. The dotted line indicates the difference between the observed trend and the 95% prediction interval.

We trained a time series model to include the months between March 2019 - December 2021, and a seasonal ARIMA(2, 0, 0)(0, 1, 1)[12] with draft fit best with AIC = −307.85. The equation is shown below

$$\left(1-\phi_1 B-\phi_2 B^2\right)\left(X_t-X_{t-12}\right)=\left(1-\theta_1 B^{12}\right)\varepsilon_t+d.$$

Where $\phi_1 = 1.046$, $\phi_1 = -0.301$, $\theta_1 = -0.5271$, and $d = 0.0065$. This model has non-seasonal autoregressive (AR) part order 2 with no differencing, and a seasonal part with AR order 1, MA order 1 and a 1st and 12th order difference. The next month's absence rate is a linear combination of the last two previous months, the same month in the previous year, and the two months prior from the previous year, plus a new white noise term and the noise terms for that month in the previous year.

The observed mental health sickness absence trend for January-March 2022 falls within this model's 95% prediction interval, suggesting an intial good fit. The model captures the decreasing trend but slightly overestimate absences for each month, Figure 4.11. However, the model also estimates a rise in absence rates in March/April 2022 when instead, there was a deeper drop than usual after the winter peak, Figure 4.11 (a). The model then to overestimates absence rates through the summer of 2022 until the last data point in September, with May-September falling outside the 95% prediction interval.

4.5 Discussion

The COVID-19 pandemic has significantly impacted various aspects of society, including healthcare systems and workforce dynamics. In the context of sickness absence rates in NHS England staff, the pandemic has introduced new challenges and considerations. In this study, we identified the trends and sources of vari-

ability over this period and developed multivariate regression and time series predictive models to explain the absence rates. Our findings can inform the development of targeted interventions and policies to reduce sickness absence and improve workforce health and productivity.

Our analysis revealed a notable increase in sickness absence rates starting around March 2020, which coincided with the establishment of COVID-19 in England. Additionally, we observed another surge in absence during late 2020, and since mid-2021, the levels have remained consistently high. Consequently, except for a few months in 2020 and 2021, the overall sickness absence rate has progressively risen since March 2020. To illustrate, the average monthly absence rate between 2015 and 2019 stood at 4.19%, whereas between 2020 and 2022, this figure increased by approximately 0.45%. As a result, each month's absence rates are at least 1% higher in 2022 compared to pre-pandemic levels. The primary contributors to this variation in sickness absence rates, notably during the surge around March 2020, were the mental health and COVID-19-related categories. We conducted further analysis to evaluate the relationship between these categories of sickness absence and COVID-19 and describe our findings in the following paragraphs.

To examine the impact of COVID-19 on mental health-related sickness absence, we employed a deterministic time series model that utilized pre-March 2020 data for extrapolation. An unexplained deviation from the model was observed during

the initial shock in sickness absence in March and April 2020. Apart from this period, the deterministic model, characterized by an increasing trend with bi-annual peaks, effectively accounted for the overall trend between May 2020 and November 2022. These findings suggest that COVID-19 likely played a role in causing a sudden surge or shock in mental health-related sickness absence among NHS staff during the first wave of the pandemic.

Supporting this, Van der Plaat et al. [13] demonstrated a regional correlation between the increase in mental health sickness absence in March-April 2020, compared to 2019, and the cumulative prevalence of COVID-19-related sickness absence during the same period. The authors hypothesized that the spike in mental health-related sickness absence was driven by the combined stresses experienced by healthcare workers, both at work and in their personal lives, as a consequence of the epidemic.

Furthermore, our time series model accuracy also diverges between January 2022 and April 2022, where the model severely overestimates the absence rate. This underestimation suggests a change in dynamics, and one interpretation is that the continuous rise in mental health absences from 2016 was beginning to plateau by 2020. However, the arrival and constant pressure from COVID-19 had knock-on effects on the well-being of staff and caused a significant increase in mental health absences. Then in 2022, these alleviated somewhat.

It is important to stress that we have not explicitly linked the indicators of COVID-19 activity to mental health-related absence. How to define the relationship and incorporate a shock function to reflect the pressure on NHS staff is unclear. A step for future investigation could involve incorporating COVID-19 activity indicators as regressors in the time series model, thus providing further insight into the relationship between COVID-19 and mental health-related absences.

We developed regression models with COVID-19 activity indicators as predictors to understand the relationship between COVID-19 and NHS staff absences in the COVID-19-related category (S13, 15, 27). The indicators of COVID-19 activity correlated positively and strongly with staff absence, but the relationship was dynamic. The multivariate model, in particular, could explain most of the variability in absence rates in these categories, suggesting COVID-19 was the main driver behind them. Furthermore, the results highlighted two sources providing independent information about the sources of infection for NHS staff, hinting at the dynamics behind NHS staff becoming ill due to COVID-19.

The estimated COVID-19 positivity, according to the ONS, was the strongest univariate predictor of COVID-19 related sickness absence over the July 2020 to December 2021 time period, followed by the number of new COVID-19 hospitalisations, then the number of PCR-positive tests. This suggests that the ONS positivity data stream contained the most information about how NHS staff caught

COVID-19 infections (becoming ill and absent) and that this source is essential to estimate absences. This result supports Zheng et al. [27], who showed that COVID-19 rates in NHS staff mainly rose and declined in parallel with the number of community cases. The ONS data reflects the COVID-19 positivity rate of people in private households, which includes the households of NHS staff [22]. Therefore, it is intuitive that it correlates so strongly with COVID-19 sickness absence. However, as a univariate model of this data stream alone does not best explain the staff absence rate, which is essentially a proxy for staff infections, this supports previous findings that staff face more significant infection pressure than a random member of the general population (living in private households) [10] [27]. Furthermore, the reported PCR-positive tests/cases include people who are (or will end up) hospitalised [23]. They are likely not the strongest predictor between March 2020 and December 2021 because the policy around groups tested and the testing capacity changed over this period [28]. Tests may add slightly different information than the ONS estimated positivity to estimate sickness absence since they were widely reported and could influence behaviour.

Our results suggest that the ONS estimated positivity and the positive test data streams contain similar information about COVID-19-related sickness absence in NHS staff. First, the univariate models of tests and estimated ONS estimate a similar absence trend. Second, the multivariate models combining hospitalisations with either the ONS or testing stream perform similarly, and including the third predictor does not add to the performance. Furthermore, the vari-

ance inflation factor was severely high for these predictors when included in the same model, suggesting a strong correlation. We hypothesise that the ONS/tests predictor provides information about absences due to infection pressure in the community.

The number of new COVID-19 hospitalisations was the second strongest univariate predictor of COVID-19-related sickness absence over the July 2020 and December 2021 period. COVID-19 hospital admissions are a direct source of infection for NHS staff since they primarily work in hospitals and will experience nosocomial infection /infection pressure there. Furthermore, a surge in hospitalisations should correlate with (can be driven by) a large number of infections in the community, another source of infection. This predictor may indirectly capture this path. In previous work using multivariate regression to link influenza activity to sickness absence, Schanzer et al. [16] used laboratory-confirmed H1N1/2009 hospital admissions as a proxy for influenza activity instead of the number of laboratory-confirmed cases overall. Although the scale of COVID-19 has far exceeded seasonal influenza, there is a parallel.

We found that a univariate regression model of new COVID-19 hospital admission better estimated the December 2020 – February 2021 wave. However, the positive tests or ONS-estimated infections better estimated the September 2021 – December 2021 wave. One possible contributing factor to this dynamic is the rollout of COVID-19 vaccines (began in December 2020, and by March and

September 2021 ≈ 30% and ≈ 70% of adults had their first dose [29]) reduced the number of people hospitalised (or the number of severe illnesses) relative to the number infected [30]. The rise in milder but more transmissible sub-variants of COVID-19 may have had a similar effect [31, 32]. The alpha variant was most common between 8th December 2020 to 17th May 2021. Then the slightly more transmissible but equally severe delta variant was most common from 17th May to 19th December 2021, and the far more transmissible but less severe Omicron variant dominated from 20th December 2021 onwards [33]. Furthermore, suppose we assume admissions are a proxy for the infection pressure in the workplace for most NHS staff (hospitals). Vaccines reducing the number of admissions reduces the infection pressure from this source and may explain why hospitalisation is a stronger predictor for the December 2020 – February 2021 wave. To conclude, each information stream (i.e., ONS estimated positivity or positive tests vs admissions) explaining different periods during March 2020-December 2021 also I) explains why they perform so well when combined in a multivariate model and II) highlights that the relationship between COVID-19 absence and surveillance data was dynamic over the study period.

The regression model explaining COVID-19 sickness absence between July 2020 and December 2021 best combines these two streams (ONS estimated positivity, new admissions) into a multivariate model and can explain most of the variability in COVID-19-related sickness absence rates between March 2020 and December 2021. Compared to the univariate ones, the strong performance of this model

suggests that each stream contains independent information about the sources of infection for NHS staff (pathways NHS staff can catch infection). These are infection pressure from the community and additional infection pressure from working in hospitals. This multivariate model combining admission and ONS estimated positivity estimates COVID-19-related sickness absence between July 2020 and early 2022 well before significantly overestimating March-May. This may suggest that other dynamics emerged that a statistical model with constant parameters could not capture. For example, the rise in milder but more transmissible subvariants of COVID-19 in late 2021 [33], and changes in policy or perception of COVID-19 risk. Furthermore, the regression coefficients hide changes in the dynamics. This is evident in the univariate model of hospitalisations explaining the trend between July 2020 and 2021 well, but not afterwards. Our multivariate model could be used to estimate the impact of future waves of COVID-19 on absence rates in staff. For example, we could use predictions of hospitalisations and incidence (with 95% confidence intervals) to create short-term projections for staff absence rates. However, given that our results suggest changes in how the predictors correlate with COVID-19 sickness absence rates, we suggest that models used for prediction are trained using only the most recent surveillance data (e.g., 5 months), and not the entire history. The most recent surveiellance data reflect the current picture regarding COVID-19 variants, vaccination, NPI's, and policy. Whereas regression models trained over longer periods, hide the recent dynamics in their coefficients.

A further limitation of our linear regression models is that they do not explain the dynamics behind COVID-19-related absences. They do not give us a way to understand the relationship between hospitalisations and the community on absences other than highlighting that these two sources play a role. We explored transforming the predictor variables (using natural log and quadratic power function - not shown), but this did not improve the performance of the models. Therefore it is unclear what the exact relationship is between COVID-19 absences and the COVID-19 activity indicators/data streams outwith it being strong and positive. Furthermore, there are variations in local epidemiology and different categories of staff, which our approach does not consider. For example, sickness absence rates are consistently higher in ambulance staff [14] and lower in non-patient-facing roles [21]. We also do not consider COVID-19 vaccination, which reduces the likelihood of NHS staff and their families falling ill with COVID-19 [10]. Our framework does not include other seasonal respiratory infections, such as influenza and rhinovirus, which will contribute to absence in the S13, 15, and 27 categories (our COVID-19-related category).

Bibliography

[1] R. P. Jones, "NHS sickness absence in England: Hidden patterns," *British Journal of Health Care Management*, vol. 26, no. 4, 2020. [Online]. Available: https://doi.org/10.12968/bjhc.2019.0026

[2] NHS Digital, "NHS Workforce Statistics - April 2022 (Including selected provisional statistics for May 2022)," NHS Digital, Tech. Rep., 2022. [Online]. Available: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-workforce-statistics/april-2022

[3] A. Wilmot and D. Leaker, "Sickness absence in the UK labour market 2021," Office for National Statistics, Tech. Rep., 2022. [Online]. Available: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/articles/sicknessabsenceinthelabourmarket/2021

[4] B. Palmer, "Chart of the week: The rise, fall and rise again of NHS staff sickness absences," Nuffield Trust, Tech. Rep., 2022. [Online]. Available: https://www.nuffieldtrust.org.uk/resource/chart-of-the-week-the-rise-fall-and-rise-again-of-nhs-staff-sickness-absences

[5] C. E. Coltart, D. Wells, E. Sutherland, and A. Fowler, "National cross-sectional survey of 1.14 million NHS staff SARS-CoV-2 serology tests: a comparison of NHS staff with regional community seroconversion rates," *BMJ Open*, vol. 11, no. 7, 2021. [Online]. Available: http://dx.doi.org/10.1136/bmjopen-2021-049703

[6] A. S. V. Shah, R. Wood, C. Gribben, D. Caldwell, J. Bishop, A. Weir, S. Kennedy, M. Reid, A. Smith-Palmer, D. Goldberg, J. Mcmenamin, C. Fischbacher, C. Robertson, S. Hutchinson, P. Mckeigue, H. Colhoun, and D. A. Mcallister, "Risk of hospital admission with

coronavirus disease 2019 in healthcare workers and their households: nationwide linkage cohort study," *BMJ*, vol. 371, 2020. [Online]. Available: https://doi.org/10.1136/bmj.m3582

[7] J. Appleby, "NHS sickness absence during the covid-19 pandemic," *BMJ*, vol. 372, 2021. [Online]. Available: http://dx.doi.org/10.1136/bmj.n471

[8] H. Mooney, "NHS faces double emergency of unprecedented demand and high COVID-19-related staff absences," Tech. Rep., 2022. [Online]. Available: https://www.nhsconfed.org/news/nhs-faces-double-emergency-unprecedented-demand-and-high-covid-related-staff-absences

[9] L. McCay, "Covid continues to disrupt: what is the plan to deal with it?" *BMJ*, vol. 378, 2022. [Online]. Available: https://www.bmj.com/content/378/bmj.o1780

[10] A. S. Shah, C. Gribben, J. Bishop, P. Hanlon, D. Caldwell, R. Wood, M. Reid, J. McMenamin, D. Goldberg, D. Stockton, S. Hutchinson, C. Robertson, P. M. McKeigue, H. M. Colhoun, and D. A. McAllister, "Effect of Vaccination on Transmission of SARS-CoV-2," *New England Journal of Medicine*, vol. 385, no. 18, pp. 1718–1720, 2021. [Online]. Available: https://www.nejm.org/doi/10.1056/NEJMc2106757

[11] Independent SAGE, "A seven point plan to suppress covid infections and reduce disruptions," *BMJ*, vol. 378, 7 2022. [Online]. Available: https://www.bmj.com/content/378/bmj.o1793

[12] R. Edge, D. A. Van Der Plaat, V. Parsons, D. Coggon, M. Van Tongeren, R. Muiry, I. Madan, and P. Cullinan, "Changing patterns of sickness absence among healthcare workers in England during the COVID-19 pandemic," *Journal of Public Health (United Kingdom)*, vol. 44, 3 2022. [Online]. Available: https://doi.org/10.1093/pubmed/fdab341

[13] D. A. Van Der Plaat, R. Edge, D. Coggon, M. van Tongeren, R. Muiry, V. Parsons, P. Cullinan, I. Madan, and S. Thomas, "Impact of COVID-19 pandemic on sickness absence for mental ill health in National Health Service staff," *BMJ Open*, vol. 11, 2021. [Online]. Available: http://dx.doi.org/10.1136/bmjopen-2021-054533

[14] Z. B. Asghar, P. Wankhade, F. Bell, K. Sanderson, K. Hird, V.-H. Phung, A. N. Siriwardena, D. Zahid, and B. Asghar, "Trends, variations and prediction of staff sickness absence rates among NHS ambulance services in England: a time series study," *BMJ Open*, vol. 11, no. 9, 2021. [Online]. Available: http://dx.doi.org/10.1136/bmjopen-2021-053885

[15] D. K. Ip, E. H. Lau, Y. H. Tam, H. C. So, B. J. Cowling, and H. K. Kwok, "Increases in absenteeism among health care workers in Hong Kong during influenza epidemics, 2004-2009," *BMC Infectious Diseases*, vol. 15, no. 1, pp. 1–9, 2015. [Online]. Available: https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-015-1316-y

[16] D. L. Schanzer, H. Zheng, and J. Gilmore, "Statistical estimates of absenteeism attributable to seasonal and pandemic influenza from the Canadian Labour Force Survey," *BMC Infectious Diseases*, vol. 11, no. 1, pp. 1–9, 2011. [Online]. Available: https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-11-90

[17] D. W. Challener, L. E. Breeher, J. Frain, M. D. Swift, P. K. Tosh, and J. O'horo, "Healthcare personnel absenteeism, presenteeism, and staffing challenges during epidemics," *Infection Control & Hospital Epidemiology*, vol. 42, pp. 388–391, 2021. [Online]. Available: https://doi.org/10.1017/ice.2020.453

[18] K. Gémes, P. Frumento, G. Almondo, M. Bottai, J. Holm, K. Alexanderson, and E. Friberg, "A prediction model for duration of sickness absence due to stress-related disorders," *Journal of Affective Disorders*, vol. 250, pp. 9–15, 5 2019. [Online]. Available: https://doi.org/10.1016/j.jad.2019.01.045

[19] M. Laaksonen, L. He, and J. Pitkäniemi, "The durations of past sickness absences predict future absence episodes," *Journal of Occupational and Environmental Medicine*, vol. 55, no. 1, pp. 87–92, 1 2013. [Online]. Available: https://journals.lww.com/joem/Fulltext/2013/01000/The_Durations_of_Past_Sickness_Absences_Predict.13.aspx

[20] J. Bowers, A. Kleczkowski, and M. Macpherson, "Modelling the effect of Covid-19 on NHS absences and possible mitigating actions to ensure staff availability," 2021, unpublished draft.

[21] D. A. van der Plaat, I. Madan, D. Coggon, M. van Tongeren, R. Edge, R. Muiry, V. Parsons, and P. Cullinan, "Risks of covid-19 by occupation in nhs workers in england," *Occupational and Environmental Medicine*, vol. 79, no. 3, pp. 176–183, 2022. [Online]. Available: https://oem.bmj.com/content/79/3/176

[22] K. Steel and E. Fordham, "Coronavirus (COVID-19) Infection Survey: methods and further information," Office for National Statistics, Tech. Rep., 2022. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveypilotmethodsandfurtherinformation#incidence

[23] UK Health Security Agency, "England Summary | Coronavirus (COVID-19) in the UK," Tech. Rep., 2022. [Online]. Available: https://coronavirus.data.gov.uk/

[24] ——, "Metrics documentation | Coronavirus in the UK," Tech. Rep., 2022. [Online]. Available: https://coronavirus.data.gov.uk/metrics/doc/newAdmissions#england

[25] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. Melbourne, Australia: OTexts, 2018. [Online]. Available: http://www.otexts.com/fpp2/

[26] R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, 2023. [Online]. Available: https://www.r-project.org/

[27] C. Zheng, N. Hafezi-Bakhtiari, V. Cooper, H. Davidson, M. Habibi, P. Riley, and A. Breathnach, "Characteristics and transmission dynamics of COVID-19 in healthcare workers at a London teaching hospital," *Journal of Hospital Infection*, vol. 106, no. 2, pp. 325–329, 2020. [Online]. Available: https://doi.org/10.1016/j.jhin.2020.07.025

[28] P. Dunn, L. Allen, G. Cameron, A. M. Malhotra, and H. Alderwick, "COVID-19 policy tracker 2020: A timeline of national policy and health system responses to COVID-19," Tech. Rep., 2023. [Online]. Available: https://www.health.org.uk/news-and-comment/charts-and-infographics/covid-19-policy-tracker

[29] E. Mathieu, H. Ritchie, L. Rodés-Guirao, C. Appel, C. Giattino, J. Hasell, B. Macdonald, S. Dattani, D. Beltekian, E. Ortiz-Ospina, and M. Roser, "Coronavirus Pandemic (COVID-19)," *Our World in Data*, 2020. [Online]. Available: https://ourworldindata.org/coronavirus

[30] T. Shi, C. Robertson, and A. Sheikh, "Effectiveness and safety of coronavirus disease 2019 vaccines," *Current Opinion in Pulmonary Medicine*, vol. 29, no. 3, 2023. [Online]. Available: https://doi.org/10.1097/mcp.0000000000000948

[31] S. Islam, T. Islam, and M. R. Islam, "New Coronavirus Variants are Creating More Challenges to Global Healthcare System: A Brief Report on the Current Knowledge," *Clinical Pathology*, vol. 15, 2022. [Online]. Available: https://doi.org/10.1177/2632010x221075584

[32] M. Mohsin and S. Mahmud, "Omicron SARS-CoV-2 variant of concern: A review on its transmissibility, immune evasion, reinfection, and severity," *Medicine (United States)*, vol. 101, no. 19, 2022. [Online]. Available: https://doi.org/10.1097/md.0000000000029165

[33] Office for National Statistics, "Coronavirus (COVID-19) latest insights," Office for National Statistics, Tech. Rep., 2022. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/infections

Chapter 5

A Mechanistic Model of COVID-19 Sickness Absence in NHS England Staff

This chapter is a manuscript produced conjointly with Itamar Megiddo (IM) and Adam Kleczkowski (AK). We plan to submit it as a paper. We developed a mechanistic model of COVID-19 sickness absence and assessed the impact and cost-effectiveness of interventions to reduce absence. Each author's contributions are outlined briefly below and shown in Table 5.1.

AK and IM played supervisory roles and provided feedback on the manuscript. The PhD author (EM) worked on; the literature review, writing and editing the manuscript, developing the model equations, developing the source code to simulate the model, and on formal model analysis and investigation (parameterisation, sensitivity analysis, and data visualisation).

Table 5.1: Description of authors' contributions to the manuscript.

Contributor Role	Role Definition	Name
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	AK
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	EM
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	EM
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	AK
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	EM
Methodology	Development or design of methodology; creation of models	AK, EM, IM
Project Administration	Management and coordination responsibility for the research activity planning and execution.	AK, IM, EM
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	AK, EM
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	AK, EM
Supervision	Oversight and leadership responsibility for the research activity planning and, execution including mentorship external to the core team.	AK, IM
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.	EM
Visualisation	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	EM
Writing- Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	EM
Writing- Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.	AK, IM, EM

5.1 Abstract

The ongoing COVID-19 pandemic has significantly impacted health-care systems worldwide, leading to a rise in staff absences due to illness. In this paper, we developed a mechanistic compartmental

model that uses publicly available COVID-19 surveillance data to estimate the trajectory of COVID-19-related sickness absence in the NHS England workforce. Our study combines an economic model with an epidemiological system, creating a framework that helps evaluate the cost-effectiveness of interventions to control such absences. We identify strategies to reduce disease-related absenteeism while considering resource constraints and epidemiological dynamics by analysing staff-related transmission dynamics and employing optimisation techniques. Our results indicate that interventions that reduce staff's contact with COVID-19-infected individuals in the community are the most cost-efficient. Additionally, our research highlights the crucial relationship between hospitalisations, the number of COVID-19 infections, and staff sickness absence, emphasising the need to prioritise measures that effectively mitigate transmission rates and minimise the burden on healthcare workers. In summary, our study contributes to evidence-based decision-making in public health and provides a framework to help manage future disease outbreaks.

5.2 Introduction

The National Health Service England bears a significant social and economic burden from sickness absence. In 2022, 27 million days were lost to sickness absence, around 74,500 full-time equivalent staff in an organisation of 1.2m [1, 2]. Sickness

absence not only hampers the provision and management of patient care, but it also negatively affects the working conditions of the remaining staff and results in additional costs through sick pay and hiring agency staff to cover shifts [3].

The arrival of the COVID-19 pandemic profoundly impacted the National Health Service (NHS) in England. The waves of COVID-19 infections created a double burden for NHS England: causing a surge in demand for treatment and an increase in absence due to illness. This combination resulted in insufficient staff to provide adequate care, placing immense pressure on the system [4]. Furthermore, COVID-19 outbreaks recurred in waves that deviated from the known pattern, intensity and duration associated with seasonal influenza [5], thus resulting in a constant strain on the NHS [6].

As a result, healthcare workers (HCWs) have experienced increased workloads, levels of stress, and risk of infection [7]. Patient-facing healthcare workers in NHS Scotland and England were three times more likely to be hospitalised [8, 9] or infected [9] with COVID-19 compared to the general population. Sickness absence rates in NHS England staff have risen dramatically. Pre-pandemic (2015-2019), the monthly average overall sickness absence rate was 4.19% for NHS England, and between 2020 and 2022, this increased from 4.66% to 5.02% to 5.57%, primarily driven by COVID-19-related and mental health sickness categories [1, 10]. During the first wave (March-May 2020), significant increases in sickness absence rates were observed for NHS England staff compared to the previous ten-year

average for each month, coinciding with the initial arrival and establishment of COVID-19 in England [11]. Furthermore, Van der Plaat et al. [12] observed a regional correlation between the percentage change in new mental health absences and absences attributed to COVID-19 in NHS England staff during the first wave, compared to pre-pandemic levels.

There is a need to use mathematical models to understand the relationship between COVID-19 transmission dynamics and the behaviour of HCWs, in particular, the impact of interventions to mitigate HCW illness and absence. Models developed to meet this goal typically operate at the hospital level and investigate the impact of these interventions on absenteeism and health-related outcomes [13–15], but neglect the cost and related economic effects of interventions. For instance, in [13, 14], compartmental SEIR models with a patient and HCW cohort demonstrate the importance of periodic testing and adequate PPE in reducing HCW absenteeism. Pham et al. [16] demonstrate similar results with a different approach. They used an agent-based model to estimate the impact of these interventions combined with contact tracing and restricting HCWs' movement between wards. Furthermore, Aguilar et al. [15] developed a pair approximation network model to explore strategies to handle HCW absences (replacement, redistribution) for a hospital during a COVID-19 outbreak.

Before the COVID-19 pandemic, similar models were developed to predict sickness absence rates for healthcare workforces while including seasonal and pan-

demic influenza effects. Le et al. [17] developed a general system dynamics model that depends on a compartmental epidemiological model, which assumes absenteeism is a function of the fraction of infected employees, caring responsibilities, risk perception, and infrastructure availability. The Centers for Disease Control (CDC) also developed a Monte Carlo simulation model to estimate days lost for a workforce during an influenza outbreak, FluWorkLoss 1.0 [18]. Their model uses probability distributions of key input variables (estimated duration of flu outbreak, attack rate, age structure of population) to produce an estimated range for the workdays lost by different reasons, age groups and health outcome categories. Wilson et al. [19] and Nap et al. [20] used this software to assess the impact of potential influenza pandemics on the healthcare workforces in New Zealand and the Netherlands, respectively. Furthermore, using multivariate regression models to estimate sickness absence attributed to influenza (seasonal and pandemic) from proxy variables for influenza activity (e.g., hospitalisations or tests) was a common approach in the literature [21–23]. However, these models do not consider the impact of interventions to reduce staff transmission pathways, for example.

As Basurto et al. [24], Duarte et al. [25], and Haw et al. [26] argued, while the well-established epidemiological models can address the health-related aspects of the COVID-19 pandemic, there is a need for rigorous approaches that consider the economic effects of interventions and containment measures. In the context of HCW's absence, we found no papers that completely met this need - costs need

to be addressed. The existing approaches that integrate epidemiological and economic factors investigate the impact of social distancing and suppression policies (e.g., lockdowns) on health-related outcomes and GDP losses, typically at the national level [24, 27–29]. For example, Basurto et al. [24] use an agent-based model that incorporates the labour market, the goods market, and household interactions to highlight the negative relationship between GDP losses and COVID-19 mortality. Similarly, Coulbourn et al. [30] used an SEIR framework inside an economic model to estimate the impact of various population-wide testing, contact tracing and isolation strategies for COVID-19 in the UK in terms of the resulting hospitalisations, deaths, costs to the UK National Health Service, reduction in GDP, and intervention costs of each strategy.

In this paper, we integrate a deterministic compartmental model with an economic framework and explore the dynamics of COVID-19 sickness absence in the NHS England workforce, using publicly available data for calibration. One key research aim is to investigate whether we can explain NHS England Sickness absence rates using this mechanistic model combined with COVID-19 surveillance data. The second aim is to employ the model to evaluate the effectiveness and cost-effectiveness of interventions aimed at reducing COVID-19-related sickness absence. Understanding the intricate relationship between COVID-19 outbreaks and sickness absence rates is crucial for effective workforce management, alleviating the strain on healthcare staff, and ensuring the provision of high-quality patient care. Additionally, it can provide insights into the impact of future waves

and the potential effects of different policy interventions [31].

This paper is structured as follows. In Section 5.3 we describe the model and methods deployed to analyse it. The corresponding results are given in Subsection 5.3.3 and discussed in Section 5.4 where we give our concluding remarks.

5.3 Model framework

We developed a compartmental system that produces a time series of COVID-19-related sickness absences in the NHS staff workforce. The model captures the unique timeline of COVID-19 disease and allows for multiple infection sources for staff. It also accounts for institutional testing and self-isolation policy. Furthermore, the model estimates the economic cost incurred by days lost due to COVID-19 absence, after accounting for interventions to control the sources of infection for NHS staff. We describe the compartmental model in Subsection 5.3.1 and the economic model in Subsection 5.3.2. All relevant parameter definitions for the model are in Table 5.2.

Table 5.2: Parameter definitions, alongside their base case values.

Parameter	Description	Base case value	Source [e]
t	Units of time (days)		
$S(t)$	Susceptible		
$P(t)$	Presymptomatic		
$A(t)$	Asymptomatic		
$I(t)$	Isolating		
$R(t)$	Recovered		
N_0	Initial staff population	1319400	[32]
Λ	Force of infection	Eq. (5.2)	
$\lambda_o(t)$	Number of people in community testing positive for COVID-19 on day t		[33]
$\lambda_h(t)$	Number of new COVID-19 hospital admissions on day t		[34]
β_s	Transmission rate from infectious staff at work (t^{-1})	0	Estimated - Section 5.3.1
β_o	Transmission rate dependent on infection pressure in community (t^{-1})	0.000000007060693	Estimated - Section 5.3.1
β_h	Transmission rate dependent on new COVID-19 hospital admissions (t^{-1})	0.000000496067891	Estimated - Section 5.3.1
τ	Asymptomatic infectious period (days, t)	10	[35]
ρ	Presymptomatic period (days, t)	2	[36, 37]
α	Fraction of asymptomatic infections	0.405	[38]
γ	Testing frequency per week	2	[39]
δ	Isolation period (days, t)	10	[40]
p_o	Community intervention effort		
p_h	Hospitalisation intervention effort		
c_o^d	Community intervention cost per staff member per day (t^{-1}£'s)		
c_h^d	Hospitalisation intervention cost per staff member per day (t^{-1}£'s)		
c_I^d	Cost per worker per day lost due to isolating with COVID-19	1	
c_o	Community intervention cost per staff member (£'s)		
c_h	Hospitalisation intervention cost per staff member (£'s)		
T	Length of the study/simulation period (days)	660	

The base case values (e) produce the dynamic model fitted to NHS England sickness absence rates between 2020-06-14 and 2022-06-12.

5.3.1 Dynamic model of absence

We now develop the deterministic compartmental model of COVID-19-related sickness absence in the NHS England staff workforce. Staff are categorised into six compartments of COVID-19 infection status: Susceptible (S), i.e., everyone who is not infected; Pre-symptomatic (P), those unknowingly infectious with the virus but not yet showing symptoms; Asymptomatic (A), undetected infectious individuals who are asymptomatic and at work; Isolating (I), infectious individuals that have been identified by a positive test or developed symptoms and stayed at home; Recovered (R), those who had COVID-19 and recovered or returned to work after a self-isolation period. The daily dynamics are governed by the following set of equations

$$\begin{aligned}
\frac{dS}{dt} &= -\Lambda S \\
\frac{dP}{dt} &= \Lambda S - \frac{P}{\rho} - \frac{\gamma P}{7} \\
\frac{dA}{dt} &= \frac{\alpha P}{\rho} - \frac{\gamma A}{7} - \frac{A}{\tau - \rho} \\
\frac{dI}{dt} &= (1-\alpha)\frac{P}{\rho} + \frac{\gamma(P+A)}{7} - \frac{I}{\delta} \\
\frac{dR}{dt} &= \frac{I}{\delta} + \frac{A}{\tau - \rho}
\end{aligned} \tag{5.1}$$

Susceptible staff become infected with a force of infection (FOI) Λ_i, and are moved to the pre-symptomatic class (P). We assume the pre-symptomatic period lasts $\rho > 0$ days on average, after which a fraction $(1-\alpha)$ of the infected staff will develop COVID-19 symptoms and immediately self-isolate at home, moving into

the Isolation class (I). The remaining staff are moved into the asymptomatic class (A). We assume all staff are regularly tested γ times per week. Therefore, presymptomatic and asymptomatic workers are identified by testing at a rate of $\frac{\gamma}{7}$ per day, after which they isolate at home. Staff isolate for $\delta > 0$ days and then return to work, moving into the recovered class (R). Furthermore, we assume workers who develop an asymptomatic infection will be infectious for $\tau > 0$ days in total, including the presymptomatic period ($\rho > 0$), during which they can test positive.

Specifically, we assume that the duration a newly infected person takes to develop symptoms (ρ) is 2 days on average [36, 37]. Furthermore, the fraction of COVID-19 infections that remain asymptomatic (α) is 0.405 [38], and these people will be infectious for $\tau = 10$ days in total [35]. Therefore, the infectious period exceeds the presymptomatic period ($\tau - \rho > 0$). NHS England policy in April 2020 was to test all symptomatic staff and staff who had contact with COVID-19-positive individuals [41]. However, all NHS England staff were tested twice weekly from November 2020 [39]. Therefore, we assume all staff are tested twice weekly ($\gamma = 2$) for our model. Furthermore, we assume staff isolate for ten days as per the UK government guidance between May 2020 and January 2022 [40].

The force of infection is partially data-driven and takes the following form,

$$\Lambda(t) = \beta_s(P + A) + \beta_o\lambda_o(t) + \beta_h\lambda_h(t). \tag{5.2}$$

Asymptomatic and pre-symptomatic staff attending work whilst unknowingly infected with COVID-19 (of which there are $P + A$) transmit the virus ([36, 37]) directly to susceptible staff at a rate of β_s. Furthermore, we assume staff face two additional sources of infection pressure: I) patients in the hospital who are COVID-19 positive and II) interactions outside of work with the general population. We express the rate at which susceptible staff become infected due to their interactions with the general population as the product between the number of people in the community newly infected with COVID-19, $\lambda_o(t)$, and a transmission rate dependent on the number of new community infections, β_o. The variable, $\lambda_o(t)$, acts as a proxy for the number of people who currently have COVID-19 in the general population. Similarly, we use the number of new COVID-19 hospital admissions ($\lambda_h(t)$) as a proxy for the number of hospital patients currently infected with COVID-19. We let the transmission rate dependent on the number of new hospitalisations be β_h. Therefore, we assume the rate at which susceptible staff become infected due to interactions with COVID-19-positive patients is $\beta_h\lambda_h(t)$. The force of infection, Eq. (5.2) assume density-dependent transmission between all three sources (staff, general population and patients) and the susceptible staff population.

In our model, the daily number of new patients admitted to hospital with COVID-19 hospital in England, $\lambda_h(t)$, acts as a proxy for the number of patients in hospitals that are currently infected with COVID-19. This variable, $\lambda_h(t)$, was taken directly from the UK Health Security Agency (UKHSA) Coronavirus Dashboard

[34], Figure 5.1. This data counts people admitted to hospitals who tested positive for COVID-19 14 days before admission or during their stay in the hospital. Inpatients diagnosed with COVID-19 after admission are reported as being admitted the day before their diagnosis [42].

The number of people in the community newly infected with COVID-19, $\lambda_o(t)$, acts as a proxy for the number of people who currently have COVID-19 in the general population in our model. We use COVID-19 incidence estimates from the Office for National Statistics Coronavirus (COVID-19) Infection Survey [33] to estimate $\lambda_o(t)$. This data reflects COVID-19 infections of people living in private households (general population households and households of NHS staff) but not patients in hospitals or care homes. Additionally, the ONS incidence rate estimates the number of new COVID-19 infections at bi-weekly intervals. Therefore we interpolated it to obtain daily estimates for the number of people in the community newly infected with COVID-19, i.e., $\lambda_o(t)$, Figure 5.1.

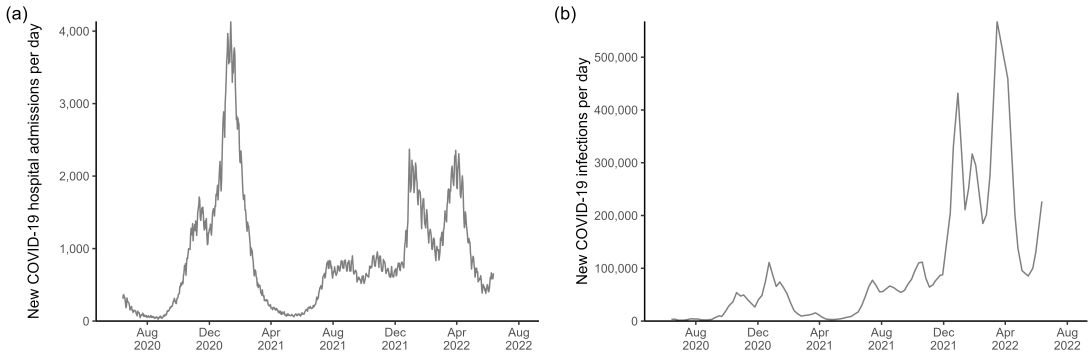

Fig. 5.1: COVID-19 surveillance data for England. Daily observations of (a) new COVID-19 hospitalisations, $\lambda_h(t)$; (b) new COVID-19 infections according to ONS, $\lambda_h(t)$.

Calibration of dynamic model to NHS England COVID-19 sickness absence data

We attempt to reproduce the dynamics of COVID-19-related sickness absence in NHS England staff between June 2020 and June 2022 by estimating the three transmission parameters for the force of infection (β_s, β_o and β_h), Eq. (5.2). As discussed in the previous section, each transmission parameter modifies the COVID-19 infection pressure acting on staff from a different source. The signal of staff-staff transmission is controlled by β_s, and the strength of the two data-driven signals from the general population and patients are modified by β_o and β_h, respectively. All other parameters were found through a literature search, as discussed in Subsection 5.3.1 and Table 5.2.

We first obtained data on absences in NHS England staff from their Sickness Absence Rates publication. This contains monthly observations of sickness absence rate broken down by reason from January 2015 until April 2022, taken directly from the electronic staff record. The sickness absence rate is defined as the ratio between the "full-time equivalent (FTE) number of days lost" and "FTE number of days available" [1], multiplied by 100 to give a percentage. NHS England started recording COVID-19-related sickness absence as a reason for absence in March 2020, and we calibrated our model to these observations.

We modelled the period between the 14th of June 2020 ($t = 0$) and the 12th of June 2022 since these were the dates for which data was available for both

new COVID-19 hospitalisations (giving $\lambda_h(t)$, a proxy for the number of COVID-19 infected patients) and ONS estimated incidence (from which we derive $\lambda_o(t)$, our proxy for the number of COVID-19 infections in the general population) Figure 5.1. On the 14th of June 2020, there was $N_0 = S(0)$ =1,319,400 staff employed by NHS England [32].

Daily estimates of the percentage of NHS staff ill and absent with COVID-19 according to our model, $100\times \frac{I(t)}{N_0}$, were converted to monthly estimates of absence rates for comparability to the observed COVID-19-related NHS sickness absence rates over the study period by taking a mean for each month. We then estimated our model's three transmission rate parameters (β_s, β_o and β_h) by minimising the sum of squared residuals between the observed and modelled absence between July 2020 and May 2022. We used the "FME" package in R [43] for this task, an inverse modelling package targeted towards ODE systems. The specific optimisation algorithm used was the Levenberg-Marquardt algorithm. This and all further analyses were performed in Rstudio version 4.2.1 [44].

5.3.2 Economic model: net cost of COVID-19 sickness absence to NHS England

We now develop the model estimating the total cost of COVID-19 sickness absence to NHS England. This includes the cost of the working days lost due to illness, which incurs the financial burden of sick pay and covering shifts, plus the

additional cost of any interventions implemented to reduce the likelihood of staff falling ill with COVID-19.

We assume there are two interventions which will independently reduce the risk of COVID-19 infection for each staff member over the entire study period (14 June 2020 - 12 June 2022). The impact of the first is to reduce the force of infection acting on staff that depends on the number of new COVID-19 hospital admissions. This represents improvements in PPE and campaigns to encourage good hygiene practices such as hand washing and social distancing in the workplace. Similarly, the second reduces the force of infection acting on staff that depends on the number of people testing positive for COVID-19 in the community. This represents social distancing, hand-washing, and encouraging staff to mix in small groups outside work and not mix with people who may be infected with COVID-19. The scale and efficiency of each intervention are captured by the effort parameters $p_h \in [0,1]$ and $p_o \in [0,1]$. Each intervention is modelled by replacing the β_o and β_h in Eq. (5.2) with the reduced transmission rates,

$$\begin{aligned} \hat{\beta}_o &= (1-p_o)\beta_o \\ \hat{\beta}_h &= (1-p_h)\beta_h \end{aligned} \tag{5.3}$$

We write the total cost of absences and interventions to NHS England over the study period as

$$\hat{J} = \int_0^T p_o c_o^d N_0 \, dt + \int_0^T p_h c_h^d N_0 \, dt + C_{\text{days lost}}(p_o, p_h). \tag{5.4}$$

We assume each intervention has a constant cost per staff member per day which depends linearly on the effort (scale and efficiency of the intervention). We call these the daily community cost, c_o^d, and the daily hospital cost, c_h^d. Furthermore, the cost per worker per day lost due to isolating with COVID-19 is given by c_I^d. Therefore, the total cost of days lost over the study period (length T days), $C_{\text{days lost}}$, is given by

$$C_{\text{days lost}}(p_o, p_h) = c_I^d \int_0^T I(t)\, dt. \tag{5.5}$$

A key priority of the health service is to determine the intervention strategy that will minimise the total cost of absences. There are two control variables; the community effort time (p_o) and rotation length (p_o). Their optimal values that minimise the objective function (5.4) are given by

$$\begin{aligned} \mathbf{s} = \ \arg\min_{\mathbf{x}} \quad & \hat{J}(\mathbf{x}), \quad \text{where} \quad \mathbf{x} = (p_o, p_h) \\ \text{subject to} \quad & 0 \leq p_o, p_h \leq 1 \end{aligned} \tag{5.6}$$

We do not estimate the intervention efforts (p_o, p_h) or their costs (c_o, c_h) using literature or a formal parameterisation process. Instead, we investigate their impact on the model outcomes through a sensitivity analysis with analytic and numerical results. We first proceed by factoring c_I^d from the objective function ($\hat{J}$) in Eq. (5.4), and setting $c_I^d = 1$, since the optimum of the objective function (intervention efforts; p_o, p_h) is unchanged by scaling the objective. This allows us to investigate the impact of the two new variables on the optimal strategy:

the costs of each intervention (per staff member over the study period) relative to the cost of a day lost (per staff member over the study period). We call these the hospital intervention cost relative to absence, $c_h = Tc_h^d$, and the community intervention cost relative to absence, $c_o = Tc_o^d$.

5.3.3 Results

In Section 5.3.3, we first show how the model captures the NHS England absence data between June 2020 to June 2022 and subsequently show how sensitive the results are to changes in key parameters. In Section 5.3.3, we investigate optimal strategies to minimise the economic cost of sickness absence to NHS England.

Compartmental model dynamics

Observed COVID-19 sickness absence was low in July 2020 (0.2%) and then rose steadily towards a peak of 1.6% in January 2021, before dropping back to 0.2% by May 2021, Figure 5.2(b). Sickness absence started to rise slowly between July and October 2021, reaching 0.5% by the end of October. They increased massively over the winter and peaked at just over 2% in January 2022. They dropped to 1.2% in February and increased slightly in March 2022 (1.6%) before falling further to 0.5% by May 2022.

We fitted the dynamic model to the NHS sickness absence data between June 2020 and June 2022 using the process discussed in Section 5.3.1, Figure 5.2. The

model captures the overall trend reasonably well with a sum of squared residuals (SRR) of 0.484. The model slightly underestimates observed absences in the June-September trough, then matches the rise into January 2021 peak (1.6%) in absences over the winter well. The model estimates the fall into April 2021 (0.4%) well but underestimates the second trough in the data around June 2021 (0.04% vs 0.22%). The model estimates a slightly steeper increase in absence rates between July and October 2021 (delta wave) than observed, and as a result, slightly overestimates October 2021 (0.62% vs 0.5%). The model matches the sharp increase from November 2021 - January 2022 during the omicron period. However, it underestimates the peak in January 2022 (1.8% vs 2.2%) and overestimates the drop in February 2022 (1.4% vs 1.2%). Furthermore, the model estimates an increase March-May 2022 then a fall to 0.6% by June, when instead, there was an increase March-April before a fall to 0.6% by June.

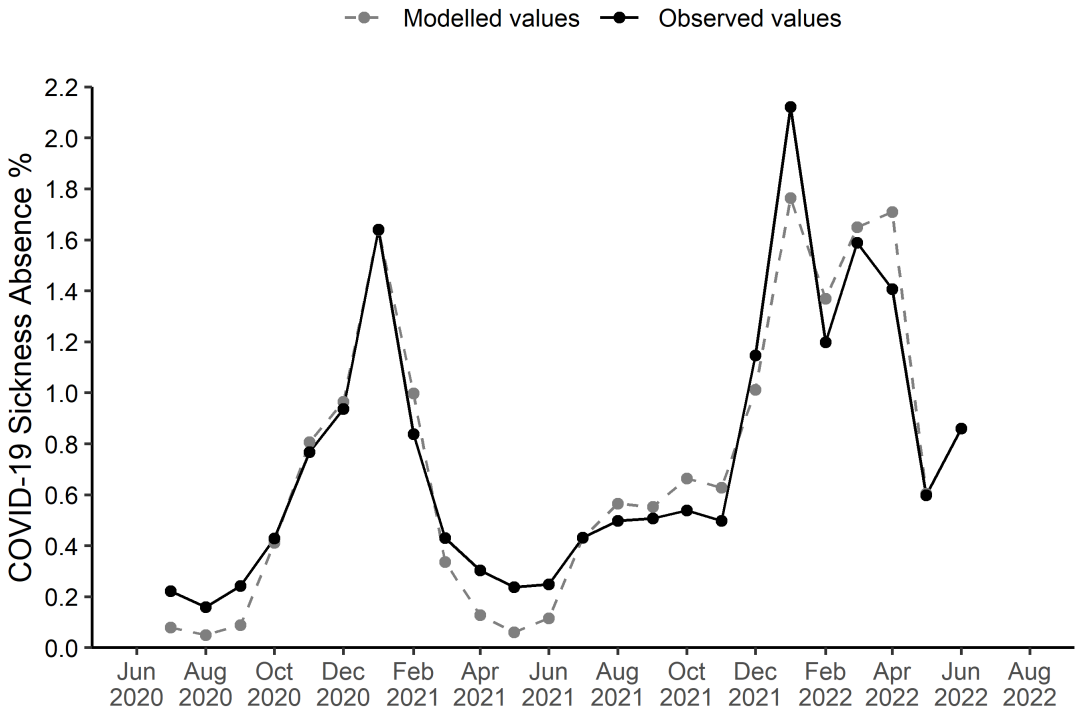

Fig. 5.2: Dynamic model of COVID-19 sickness absence fitted to NHS England. The solid black line is the monthly observed COVID-related absence rate in NHS staff, and the dashed line is the corresponding estimated rate according to our model. This modelled rate is given by converting the daily percentage of NHS staff ill and absent with COVID-19 in the model, i.e., $100\times \frac{I(t)}{N_0}$ after solving Eq. (5.1) with the parameter values shown in Table 5.2, to monthly estimates by taking a mean for each month. Further details on parameter estimation are given in Section 5.3.1.

The fitted transmission parameters (β_s, β_o, β_h) are given in Table 5.2. In particular, note that the staff-staff transmission parameter was not required to fit the data ($\beta_s = 0$). This means that the transmission dynamics in the fitted model (the FOI, Eq. (5.2)) of COVID-19 absence are governed completely by the two data-driven infection rates. One signal dependent on community COVID-19

infections (controlled by $\beta_o \lambda_o(t)$) and another dependent on new COVID-19 hospitalisations (controlled by $\beta_h \lambda_h(t)$).

We performed local sensitivity analysis w.r.t. the three transmission parameters (β_s, β_o, β_h) by calculating sensitivity functions using the FME package in R [43]. For each time point, this estimates the derivative of the modelled value for each of the selected parameters. We found strong collinearity and positive correlation (r=0.94) between the corresponding sensitivity functions for β_s and β_h. This suggests the hospitalisation stream (implicitly) captures all information about infections contained within the staff-staff transmission stream of the model. There were weaker correlations between all other parameters ($r = -0.12, 0.21$). In particular, the community and hospitalisation transmission sources were not strongly correlated. Given the linear form of the force of infection, Eq. (5.2), we could estimate the fraction of infections each source was responsible for over the study period. The ONS/community source was responsible for 57% of infections, and hospitalisations were responsible for 43%. The ONS source slightly dominated the dynamics over the study period.

Individually modifying the transmission rates from each source of infection, Eq. (5.2), impacts the magnitude of sickness absence at different periods in our simulated absence timeseries, Figure 5.3. We varied the community (β_o) and hospitalisation (β_h) transmission parameters from their base case (fitted values, Table 5.2) by eleven different percentage changes. We did the same for eleven staff-staff

transmission parameters (β_s) equally spaced between 0 and 0.0000004. The community (β_o) transmission parameter has a huge impact on the sickness absence rate around February-April 2022, Figure 5.3 (a). Its impact decreases for the earlier months between September 2021-January 2022 and is even smaller than those in January 2021. This is due to the shape of the community indigence data, Figure 5.1. Furthermore, a reduction of 50% in the community (β_o) transmission parameter causes a decrease larger in magnitude than the increase in sickness absence caused by increasing the parameter by 50%. Changes in the community (β_o) transmission parameter correlate positively with the sickness absence rate over the study period.

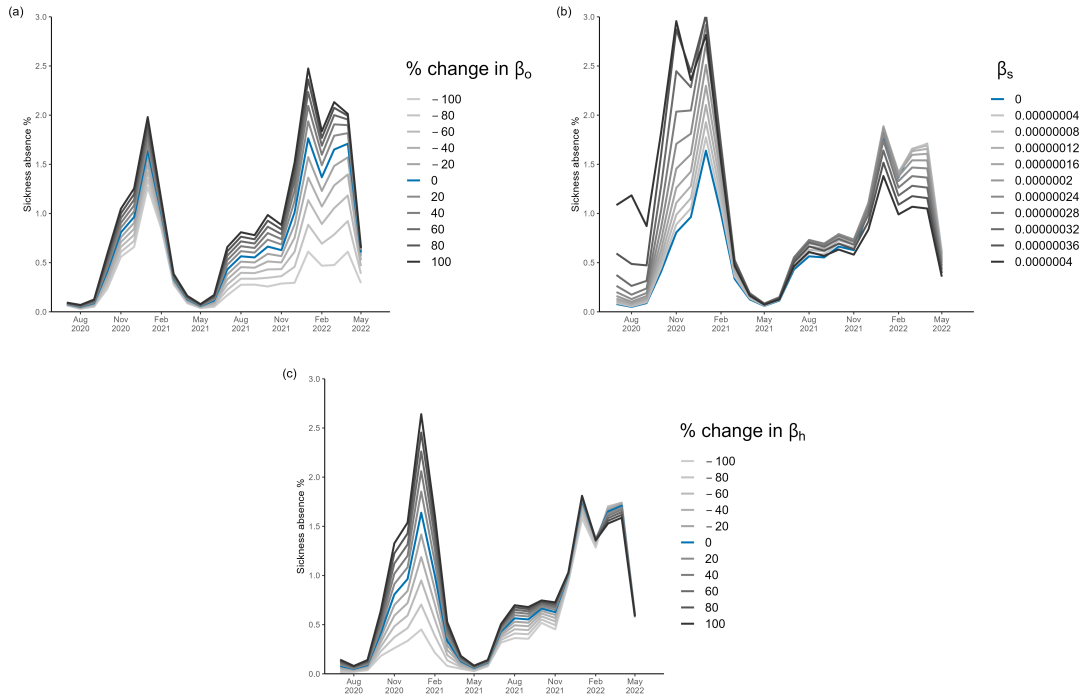

Fig. 5.3: Sensitivity analysis of the modelled COVID-19-related absence rate to the transmission parameters. Each panel shows the modelled absence rate timeseries for eleven different values of a transmission parameter (β_o, β_s, or β_h), keeping all other parameter values at the base case values in Table 5.2 (i.e., those fit to NHS England). The following parameters are varied in each panel: (a) the community transmission parameter (β_o); (b) the staff-staff transmission parameter (β_s); the hospitalisation transmission parameter (β_h). In panels (a) and (c), the modelled rate was plotted for transmission rates after a percentage change from the base case value (Table 5.2), whereas in (b), the modelled rate was plotted for the different values shown. Darker lines within panels indicate higher values of each transmission parameter. The blue line highlights the modelled absence rate fitted to NHS England. Note that the modelled rate is given by converting $\frac{A(t)}{N_0}$ to monthly observations after solving Eq. (5.1).

On the other hand, the hospitalisation (β_h) transmission parameter significantly impacts the sickness absence rate during the September 2020 - February 2021 wave, Figure 5.3. At this peak, a reduction of 50% in the hospitalisation (β_h)

transmission parameter causes a decrease in sickness absence larger in magnitude than the increase caused by increasing the parameter by 50%. The hospitalisation (β_h) transmission parameter has a more minor impact on sickness absence during the resurgence in absence from summer 2021 into 2022, Figure 5.3. Again, this is due to the shape of the hospitalisation incidence data, Figure 5.1. The hospitalisation (β_h) transmission parameter positively correlates with sickness absence for most of the study period. However, in the last few months, this relationship changes, and there is a negative relationship. Increasing the hospitalisation (β_h) transmission parameter slightly decreases the sickness absence rate. This is because the susceptible staff population is massively depleted by then.

The staff-staff transmission parameter (β_h) mainly impacts the dynamics of the start of the study period (September 2020-January 2021), Figure 5.3. Increasing the staff-staff transmission parameter, β_s, primarily increases the level of sickness absence between July 2020 - January 2021. Furthermore, when increased significantly ($\beta_s > 0.00000035$) it accelerates the initial outbreak in the model and causes a peak in August 2020. However, when the staff-staff transmission parameter is increased above a certain threshold, it causes fewer absences between January and June 2022. In the model, there susceptible staff pool has depleted and most have recovered from the virus, causing a drop in the force of infection.

When the staff-staff transmission parameter (β_s) equals zero, then increasing the PCR testing frequency (γ)) increases the absence rate. This is because

testing does not affect the internal staff-staff infection dynamics in this model configuration. Staff still move to infected classes at a rate determined by the exogenous variables dependent on the COVID-19 data (hospitalisations, community incidence). Therefore, testing identifies more of the infected staff. Whereas if the staff-staff transmission parameter (β_s) is increased from nonzero (results not shown), we see a switch to a negative relationship, where increasing testing decreases absences since there's a benefit to removing staff from work. Then there is a switch back to a positive relationship once the staff-staff transmission parameter (β_s) is increased high enough that all staff are infected very quickly.

Sensitivity of the optimal intervention strategy to costs

To guide our intuition, we first find the first-order condition for the optimal hospitalisation intervention effort (p_h) that minimises the total absences cost to NHS England, Eq. (5.4). We assume the community intervention effort, p_o, is fixed. We derive the condition by differentiating Eq. (5.4) with respect to p_h,

$$\frac{\partial \hat{J}}{\partial p_h} = c_h N_0 + \frac{\partial}{\partial p_h} C_{\text{days lost}}(p_o, p_h). \tag{5.7}$$

Setting Eq. (5.7) equal to zero and rearranging gives the optimal hospital intervention effort condition

$$c_h N_0 = -\frac{\partial}{\partial p_h} C_{\text{days lost}}(p_o, p_h) \tag{5.8}$$

We conclude that the optimal hospital intervention effort, p_h, is the effort where the increase in hospital intervention cost ($c_h N_0$) for a unit of effort equals the decrease in days lost cost caused by a unit increase of effort in the hospital intervention. It will be optimal to invest a unit of effort in the hospital intervention when it causes a larger decrease in the costs of days lost due to absence than it costs to implement.

Furthermore, we find that the absence objective function ($\hat{J}$) is essentially linear (shown later) and therefore $\frac{\partial \hat{J}}{\partial p_h} > 0$ means the optimal $p_h = 0$, and $\frac{\partial \hat{J}}{\partial p_h} < 0$ means optimal $p_h = 1$. The point of indifference between a strategy of full intervention compared to no intervention is given by Eq. (5.8).

We now find the first-order condition for optimal community intervention effort (p_o) that minimises the total cost, Eq. (5.4). We assume the hospital intervention effort, p_h, is fixed and derive the condition by differentiating Eq. (5.4) with respect to p_o,

$$\frac{\partial \hat{J}}{\partial p_o} = c_o N_0 + \frac{\partial}{\partial p_o} C_{\text{days lost}}(p_o, p_h). \tag{5.9}$$

Setting Eq. (5.9) equal to zero and rearranging gives the optimal community intervention effort condition

$$c_o N_0 = -\frac{\partial}{\partial p_o} C_{\text{days lost}}(p_o, p_h) \tag{5.10}$$

This means the optimal community intervention effort, p_o, is the effort where the increase in community intervention cost for an additional unit of effort equals the decrease in days lost cost caused by that unit increase of effort in the community intervention. It will be optimal to invest a unit of effort in the community intervention when it causes a larger decrease in the costs of days lost due to absence than it costs to implement.

We investigated the sensitivity of the optimal combined intervention strategy (minimising the total cost of absences, Eq. (5.6)) to the costs, Figure 5.4. The optimal strategy, in terms of the level of investment in either intervention, consisted of boundary solutions (either full or no intervention in each), Figure 5.4 (a). This implies that the cost of days lost due to absence ($C_{\text{days lost}}(p_o, p_h)$, Eq. (5.5)), and therefore the total cost of absence ($\hat{J}$, Eq. (5.4)), changes linearly with respect to changes in the intervention efforts (p_o, p_h). There are four possible optimal strategies; fully investing in both the hospital and community intervention ($p_o = p_h = 1$), investing fully in the hospital intervention but not at all in the community intervention ($p_o = 0$, $p_h = 1$), investing fully in the community intervention but not at all in the hospital intervention ($p_o = 1$, $p_h = 0$), and not investing any effort into either intervention ($p_o = p_h = 0$). There are four corresponding expressions for the total cost of absences ($\hat{J}$, Eq. (5.4)). We can use these to determine the exact restrictions on the intervention costs (c_o, c_h)

where each strategy in Figure 5.4 (a) is optimal. Furthermore, note that since the strategy determines the effective transmission rates, the absence time-series under each strategy can be deduced from Figure 5.3 (a) or (c).

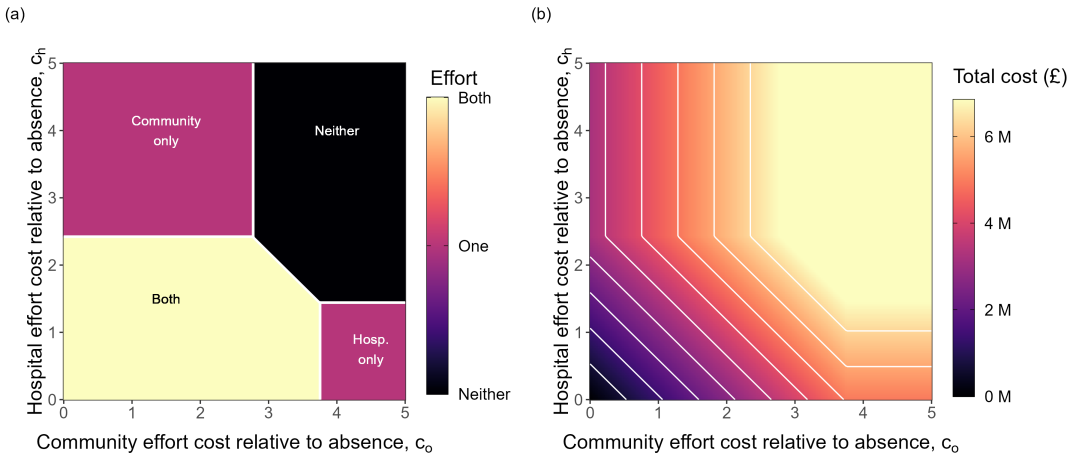

Fig. 5.4: Sensitivity of the optimal intervention strategy to intervention costs relative to absence cost. We calculated the optimal intervention strategy in an intervention cost ($c_o - c_h$) parameter space by solving Eq. (5.6) for different combinations of c_o and c_h. All other parameter values are given in Table 5.2. Panel (a) shows the optimal strategy in a $c_o - c_h$ parameter space; panel (b) the total cost under the optimised strategy, $\hat{J}$, given by Eq. (5.4). The optimal strategy consisted of boundary solutions (either full or no investment of effort in each intervention). Therefore, there are four possible optimal strategies; fully investing in both the hospital and community intervention ($p_o = p_h = 1$), investing fully in the hospital intervention but not at all in the community intervention ($p_o = 0$, $p_h = 1$), investing fully in the community intervention but not at all in the hospital intervention ($p_o = 1$, $p_h = 0$), and not investing any effort into either intervention ($p_o = p_h = 0$).

For example, if it is optimal to invest in both the hospital and community intervention fully, then the total cost of absences ($\hat{J}$, Eq. (5.4)) is given by

$$\hat{J}(p_o = 1, p_h = 1) = c_o N_0 + c_h N_0 + C_{\text{days lost}}(p_o = 1, p_h = 1) = c_o N_0 + c_h N_0. \quad (5.11)$$

If it is optimal to invest fully in the hospital intervention but not in the community intervention, then the total cost becomes

$$\hat{J}(p_o = 0, p_h = 1) = c_h N_0 + C_{\text{days lost}}(p_o = 0, p_h = 1). \tag{5.12}$$

If it is optimal to invest fully in the community intervention but not in the hospital intervention, then the total cost becomes

$$\hat{J}(p_o = 1, p_h = 0) = c_o N_0 + C_{\text{days lost}}(p_o = 1, p_h = 0). \tag{5.13}$$

If it is optimal to invest in neither intervention, then the total cost of absences ($\hat{J}$, Eq. (5.4)) is given by

$$\hat{J}(p_o = 0, p_h = 0) = C_{\text{days lost}}(p_o = 0, p_h = 0) \tag{5.14}$$

Using Eq. (5.11) - Eq. (5.13) we can find the restrictions on the intervention costs (c_o, c_h) where each strategy in Figure 5.4 (a) is optimal. For example, it is optimal to invest in both interventions ($p_o = 1, p_h = 1$) when Eq. (5.11) $\leq$ Eq. (5.12), Eq. (5.11) $\leq$ Eq. (5.14), and Eq. (5.11) $\leq$ Eq. (5.13). These restrictions define the lower left region of Figure 5.4 (a) and simplify to

$$\begin{aligned} c_o + c_h &\leq \frac{C_{\text{days lost}}(p_o = 0, p_h = 0)}{N_0} \approx 5.3 \\ c_h &\leq \frac{C_{\text{days lost}}(p_o = 1, p_h = 0)}{N_0} \approx 2.4 \\ c_o &\leq \frac{C_{\text{days lost}}(p_o = 0, p_h = 1)}{N_0} \approx 3.8 \end{aligned} \tag{5.15}$$

This means when the community intervention cost relative to the cost of a day lost due to absence (c_o) is below 3.8, the hospital intervention cost relative to the cost of a day lost due to absence (c_h) is less than 1.3, and the relative combined cost of both interventions is less than 5.3 times the community cost, full investment in both the hospital and community intervention ($p_o = 1, p_h = 1$) is optimal.

It is optimal to invest fully in the hospital intervention but not in the community intervention ($p_o = 0, p_h = 1$) when Eq. (5.12) $\leq$ Eq. (5.11), Eq. (5.12) $\leq$ Eq. (5.14), and Eq. (5.12) $\leq$ Eq. (5.13). These restrictions define the lower right region of Figure 5.4 (a) and simplify to

$$\begin{aligned} c_o N_0 &\geq C_{\text{days lost}}(p_o = 0, p_h = 1) \\ c_h N_0 &\leq C_{\text{days lost}}(p_o = 0, p_h = 0) - C_{\text{days lost}}(p_o = 0, p_h = 1) \\ c_h N_0 &\leq c_o N_0 - C_{\text{days lost}}(p_o = 0, p_h = 1) + C_{\text{days lost}}(p_o = 1, p_h = 0) \end{aligned} \tag{5.16}$$

We have that $C_{\text{days lost}}(p_o = 0, p_h = 0) - C_{\text{days lost}}(p_o = 0, p_h = 1) \approx £7m - £5m = £2m \leq C_{\text{days lost}}(p_o = 1, p_h = 0) \approx £3m$. Therefore, Eq. (5.16) simplifies and there are two restrictions where full investment in the hospital intervention but none in the community ($p_o = 0, p_h = 1$) is optimal,

$$\begin{aligned} c_o &\geq \frac{C_{\text{days lost}}(p_o = 0, p_h = 1)}{N_0} \approx 3.8 \\ c_h &\leq \frac{C_{\text{days lost}}(p_o = 0, p_h = 0) - C_{\text{days lost}}(p_o = 0, p_h = 1)}{N_0} \approx 1.3. \end{aligned} \tag{5.17}$$

In other words, when the community cost relative to the cost a day lost due to absence c_o is greater than 3.8, and the hospital cost relative to the cost a day lost due to absence c_h is less than 1.3, full investment in the hospital intervention but none in the community ($p_o = 0, p_h = 1$) is optimal.

Similarly, it is optimal to invest fully in the community intervention but not in the hospital intervention ($p_o = 1, p_h = 0$) when Eq. (5.13) $\leq$ Eq. (5.11), Eq. (5.13) $\leq$ Eq. (5.12), and Eq. (5.13) $\leq$ Eq. (5.14). These restrictions define the upper left region of Figure 5.4 (a) and simplify to

$$\begin{aligned} c_h N_0 &\geq C_{\text{days lost}}(p_o = 1, p_h = 0) \\ c_h N_0 &\leq C_{\text{days lost}}(p_o = 0, p_h = 0) - C_{\text{days lost}}(p_o = 1, p_h = 0) \\ c_o N_0 &\leq c_h N_0 - C_{\text{days lost}}(p_o = 1, p_h = 0) + C_{\text{days lost}}(p_o = 0, p_h = 1) \end{aligned} \tag{5.18}$$

Since $C_{\text{days lost}}(p_o = 0, p_h = 0) - C_{\text{days lost}}(p_o = 1, p_h = 0) \approx £7m - £3m = £4m \leq C_{\text{days lost}}(p_o = 0, p_h = 1) \approx £5m$, Eq. (5.18) simplifies, leaving two restrictions where full investment in the community intervention but none in the hospital intervention ($p_o = 1, p_h = 0$) is optimal,

$$\begin{aligned} c_h &\geq \frac{C_{\text{days lost}}(p_o = 1, p_h = 0)}{N_0} \approx 2.4 \\ c_o &\leq \frac{C_{\text{days lost}}(p_o = 0, p_h = 0) - C_{\text{days lost}}(p_o = 1, p_h = 0)}{N_0} \approx 2.8. \end{aligned} \tag{5.19}$$

When the community cost relative to the cost a day lost due to absence c_o is less than 2.8, and the hospital cost relative to the cost a day lost due to absence c_h is greater than 2.4, no investment in the hospital intervention but full in the community ($p_o = 1, p_h = 0$) is optimal.

The restrictions on the costs where it is not optimal to invest at all in either intervention ($p_o = 0, p_h = 0$) are given by Eq. (5.14) $\leq$ Eq. (5.11), Eq. (5.14) $\leq$ Eq. (5.12), and Eq. (5.14) $\leq$ Eq. (5.13). IN the following parameter space, the cost of the interventions outweigh the benefit in terms of reduction in absences,

$$\begin{aligned} c_o + c_h &\geq \frac{C_{\text{days lost}}(p_o = 0, p_h = 0)}{N_0} \approx 5.3 \\ c_h &\geq \frac{C_{\text{days lost}}(p_o = 0, p_h = 0) - C_{\text{days lost}}(p_o = 0, p_h = 1)}{N_0} \approx 1.3 \\ c_o &\geq \frac{C_{\text{days lost}}(p_o = 0, p_h = 0) - C_{\text{days lost}}(p_o = 1, p_h = 0)}{N_0} \approx 2.8. \end{aligned} \tag{5.20}$$

Figure 5.4 (b) shows how the total cost of absences changes under the optimal combination strategy in the cost ($c_o - c_h$) parameter space. The optimal strategy is to invest in both interventions when the costs are below the previously discussed thresholds (Eq. (5.15)) This region is the lower left corner of Figure 5.4 (b), and the cost is given by Eq. (5.11). It is clear in that region that the total cost increases with each individual intervention cost. When the optimal strategy is to invest in the community intervention fully but not the hospital intervention, then the total cost is only affected by the community intervention cost, Eq. (5.13), as evident in the top left corner of Figure 5.4 (b). Similarly, in the lower right corner of Figure 5.4 (b) where the optimal strategy is to invest in the

hospital intervention fully but not the community intervention, the total cost is increased only by the community intervention cost, Eq. (5.12). Whereas when it is optimal to invest in neither intervention (top right of Figure 5.4 (b)), the total cost does not change with increases in the intervention costs. The total cost equals the cost of days lost with no intervention, Eq. (5.14).

When the hospital cost relative to an absence (c_h) is greater than ≈ 2.4, it is never optimal to invest at all in the hospital intervention ($p_h = 0$), independent of the community cost and effort, Figure 5.4 (a). A unit of hospital intervention costs more than the decrease the intervention causes in the cost of days lost, and therefore $\frac{\partial \hat{J}}{\partial p_h} > 0$ in Eq. (5.7). Comparing Figure 5.5 (a) to (b) (or Figure 5.6 (a) to (b)) in the $c_h > 2.4$ region shows an example of how the increase in the total intervention cost ($p_o c_o N_0 + p_h c_h N_0$) exceeds the decrease in the cost of days lost due to absence ($C_{\text{days lost}}$) for an increase in the hospital intervention effort. Similarly, when the community cost relative to absence is greater than ≈ 3.8, it is never optimal to invest in the community intervention ($p_o = 0$). A unit of community intervention costs more than the decrease the intervention causes in the cost of days lost, and therefore $\frac{\partial \hat{J}}{\partial p_o} > 0$ in Eq. (5.9). This threshold cost, where it becomes no longer optimal to invest at all in the community intervention ($c_o \approx 3.8$), is larger than the threshold cost for the hospital intervention ($c_h \approx 2.4$), Figure 5.4 (a). This asymmetry is because a unit of change in the community intervention (p_o) decreases the cost of days lost due to COVID-19 illness ($C_{\text{days lost}}$) more than a unit change in the hospital intervention. This is

because the community force of infection dominates the dynamics (accounts for more infections) after fitting the model to the NHS absence data (discussed in Section 5.3.3).

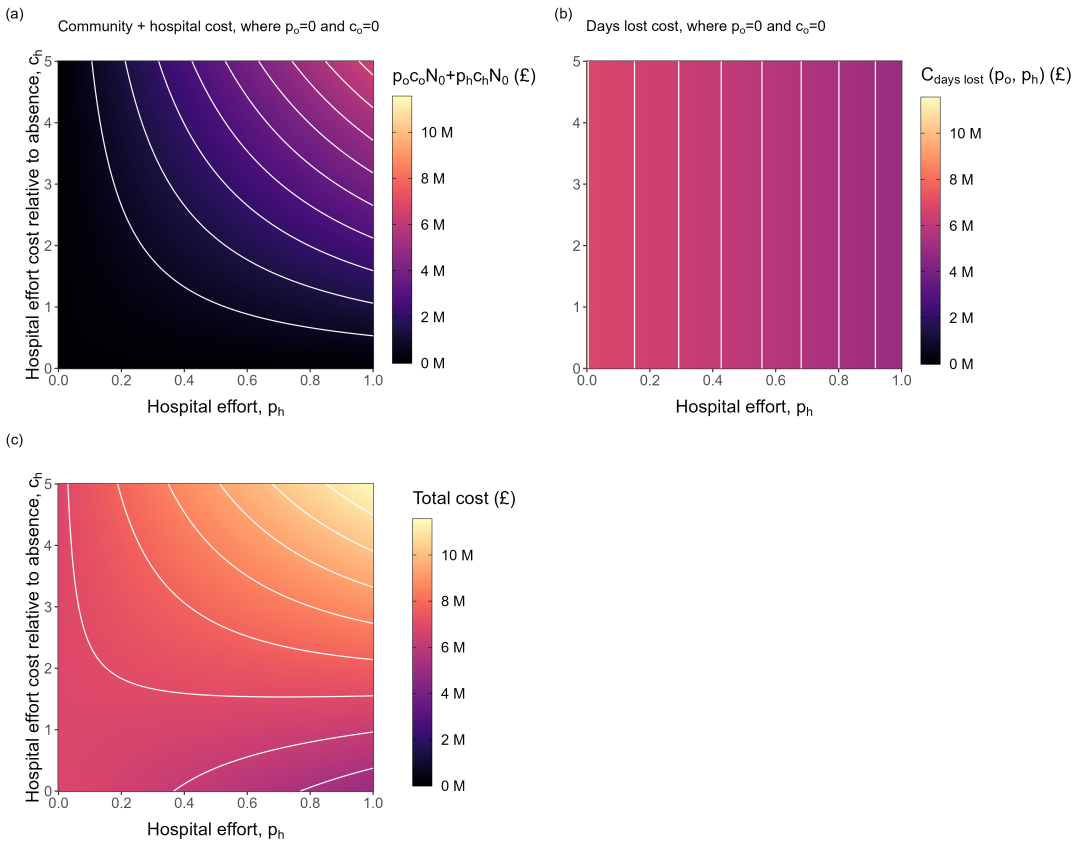

Fig. 5.5: Sensitivity of the total cost of COVID-19 absence to changes in hospital intervention effort and cost. We calculated the total cost in a hospital effort - hospital intervention cost ($p_h - c_h$) parameter space using Eq. (5.4) for the different combinations. The community effort and cost were fixed at $p_o = 0$ and $c_o = 0$. All other parameter values are given in Table 5.2. Panels (a)-(c) give a breakdown of the total cost (Eq. (5.4)): (a) shows the combined cost of the interventions ($p_o c_o N_0 + p_h c_h N_0$); (b) the cost of the days lost to COVID-19 absence ($C_{\text{days lost}}(p_o = 0, p_h)$; (c) the total cost of interventions and absences to the NHS, $\hat{J}$, given by Eq. (5.4).

As discussed, community intervention is optimal ($p_o = 1$) only when the community cost is sufficiently low. Furthermore, when the community cost is high ($c_o > 3,8$)), the threshold hospital cost that causes a switch from full hospital intervention to none is lower compared to when the community cost is low ($c_o < 2.8$)). This switch occurs at $c_h \approx 1.3$ instead of $c_h \approx 2.4$, Figure 5.4 (a). For $\frac{\partial \hat{J}}{\partial p_h} < 0$ in Eq. (5.7), this means the decrease in the cost of days lost due to absence that comes with an increase in hospital intervention effort must be larger when the community intervention effort is higher. In other words, $\frac{\partial}{\partial p_h} C_{\text{days lost}}(p_o = 1, p_h) < \frac{\partial}{\partial p_h} C_{\text{days lost}}(p_o = 0, p_h)$. When the community intervention effort is larger, the FOI from the community in Eq. (5.3) and Eq. (5.2) is smaller. Therefore the hospital FOI acts on a larger pool of susceptibles, so the same percentage increase in hospitalisation effort (p_h), has more effect in reducing the overall force of infection. Since the benefit/impact of a unit of intervention (decrease in days lost cost) is higher, the cost of a unit of intervention must be higher to incentivise not implementing it.

This result introduced in the last paragraph is also evident by comparing Figure 5.5 (c) and Figure 5.6 (c). By looking at a transect from a value of the y-axis in each of these panels (i.e., a value of the hospital intervention cost relative to absence, c_h, we can identify the optimal hospital effort (p_h) by the value with the lowest total cost. The community intervention (p_o) is zero in Figure 5.5 (c), representing a scenario with high community intervention cost. On the other hand, it is low in Figure 5.6 (c) ($c_o = 0.5$) - low enough for community intervention to

be optimal.

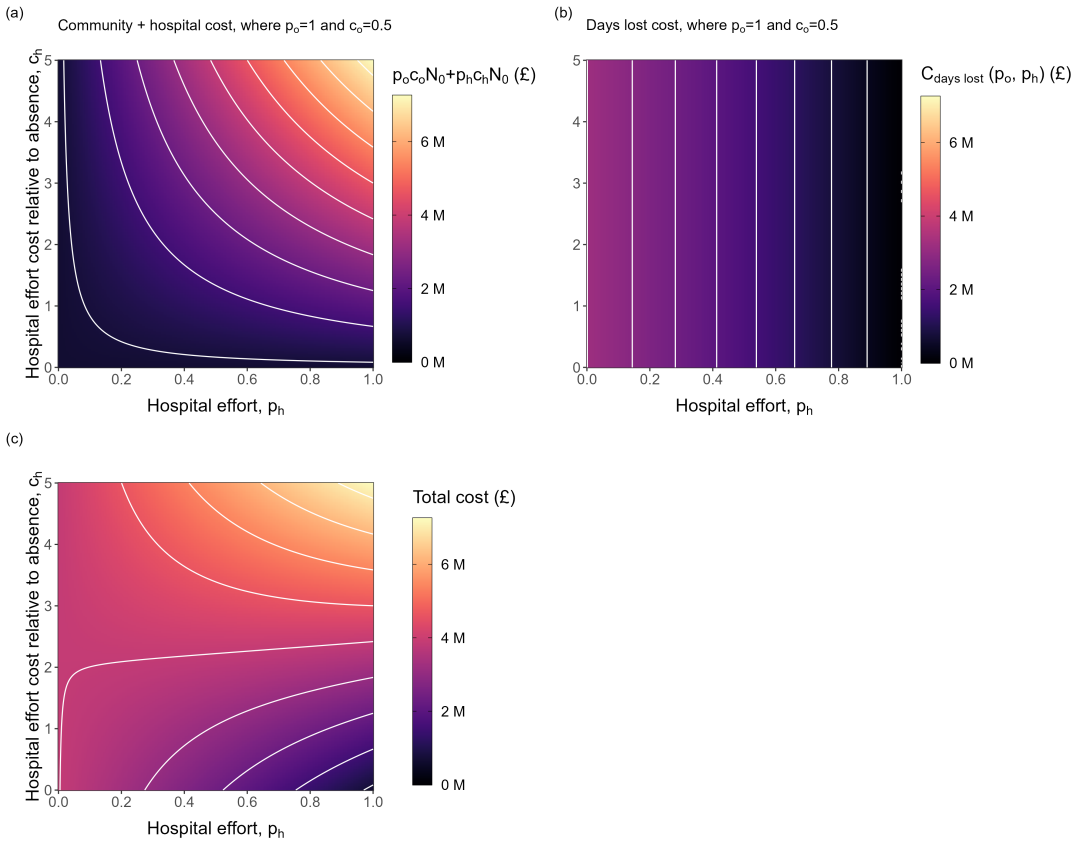

Fig. 5.6: Sensitivity of the total cost of COVID-19 absence to changes in hospital intervention effort and cost. We calculated the total cost in a hospital effort - hospital intervention cost ($p_h - c_h$) parameter space using Eq. (5.4) for the different combinations. The community effort and cost were fixed at $p_o = 1$ and $c_o = 0.5$. All other parameter values are given in Table 5.2. Panels (a)-(c) give a breakdown of the total cost (Eq. (5.4)): (a) shows the combined cost of the interventions ($p_o c_o N_0 + p_h c_h N_0$); (b) the cost of the days lost to COVID-19 absence ($C_{\text{days lost}}(p_o = 0, p_h)$); (c) the total cost of interventions and absences to the NHS, $\hat{J}$, given by Eq. (5.4).

There is a combination of costs that produces indifference between investing in either i) both interventions or ii) neither intervention, and investing in only one intervention is suboptimal, Figure 5.4 (a) and Figure 5.7 (c). Using Eq. (5.11) and Eq. (5.14), the combinations of intervention costs that result in this phenomenon satisfy $c_o + c_h = \frac{C_{\text{days lost}}(p_o=0,p_h=0)}{N_0} \approx 5.3$ for $c_o < 3.8$ and $c_h < 2.4$. Comparing Figure 5.7 (a) and (b), we see a symmetry in the impact of a unit of each effort on the total intervention cost and cost of days lost due to absence. The increase in intervention cost ($p_o c_o N_0 + p_h c_h N_0$) equals the decrease in the cost of days lost due to absence ($C_{\text{days lost}}(p_o = 0, p_h = 0)$).

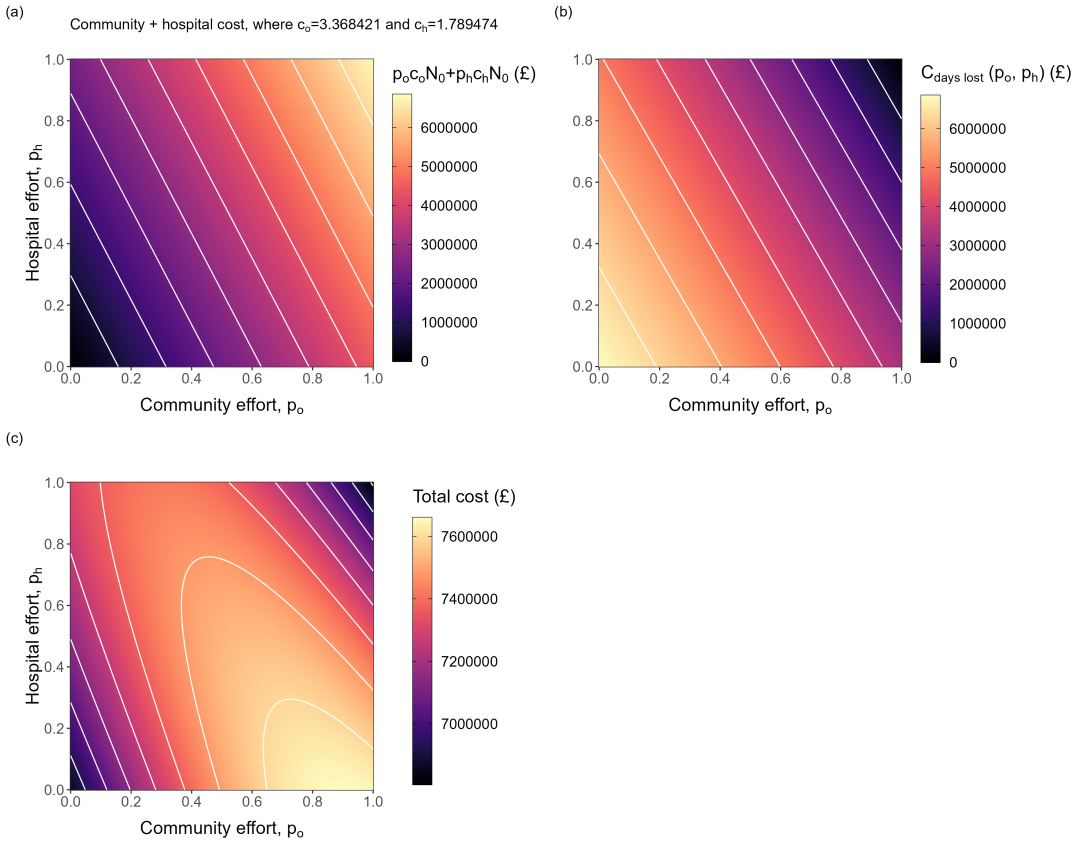

Fig. 5.7: Sensitivity of the total cost of COVID-19 absence to changes in community and hospital intervention efforts. We calculated the total cost in a community effort - hospital effort ($p_o - p_h$) parameter space using Eq. (5.4) for the different combinations. The community and hospital cost were fixed at $c_o = 3.37$ and $c_h = 1.79$. All other parameter values are given in Table 5.2. Panels (a)-(c) give a breakdown of the total cost (Eq. (5.4)): (a) shows the combined cost of the interventions ($p_o c_o N_0 + p_h c_h N_0$); (b) the cost of the days lost to COVID-19 absence ($C_{\text{days lost}}$; (c) the total cost of interventions and absences to the NHS, $\hat{J}$, given by Eq. (5.4).

Since the cost of days lost due to absence essentially changes linearly with respect to the intervention efforts (p_o, p_h), Figure 5.7 (b). Therefore $\frac{\partial}{\partial p_h} C_{\text{days lost}}(p_o, p_h)$ can be estimated by the gradient of a straight line. We estimate it for a given

community intervention effort (p_o) using

$$\frac{\partial}{\partial p_h} C_{\text{days lost}}(p_o, p_h) = \frac{C_{\text{days lost}}(p_o, p_h = 1) - C_{\text{days lost}}(p_o, p_h = 0)}{1 - 0}. \tag{5.21}$$

Similarly, we estimate the change in days lost due to absence for an increase in the community intervention effort by

$$\frac{\partial}{\partial p_o} C_{\text{days lost}}(p_o, p_h) = C_{\text{days lost}}(p_o = 1, p_h) - C_{\text{days lost}}(p_o = 0, p_h). \tag{5.22}$$

5.4 Discussion

The COVID-19 pandemic has profoundly impacted various aspects of healthcare systems worldwide, including the sickness absence rates among healthcare staff. In this paper, we developed a mechanistic compartmental model that uses publicly available COVID-19 surveillance data to estimate COVID-19-related sickness absence for the NHS England healthcare workforce. We parameterized the model specifically to NHS England and investigated the sensitivity of the transmission parameters, providing insights into the dynamics of the model. Moreover, our study demonstrates how this model can assess the impact and cost-effectiveness of interventions to reduce COVID-19-related sickness absence. We examined the sensitivity of the optimal intervention strategy in minimizing the economic cost of sickness absence to the NHS, considering both absence rate reduction benefits and intervention costs. This integration of an economic model with an epidemiological system provides a comprehensive framework for evaluating the cost-effectiveness

of interventions to control sickness absence [45]. Our findings highlight the crucial relationship between hospitalisations, the number of COVID-19 infections, and staff sickness absence, emphasising the need to prioritise measures that effectively mitigate transmission rates and minimise the burden on healthcare workers.

Our parameter estimation and sensitivity analysis showed that the transmission pathways through which staff acquire COVID-19 infections could be represented effectively with a force of infection dependent on the incidence of new infections and hospitalisations. This is consistent with findings outlined in the previous chapter and reminiscent of similar findings from influenza models [21–23]. The mechanistic model accurately captures the COVID-19 absence trend from July 2020 until November 2022. However, a slight divergence in accuracy was observed between January 2022 and April 2022, where the model exhibits an under- and overestimation of the absence rate. This discrepancy may be attributed to evolving dynamics of COVID-19, such as the impact of vaccinations and emerging variants, or changes in policy or perception of risk not accounted for in our model. For example, in February 2022, the isolation period was reduced to 6 days total, where staff could return to work after showing two negative lateral flow tests 24 hours apart [46]. Moreover, our model underestimates absences during periods of low sickness absence (the 'troughs' in the waves). An additional feature of our model is its predictive capability, enabling future sickness absence rates to be forecasted based on estimates of new hospitalisations and infections within the population.

In our model, we incorporated two COVID-19 surveillance data streams for the forces of infection, hospitalisations and ONS estimated incidence. We assumed that NHS staff primarily work in hospitals and have contact with the general community and their household outside of work. The hospitalisations stream represents their contact with patients in hospitals that results in infection, the ONS estimated incidence represents those contacts with the general community, and the staff-staff transmission part represents contact with other staff. However, parameter estimation and sensitivity analysis suggested an overlap between the streams. The staff-staff transmission parameter is not required for estimating the NHS England trend, which suggests the role of staff-staff transmission can be captured implicitly with the two proxy variables of COVID-19 activity (hospitalisations, ONS incidence).

During the sensitivity analysis to the forces of infection, we found that the community and hospitalisation transmission parameters mainly impacted different points during the study period. This is inherently linked to the changes in dynamics in the ONS and hospitalisation data streams. The rollout of COVID-19 vaccines (began in December 2020, and by March and September 2021 $\approx$ 30% and $\approx$ 70% of adults had their first dose [47]) reduced the number of people hospitalised (or the number of severe illnesses) relative to the number infected [48]. The rise in milder but more transmissible sub-variants of COVID-19 in late 2021 may have had a similar effect [49, 50]. If we assume admissions are a

proxy for the infection pressure in the workplace for most NHS staff (hospitals). Vaccines reducing the number of admissions lowers the infection pressure from this source and may explain why the hospitalisation source impacts the December 2020 – February 2021 wave in sickness absence more strongly. To conclude, each information stream explaining different periods between June 2020-July 2022 highlights that the relationship between COVID-19 absence and the indicators of COVID-19 activity was dynamic over the study period.

The sensitivity analysis suggests that the force of infection dependent on the number of community infections dominates the infection dynamics. This suggests that the ONS incidence data stream contains more information about the pathways staff become ill with COVID-19 than new hospital admissions. Furthermore, during our analysis of interventions, we observe an asymmetry between the community and hospital interventions. The threshold cost at which investing in the community intervention is no longer optimal is higher than the hospital intervention's threshold cost. One interpretation of these findings is that the community intervention demonstrates greater cost-effectiveness, as a unit of intervention effort in this intervention has a more substantial impact on reducing the absence rate.

The community intervention reduces the force of infection acting on staff, depending on the number of people testing positive for COVID-19 in the community. Any intervention reducing staff's contact with COVID-19-infected individuals in

the community could be considered part of this intervention. For example, social distancing, hand-washing, and encouraging staff to mix in small groups outside of work, not with people in places with high COVID-19 risk. It may also involve enhancing the effectiveness of tracing and quarantining infected individuals through improved testing availability and track-and-trace systems.

The optimal intervention strategy, which combines the hospital and community interventions, was determined based on the relative costs compared to the savings from avoiding staff absence. This approach separates our study from the COVID-19 HCW intervention modelling by Evans et al. [13], Qui et al. [14], and Pham et al. [16], since they did not consider the intervention costs. Using semi-analytic methods, we identified specific cost thresholds that governed the optimal strategy. Four possible strategies emerged, varying between full investment and no investment in each intervention. When the community intervention cost (c_o) is below 3.8 times the cost of a day lost, and the hospital intervention cost (c_h) is less than 1.3 times the cost of a day lost, full investment in both interventions is optimal. In this scenario, the hospital and community interventions are assigned maximum effort. If the community intervention cost is greater than 3.8 times the cost of a day lost, but the hospital intervention cost remains below 1.3 times the cost of a day lost, it becomes optimal to invest fully in the hospital intervention while disregarding the community intervention. Conversely, when the community intervention cost is less than 2.8 times the cost of a day lost, but the hospital intervention cost exceeds 2.4 times the cost of a day lost, it be-

comes optimal to fully invest in the community intervention and not at all in the hospital intervention. In all other scenarios, it is optimal not to invest in either intervention (the "do nothing" strategy). In summary, the optimal intervention strategy depends on the relative costs involved, and there are specific thresholds that dictate the preferred investment approach. The sensitivity to costs highlights the significance of proactive planning and ensuring that decision makers are informed with accurate information.

The impact of one intervention varied depending on whether we were already investing in another intervention. For example, when we were not investing in community intervention, the absence cost was high due to the high cost of days lost due to absence. Consequently, investing in hospital intervention had a limited impact. As we assessed the optimal hospital intervention as a function of cost, it quickly became too costly to bother investing. In contrast, when we were already investing in community intervention, the cost of days lost due to absence became a minor and more manageable issue. As a result, the hospital intervention had a more significant impact on reducing absence costs. In this scenario, the cost of the hospital intervention would need to be higher before we would be discouraged from investing.

Our model assumes a static cost associated with each day of absence, capturing expenses related to sick pay and shift cover. However, it is important to acknowledge that our model does not account for non-constant externalities that arise

from staff absences. These externalities include the increased workload for staff members who cover shifts and the potential virus transmission from healthcare workers to the general community, particularly within households. Moreover, our model does not incorporate tipping points where an excessive number of staff absences could significantly impact the functioning of the health service, potentially leading to system collapse. Such scenarios could result in heightened social and economic costs, compromised infection control procedures, and increased transmission to the community.

Our assessment of new interventions does not consider pre-existing interventions/restrictions, such as lockdowns and social distancing protocols, implemented between June 2020 and 2022 in England [51]. The impact of these existing interventions is included in the data streams in the force of infection. The interventions we evaluate are additional or assumed improvements upon the existing measures. Moreover, our model assumes a linear relationship between interventions and staff contact rates, overlooking the potential nonlinear impact and the inability to eliminate transmission from any source. For instance, specific interventions may reach a saturation point of effectiveness while maintaining a constant cost. Additionally, interventions may not uniformly impact the entire staff population. The effectiveness of interventions can vary among individuals or different groups within the staff population. Furthermore, it is unlikely that interventions were predetermined and not reviewed after the waves of COVID-19. For instance, social distancing measures in England were scaled up and down between outbreaks.

Our model does not explicitly distinguish between the various infection pathways for NHS staff, such as whether they contracted the infection from a patient, family member, or colleague. Additionally, it does not explicitly assess intervention strategies related to cohorting or testing. Further research is required to accurately capture these infection pathways and evaluate more specific intervention strategies. One way to achieve this is by adapting our force of infection and expanding the model to include additional populations, such as patients and the general community. Furthermore, our model assumes that staff members acquire complete immunity after a COVID-19 infection, with no possibility of reinfection. However, there has been evidence of reinfection [52], with HCWs at a slightly higher risk than the general population [53].

Bibliography

[1] NHS Digital, "NHS Workforce Statistics - April 2022 (Including selected provisional statistics for May 2022)," NHS Digital, Tech. Rep., 2022. [Online]. Available: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-workforce-statistics/april-2022

[2] B. Palmer and L. Rolewicz, "All is not well: Sickness absence in the NHS in England," Tech. Rep., 2023. [Online]. Available: https://www.nuffieldtrust.org.uk/resource/all-is-not-well-sickness-absence-in-the-nhs-in-england

[3] B. Palmer, "Chart of the week: The rise, fall and rise again of NHS staff sickness absences," Nuffield Trust, Tech. Rep., 2022. [Online]. Available: https://www.nuffieldtrust.org.uk/resource/chart-of-the-week-the-rise-fall-and-rise-again-of-nhs-staff-sickness-absences

[4] H. Mooney, "NHS faces double emergency of unprecedented demand and high COVID-19-related staff absences," Tech. Rep., 2022. [Online]. Available: https://www.nhsconfed.org/news/nhs-faces-double-emergency-unprecedented-demand-and-high-covid-related-staff-absences

[5] G. Cao, Z. Guo, J. Liu, and M. Liu, "Change From Low to Out-of-season Epidemics of Influenza in China During the COVID-19 Pandemic: A Time Series Study," *Journal of Medical Virology*, vol. 95, 2023. [Online]. Available: https://doi.org/10.1002/jmv.28888

[6] L. McCay, "Covid continues to disrupt: what is the plan to deal with it?" *BMJ*, vol. 378, 2022. [Online]. Available: https://www.bmj.com/content/378/bmj.o1780

[7] J. B. Adams and R. M. Walls, "Supporting the Health Care Workforce During the COVID-19 Global Epidemic," *Jama*, vol. 325, no. 15, 2020. [Online]. Available: https://jamanetwork.com/journals/jama/fullarticle/2763136

[8] A. S. V. Shah, R. Wood, C. Gribben, D. Caldwell, J. Bishop, A. Weir, S. Kennedy, M. Reid, A. Smith-Palmer, D. Goldberg, J. Mcmenamin, C. Fischbacher, C. Robertson, S. Hutchinson, P. Mckeigue,

H. Colhoun, and D. A. Mcallister, "Risk of hospital admission with coronavirus disease 2019 in healthcare workers and their households: nationwide linkage cohort study," *BMJ*, vol. 371, 2020. [Online]. Available: https://doi.org/10.1136/bmj.m3582

[9] C. E. Coltart, D. Wells, E. Sutherland, and A. Fowler, "National cross-sectional survey of 1.14 million NHS staff SARS-CoV-2 serology tests: a comparison of NHS staff with regional community seroconversion rates," *BMJ Open*, vol. 11, no. 7, 2021. [Online]. Available: http://dx.doi.org/10.1136/bmjopen-2021-049703

[10] E. McTaggart, J. Bowers, I. Megiddo, and A. Kleczkowski, "Sickness Absence Rates in NHS England Staff during the COVID-19 Pandemic," 2023, unpublished draft.

[11] J. Appleby, "NHS sickness absence during the covid-19 pandemic," *BMJ*, vol. 372, 2021. [Online]. Available: http://dx.doi.org/10.1136/bmj.n471

[12] D. A. Van Der Plaat, R. Edge, D. Coggon, M. van Tongeren, R. Muiry, V. Parsons, P. Cullinan, I. Madan, and S. Thomas, "Impact of COVID-19 pandemic on sickness absence for mental ill health in National Health Service staff," *BMJ Open*, vol. 11, 2021. [Online]. Available: http://dx.doi.org/10.1136/bmjopen-2021-054533

[13] S. Evans, E. Agnew, E. Vynnycky, J. Stimson, A. Bhattacharya, C. Rooney, B. Warne, and J. Robotham, "The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics

of COVID-19 in English hospitals," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, 2021. [Online]. Available: https://doi.org/10.1098/rstb.2020.0268

[14] X. Qiu, J. C. Miller, D. R. MacFadden, and W. P. Hanage, "Evaluating the contributions of strategies to prevent sars-cov-2 transmission in the healthcare setting: a modelling study," *BMJ Open*, vol. 11, no. 3, 2021. [Online]. Available: https://bmjopen.bmj.com/content/11/3/e044644

[15] E. Aguilar, N. J. Roberts, I. Uluturk, P. Kaminski, J. W. Barlow, A. G. Zori, L. Hébert-Dufresne, and B. D. Zusman, "Adaptive staffing can mitigate essential worker disease and absenteeism in an emerging epidemic," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 34, 2021. [Online]. Available: https://doi.org/10.1073/pnas.2105337118

[16] T. M. Pham, H. Tahir, J. H. H. M. van de Wijgert, B. R. Van der Roest, P. Ellerbroek, M. J. M. Bonten, M. C. J. Bootsma, and M. E. Kretzschmar, "Interventions to control nosocomial transmission of SARS-CoV-2: a modelling study," *BMC Medicine*, vol. 19, no. 211, pp. 1–16, 2021. [Online]. Available: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-021-02060-y

[17] Y. M. Lee and L. An, "Modeling Dynamics of Workforce Absenteeism and Effectiveness of Mitigation Actions During Pandemics," in *Proceedings of the System Dynamics Conference*, 2007. [Online]. Available: https://proceedings.systemdynamics.org/2007/proceed/papers/AN375.pdf

[18] P. Dhankhar, "Software to Estimate the Impact of an Influenza Pandemic on Work Day Loss: FluWorkLoss 1.0," 2023. [Online]. Available: https://www.cdc.gov/flu/pandemic-resources/tools/fluworkloss.htm

[19] N. Wilson, M. Baker, P. Crampton, and O. Mansoor, "The potential impact of the next influenza pandemic on a national primary care medical workforce," *Human Resources for Health*, vol. 3, no. 1, pp. 1–6, 2005. [Online]. Available: https://link.springer.com/articles/10.1186/1478-4491-3-7https://link.springer.com/article/10.1186/1478-4491-3-7

[20] R. E. Nap, M. P. Andriessen, N. E. Meessen, and T. S. Van Der Werf, "Pandemic Influenza and Hospital Resources," *Emerging Infectious Diseases*, vol. 13, no. 11, pp. 1714–1719, 2007. [Online]. Available: https://wwwnc.cdc.gov/eid/article/13/11/07-0103_article

[21] D. K. Ip, E. H. Lau, Y. H. Tam, H. C. So, B. J. Cowling, and H. K. Kwok, "Increases in absenteeism among health care workers in Hong Kong during influenza epidemics, 2004-2009," *BMC Infectious Diseases*, vol. 15, no. 1, pp. 1–9, 2015. [Online]. Available: https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-015-1316-y

[22] D. L. Schanzer, H. Zheng, and J. Gilmore, "Statistical estimates of absenteeism attributable to seasonal and pandemic influenza from the Canadian Labour Force Survey," *BMC Infectious Diseases*, vol. 11, no. 1, pp. 1–9, 2011. [Online]. Available: https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-11-90

[23] D. W. Challener, L. E. Breeher, J. Frain, M. D. Swift, P. K. Tosh, and J. O'horo, "Healthcare personnel absenteeism, presenteeism, and staffing challenges during epidemics," *Infection Control & Hospital Epidemiology*, vol. 42, pp. 388–391, 2021. [Online]. Available: https://doi.org/10.1017/ice.2020.453

[24] A. Basurto, H. Dawid, P. Harting, J. Hepp, and D. Kohlweyer, "Economic and Epidemic Implications of Virus Containment Policies: Insights from Agent-Based Simulations," *Bielefeld Working Papers in Economics and Management*, no. 05, 2020. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635329

[25] A. Duarte, S. Walker, A. Metry, R. Wong, J. Panovska-Griffiths, and M. Sculpher, "Jointly Modelling Economics and Epidemiology to Support Public Policy Decisions for the COVID-19 Response: A Review of UK Studies," *PharmacoEconomics*, vol. 39, 2021. [Online]. Available: https://doi.org/10.1007/s40273-021-01045-2

[26] D. Haw, C. Morgenstern, G. Forchini, R. Johnson, P. Doohan, P. C. Smith, and K. Hauck, "Data Needs for Integrated Economic-Epidemiological Models of Pandemic Mitigation Policies," *Epidemics*, vol. 41, p. 100644, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1755436522000846

[27] T. Blakely, J. Thompson, L. Bablani, P. Andersen, D. Ait Ouakrim, N. Carvalho, P. Abraham, M.-A. Boujaoude, A. Katar, E. Akpan, N. Wilson, and M. Stevenson, "Association of Simulated COVID-19 Policy Responses for Social Restrictions and Lockdowns With Health-Adjusted Life-Years and Costs in Victoria, Australia," *JAMA Health Forum*, vol. 2, no. 7, pp. e211 749–e211 749, 07 2021. [Online]. Available: https://doi.org/10.1001/jamahealthforum.2021.1749

[28] F. E. Alvarez, D. Argente, and F. Lippi, "a Simple Planning Problem for Covid-19 Lockdown," *Geospatial Health*, vol. 17, no. s1, 2022. [Online]. Available: https://doi.org/10.4081/gh.2022.1022

[29] L. Miclo, D. Spiro, and J. Weibull, "Optimal epidemic suppression under an ICU constraint: An analytical solution," *Journal of Mathematical Economics*, vol. 101, p. 102669, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304406822000258

[30] T. Colbourn, J. Panovska-griffiths, D. Ph, G. Colbourn, D. Ph, R. A. Burgess, D. Ph, and K. Ouyang, "Modelling the health and economic impacts of Population-wide Testing, contact Tracing and Isolation (PTTI) strategies for COVID-19 in the UK," *The Lancet*, 2020. [Online]. Available: https://dx.doi.org/10.2139/ssrn.3627273

[31] J. Bowers, A. Kleczkowski, and M. Macpherson, "Modelling the effect of Covid-19 on NHS absences and possible mitigating actions to ensure staff availability," 2021, unpublished draft.

[32] NHS Digital, "NHS Workforce Statistics - June 2020," NHS Digital, Tech. Rep., 2020. [Online]. Available: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-workforce-statistics

[33] K. Steel and E. Fordham, "Coronavirus (COVID-19) Infection Survey: methods and further information," Office for National Statistics, Tech. Rep., 2022. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/covid19infectionsurveypilotmethodsandfurtherinformation#incidence

[34] UK Health Security Agency, "England Summary | Coronavirus (COVID-19) in the UK," Tech. Rep., 2022. [Online]. Available: https://coronavirus.data.gov.uk/

[35] K. A. Walsh, S. Spillane, L. Comber, K. Cardwell, P. Harrington, J. Connell, C. Teljeur, N. Broderick, C. F. de Gascun, S. M. Smith, M. Ryan, and M. O'Neill, "The duration of infectiousness of individuals infected with SARS-CoV-2," *Journal of Infection*, vol. 81, no. 6, pp. 847–856, 2020. [Online]. Available: https://doi.org/10.1016/j.jinf.2020.10.009

[36] M. D. Van Kerkhove, "COVID-19: When are you most infectious?" *The Conversation*, 6 2020. [Online]. Available: https://theconversation.com/covid-19-when-are-you-most-infectious-150760

[37] M. Cevik, M. Tate, O. Lloyd, A. E. Maraolo, J. Schafers, and A. Ho, "SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis," *The Lancet Microbe*, vol. 2, no. 1, pp. e13–e22, 2021. [Online]. Available: http://dx.doi.org/10.1016/S2666-5247(20)30172-5

[38] Q. Ma, J. Liu, Q. Liu, L. Kang, R. Liu, W. Jing, Y. Wu, and M. Liu, "Global Percentage of Asymptomatic SARS-CoV-2 Infections among the Tested Population and Individuals with Confirmed COVID-19 Diagnosis: A Systematic Review and Meta-analysis," *JAMA Network Open*, vol. 4, no. 12, pp. e2 137 257–e2 137 257, 12 2021. [Online]. Available: https://doi.org/10.1001/jamanetworkopen.2021.37257

[39] National Health Service (NHS) England, "Standard Operating Procedure for LFD Rollout for Asymptomatic Staff Testing (Phase 2 Trusts)," National Health Service (NHS) England, Tech. Rep., 11

2020. [Online]. Available: https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/11/C0873_i_SOP_LFD-rollout-for-asymptomatic-staff-testing_phase-2-trusts-v1.1_16-nov20.pdf

[40] Government of the United Kingdom, "NHS Test and Trace: what to do if you are contacted," GOV.UK, Tech. Rep., 2022. [Online]. Available: https://www.gov.uk/guidance/nhs-test-and-trace-how-it-works#:~:text=Ifyouliveinthe,thedateoftheirtest

[41] National Health Service (NHS) England, "Letter on COVID-19 Testing to Support Retention of NHS Staff," 2020. [Online]. Available: https://www.england.nhs.uk/coronavirus/publication/letter-covid-19-testing-to-support-retention-of-nhs-staff/

[42] UK Health Security Agency, "Metrics documentation | Coronavirus in the UK," Tech. Rep., 2022. [Online]. Available: https://coronavirus.data.gov.uk/metrics/doc/newAdmissions#england

[43] K. Soetaert and T. Petzoldt, "Inverse modelling, sensitivity and monte carlo analysis in R using package FME," *Journal of Statistical Software*, vol. 33, no. 3, pp. 1–28, 2010. [Online]. Available: https://doi.org/10.18637/jss.v033.i03

[44] R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, 2023. [Online]. Available: https://www.r-project.org/

[45] K. D. S. Yu and K. B. Aviso, "Modelling the Economic Impact and Ripple Effects of Disease Outbreaks," *Process Integration and Optimization for Sustainability*, vol. 4, no. 2, pp. 183–186, 2020. [Online]. Available: https://doi.org/10.1007/s41660-020-00113-y

[46] National Health Service (NHS) England, "NHS Test and Trace in the workplace," GOV.UK, Tech. Rep., 2022. [Online]. Available: https://www.gov.uk/guidance/nhs-test-and-trace-workplace-guidance?priority-taxon=09944b84-02ba-4742-a696-9e562fc9b29d#history

[47] E. Mathieu, H. Ritchie, L. Rodés-Guirao, C. Appel, C. Giattino, J. Hasell, B. Macdonald, S. Dattani, D. Beltekian, E. Ortiz-Ospina, and M. Roser, "Coronavirus Pandemic (COVID-19)," *Our World in Data*, 2020. [Online]. Available: https://ourworldindata.org/coronavirus

[48] T. Shi, C. Robertson, and A. Sheikh, "Effectiveness and safety of coronavirus disease 2019 vaccines," *Current Opinion in Pulmonary Medicine*, vol. 29, no. 3, 2023. [Online]. Available: https://doi.org/10.1097/mcp.0000000000000948

[49] S. Islam, T. Islam, and M. R. Islam, "New Coronavirus Variants are Creating More Challenges to Global Healthcare System: A Brief Report on the Current Knowledge," *Clinical Pathology*, vol. 15, 2022. [Online]. Available: https://doi.org/10.1177/2632010x221075584

[50] M. Mohsin and S. Mahmud, "Omicron SARS-CoV-2 variant of concern: A review on its transmissibility, immune evasion, reinfection, and severity," *Medicine (United States)*, vol. 101, no. 19, 2022. [Online]. Available: https://doi.org/10.1097/md.0000000000029165

[51] H. Thomas, P. Anna, P. Toby, , A. Jessica, A. d. M. Bernardo, A. Noam, B. Roy, B. Thomas, , C.-B. Emily, C. Alice, D. F. Martina, E. Benjamin, E. Lucy, , E. Jodie, F. Rodrigo, G. R. Liz, G. Kaitlyn, G. Rafael, Laura, Hallas, K. Nadezhda, K. Beatriz, L. Sandhya, L. Maria, Saptarshi, Majumdar, M. O. Thayslene, N. Radhika, P. Annalena, R. Luyao, , S. Julia, T. Helen, T. Will, W. Adam, W. Samuel, W. Andrew, , Z. Hao, Z. Yuxi, and V. Andrea, "Variation in government responses to COVID-19," Blavatnik School of Government, Tech. Rep., 2023. [Online]. Available: www.bsg.ox.ac.uk/covidtracker

[52] M. S. Graham, C. H. Sudre, A. May, M. Antonelli, B. Murray, T. Varsavsky, K. Kläser, L. S. Canas, E. Molteni, M. Modat, D. A. Drew, L. H. Nguyen, L. Polidori, S. Selvachandran, C. Hu, J. Capdevila, C.-. G. Consortium, A. Hammers, A. T. Chan, J. Wolf, T. D. Spector, C. J. Steves, and S. Ourselin, "Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: an ecological study," *The Lancet Public Health*, vol. 6, no. 5, pp. e335–e345, 2021. [Online]. Available: https://doi.org/10.1016/S2468-2667(21)00055-4

[53] X. Trujillo, O. Mendoza-Cano, M. Ríos-Silva, M. Huerta, J. Guzmán-Esquivel, V. Benites-Godínez, A. Lugo-Radillo, J. A. Bricio-Barrios, M. I. Cárdenas-Rojas, E. F. Ríos-Bracamontes, V. M. Ortega-Macías, V. Ruiz-Montes de Oca, and E. Murillo-Zamora, "Predictors of Recurrent Laboratory-Confirmed Symptomatic SARS-CoV-2 Infections in a Cohort of Healthcare Workers," *Vaccines*, vol. 11, no. 3, p. 626, 2023. [Online]. Available: https://doi.org/10.3390/vaccines11030626

Chapter 6

Conclusions and Future Work

The main objective of this doctoral thesis was to explore the use of epidemiological models in understanding disease dynamics, assessing their impact on different systems, and developing effective mitigation strategies. Through four different papers, we investigated the effects of pests and pathogens on forest harvesting regimes, the transmission dynamics of COVID-19 in care homes, and sickness absence rates among NHS England staff during the COVID-19 pandemic. Each paper has addressed unique challenges posed by disease outbreaks, and the findings contribute to the field of epidemiology by providing evidence-based insights for effective disease control and management.

The first paper (Chapter 2) investigated the impact of pests and pathogens on forest harvesting regimes using a bioeconomic model that integrates ecological, epidemiological, and economic factors. Our study provided insights for forest

managers in making informed decisions regarding control and mitigation strategies. By analysing optimal harvesting strategies and considering the interactions between harvesting, tree growth, and disease progression, we identified complex dynamics and trade-offs between maximising return on harvest yields and mitigating the spread and impact of disease. Moreover, we specifically focused on thinning as a key intervention within harvesting strategies and demonstrated its potential to improve the net present value of investments significantly. Overall, this paper offers a framework that can help design effective forest management strategies in the presence of disease, with practical implications for decision-makers and stakeholders seeking to minimise economic losses.

In the second paper (Chapter 3), we employed a compartmental metapopulation model to investigate the transmission dynamics of COVID-19 within care homes in Scotland. Our study focused on the intra-subpopulation mixing patterns between care home residents, care home staff, and the general population across a network of care homes. By analysing these dynamics, we provided insights into the factors influencing disease transmission and identified potential mitigation strategies to minimise the impact of future outbreaks. Our findings showed that a complete restriction on staff sharing across care homes or community visitation alone is insufficient as a control measure. However, having staff live exclusively in the care homes where they work, with no staff sharing, effectively reduces COVID-related deaths among care home residents. Moreover, implementing restrictions on all population movements, including visitation and

staff sharing, substantially reduces resident cases. Our findings contribute to developing targeted interventions and management strategies to mitigate the spread of COVID-19 in vulnerable care home settings. Additionally, they highlight the importance of robust planning and support for care homes and their staff, enabling effective responses to emerging pandemics.

In the third paper (Chapter 4), we investigated the impact of the COVID-19 pandemic on sickness absence rates among NHS England staff. Our research aimed to provide insights into the pandemic's effect on healthcare workers and identify factors influencing sickness absence rates. Our analysis of sickness absence trends highlighted significant increases in sickness absence rates coinciding with the arrival of COVID-19 in England, which continued to rise throughout the pandemic. High COVID-19 activity periods strongly correlated with staff absence, and the main categories driving the dynamics were COVID-19-related or mental health absences. We demonstrated that sickness absences in these two categories could be accurately estimated using multivariate regression and time series models, respectively. Moreover, we showed that indicators of COVID-19 activity contain useful information about staff infection pathways. Our findings can inform targeted interventions and policies to reduce sickness absence, improve workforce health, and enhance productivity. Additionally, they contribute to understanding the pandemic's impact on healthcare systems, guiding workforce management and public health preparedness, and highlighting areas for future research.

In the fourth and final paper (Chapter 5), we developed a mechanistic compartmental model that uses publicly available COVID-19 surveillance data to estimate COVID-19 sickness absence among the NHS England workforce. Our study integrated an economic model with an epidemiological system to create a comprehensive framework for evaluating the cost-effectiveness of interventions to control sickness absence. We identified strategies to reduce disease-related absenteeism while considering resource constraints and epidemiological dynamics by analysing staff-related transmission dynamics and using optimisation techniques. The model's outcomes highlight the importance of the relative costs of interventions compared to the savings from preventing staff absence. In our model, interventions that reduce staff's contact with COVID-19-infected individuals in the community were the most cost-efficient. Moreover, our research highlights the crucial relationship between hospitalisations, the number of COVID-19 infections, and staff sickness absence, emphasising the need to prioritise measures that effectively mitigate transmission rates and minimise the burden on healthcare workers. The findings contribute to evidence-based decision-making in public health and offer practical guidelines for managing future disease outbreaks.

The collective findings of this thesis met the research objective, which was to explore the use of epidemiological models in understanding disease dynamics, assessing their impact on different systems, and developing effective mitigation strategies. Furthermore, a common framework, and process for investigating dis-

ease outbreaks in different systems using mathematical and statistical models, underlies the four papers in this doctoral thesis. This framework seeks to understand how diseases affect the dynamics of various systems and how adjustments in human behaviour influence the epidemiological aspects - core principles of the Epi/Bio-Economic approach [48, 50]. The initial status quo of each system sets the stage, the disease disrupts the system and shapes the epidemiology, and human behaviour adapts accordingly. Epidemiological models, particularly optimisation, serve as essential tools within this framework to help understand the interaction between disease dynamics and human behaviour. They help inform recommendations or guidelines on how human behaviour should be modified to mitigate the disease's effects and reduce transmission.

In conclusion, this doctoral thesis advances our understanding of the feedback loops between disease impacts and human responses, contributing to evidence-based decision-making. Integrating epidemiological models with the relevant ecological, economic, or social factors provides a powerful framework for designing disease management strategies. Policymakers, stakeholders, and managers can leverage the insights to develop targeted interventions, allocate resources efficiently, and enhance preparedness during outbreaks.

Future Directions

Strengthening our understanding of disease dynamics, impact, and interventions will further aid policymakers, stakeholders, and managers in preventing and mitigating infectious disease outbreaks.

We identified several future research directions in Chapter 2 (Paper 1). One possible direction was increasing the heterogeneity and complexity of harvesting regimes, such as biasing thinning towards infected trees or implementing multiple thinning operations during rotations. Another would be to consider the unfavourable or unintended impacts of thinning on some forest diseases, such as the release of *Heterobasidian Annosum* spores from cut trees leading to increased spread [72]. Exploring this interaction between thinning and disease spread and optimal harvesting under these scenarios requires further study. To address the single rotation assumption, incorporating multiple rotations with re-planting and assumptions about disease persistence between rotations would provide insights into longer-term management strategies. Given more time, extending the forest model with a spatial component through a metapopulation model would be interesting. It would allow the consideration of forest management in a connected landscape and provide management insights at the regional stakeholder level.

In Chapter 3 (Paper 2), we outlined future research directions to enhance the understanding of care home outbreaks. Further research could examine the impact of different contact structures between staff and residents at different homes

and the presence of highly connected worker populations (e.g., agency staff) on disease dynamics [73]. Incorporating the explicit modelling of symptomatic and asymptomatic individuals, self-isolation, and behavioural changes after infection could allow the investigation of more specific interventions for staff. Including differences in care home sizes would improve the representation of NHS Lothian and allow for investigating the impact of care home size heterogeneity on outbreak risk.

For Chapter 4 (Paper 3), future research directions include further investigation of the relationship between mental health sickness absence and COVID-19, possibly by incorporating COVID-19 activity indicators as regressors in the time series model. Furthermore, the linear regression models used in the study do not fully explain the dynamics behind COVID-19-related absences, and there is limited consideration for local epidemiology, staff categories, COVID-19 vaccination, and other seasonal respiratory infections. Further investigations could address these limitations to better understand the sickness absence rates among NHS staff during the pandemic.

We identified several future research directions in Chapter 5 (Paper 4). To accurately capture various infection pathways for NHS staff, future research could expand the model to include additional populations, such as patients and the general community. This would enable the assessment of specific intervention strategies related to cohorting or testing. Furthermore, the model exhibited slight

divergences in accuracy between January 2022 and April 2022, possibly due to the evolving dynamics of COVID-19, such as the impact of vaccinations and emerging variants or changes in policy or perception of risk [74–76]. Future research could focus on understanding these dynamics and updating the model to account for them. The model also assumed complete immunity after COVID-19 infection, while reinfections have been reported [77]. Future research could consider the possibility of reinfection.

At the time of writing this doctoral thesis, the PhD author is currently a research assistant on a NERC-funded interdisciplinary project, NE/V019988/1, "Learning to adapt to an uncertain future: linking genes, trees, people and processes for more resilient treescapes (newLEAF)". This project addresses how quickly trees can adapt to climate change and whether human intervention is needed to protect their future. In particular, the project has highlighted the continuous and long-range pollen dispersal by trees, meaning forests impact their neighbours' genetic diversity and trait evolution. Additionally, individual decisions have landscape consequences for pest and pathogen outbreaks. There is an unexplored need to develop methodologies to understand the consequences of strategic planting decisions in landscapes with multiple forest owners and objectives and the background of a changing climate. This research direction would further build upon the methodologies and future directions highlighted in Chapter 2 of this thesis. By prioritizing this avenue of research, we can help navigate these interconnected and pressing issues in forest management.

Bibliography for Introduction and Conclusion

[1] R. E. Baker, A. S. Mahmud, I. F. Miller, M. Rajeev, F. Rasambainarivo, B. L. Rice, S. Takahashi, A. J. Tatem, C. E. Wagner, L. F. Wang, A. Wesolowski, and C. J. E. Metcalf, "Infectious disease in an era of global change," *Nature Reviews Microbiology*, vol. 20, no. 4, pp. 193–205, 2022. [Online]. Available: https://doi.org/10.1038/s41579-021-00639-z

[2] K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak, "Global Trends in Emerging Infectious Diseases," *Nature*, vol. 451, pp. 990–993, 2008. [Online]. Available: https://doi.org/10.1038/nature06536

[3] D. E. Bloom, K. Prettner, and D. E. Bloom, "Modern Infectious Diseases : Macroeconomic Impacts and Policy Responses," Tech. Rep., 2020. [Online]. Available: https://docs.iza.org/dp13625.pdf

[4] S. Savary, S. Bregaglio, L. Willocquet, D. H. Gustafson, D. M. D'Croz, A. E. T. Sparks, N. P. Castilla, A. Djurle, C. Allinne, M. Sharma, V. Rossi, L. Amorim, A. Bergamin, J. Yuen, P. D. Esker, N. McRoberts, J. Avelino, E. Duveiller, J. Koo, and K. A. Garrett, "Crop Health and Its Global Impacts on the Components of Food Security," *Food Security*, vol. 9, pp. 311–327, 2017. [Online]. Available: https://doi.org/10.1007/s12571-017-0659-1

[5] M. C. Fisher, D. A. Henk, C. J. Briggs, J. S. Brownstein, L. C. Madoff, S. L. McCraw, and S. J. Gurr, "Emerging fungal threats to animal, plant and ecosystem health," *Nature*, vol. 484, no. 7393, pp. 186–194, 2012. [Online]. Available: https://doi.org/10.1038/nature10947

[6] C. Armitage, "The high burden of infectious disease," *Nature*, vol. 598, no. s9, 2021. [Online]. Available: https://doi.org/10.1038/d41586-021-02909-5

[7] World Health Organization, "Summary of Probable SARS Cases with Onset of Illness from 1 November 2002 to 31 July 2003," 2023. [Online]. Available: https://www.who.int/publications/m/item/summary-of-probable-sars-cases-with-onset-of-illness-from-1-november-2002-to-31-july-2003

[8] L. Simonsen, P. Spreeuwenberg, R. Lustig, R. J. Taylor, D. M. Fleming, M. Kroneman, M. D. Van Kerkhove, A. W. Mounts, W. J. Paget, H. Echenique, V. Savy, D. Muscatello, C. R. MacIntyre, D. E. Dwyer, E. Azziz-Baumgartner, N. Homaira, F. E. A. Moura, C. Schuck, H. Akwar, D. Schanzer, R. Fuentes, A. Olea, V. Sotomayor, L. Feng, H. Yu, A. Mazick, K. Mølbak, J. Nielsen, F. Carrat, M. Lemaitre, U. Buchholz, B. Schweiger, M. Höhle, S. Vesenbeckh, B. Cowling, G. Leung, T. Tsang, S. K. Chuang, M. Bromberg, Z. Kaufman, N. Sugaya, K. Oka Ezoe, S. Hayashi, M. Matsuda, H. Lopez-Gatell, C. Alpuche-Aranda, D. Noyola, G. Chowell, L. van Asten, A. Meijer, K. van den Wijngaard, M. van der

Sande, M. Baker, J. Zhang, J. G. Benavides, C. Munayco, A. Laguna-Torres, D. Rabczenko, B. Wojtyniak, S. H. Park, Y. K. Lee, L. Zolotusca, O. Popovici, R. Popescu, L. W. Ang, J. Cutter, R. Lin, S. Ma, M. Chen, V. J. Lee, K. Prosenc, M. Socan, C. Cohen, A. Larrauri, S. de Mateo, L. S. Méndez, C. D. Sanz, N. Andrews, H. K. Green, R. Pebody, A. Saei, D. Shay, and C. Viboud, "Global Mortality Estimates for the 2009 Influenza Pandemic from the GLaMOR Project: A Modeling Study," *PLoS Medicine*, vol. 10, no. 11, 2013. [Online]. Available: https://doi.org/10.1371/journal.pmed.1001558

[9] Centers for Disease Control and Prevention, "2014-2016 Ebola Outbreak," 2023. [Online]. Available: https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html

[10] A. O. Adem and M. Amin, "Incidence and Severity of Major Diseases of Coffee in Highland of Eastern Ethiopia," *Journal of Food Nutrition and Agriculture*, vol. 4, 2021. [Online]. Available: https://doi.org/10.21839/jfna.2021.v4.6944

[11] A. Koutouleas and D. B. Collinge, "Coffee Leaf Rust Back with a Vengeance," British Society For Plant Pathology, Tech. Rep. September, 2022. [Online]. Available: www.bspp.org.uk.

[12] L. Ghelardini, N. Luchi, F. Pecori, A. L. Pepori, R. Danti, G. Della Rocca, P. Capretti, P. Tsopelas, and A. Santini, "Ecology of invasive forest pathogens," *Biological Invasions*, vol. 19, no. 11, pp. 3183–3200, 2017. [Online]. Available: https://doi.org/10.1007/s10530-017-1487-0

[13] A. Santini and A. Battisti, "Complex insect-pathogen interactions in tree pandemics," *Frontiers in Physiology*, vol. 10, no. MAY, pp. 1–7, 2019. [Online]. Available: https://doi.org/10.3389/fphys.2019.00550

[14] Our World in Data, "Cumulative Deaths and Cases: COVID-19," 2023. [Online]. Available: https://ourworldindata.org/grapher/cumulative-deaths-and-cases-covid-19

[15] D. U. Pfeiffer, M. J. Otte, D. Roland-Holst, and D. Zilberman, "A one health perspective on HPAI H5N1 in the Greater Mekong sub-region," *Comparative Immunology, Microbiology and Infectious Diseases*, vol. 36, no. 3, pp. 309–319, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.cimid.2012.11.005

[16] World Health Organization, "Ongoing Avian Influenza Outbreaks in Animals Pose Risk to Humans," 2023. [Online]. Available: https://www.who.int/news/item/12-07-2023-ongoing-avian-influenza-outbreaks-in-animals-pose-risk-to-humans

[17] K. F. Smith, S. R. Rosenthal, and S. Ramachandran, "Global Rise in Human Infectious Disease Outbreaks," *Journal of the Royal Society Interface*, vol. 11, 2014. [Online]. Available: https://doi.org/10.1098/rsif.2014.0950

[18] T. D. Ramsfield, B. J. Bentz, M. Faccoli, H. Jactel, and E. G. Brockerhoff, "Forest health in a changing world: Effects of globalization and climate change on forest insect and pathogen impacts," *Forestry*, vol. 89, no. 3, pp. 245–252, 7 2016. [Online]. Available: https://doi.org/10.1093/forestry/cpw018

[19] N. M. Naveed and N. D. .Syeda Sadia, "Twenty First Century- An Era of Infectious Diseases," *Annals of Pims-Shaheed Zulfiqar Ali Bhutto Medical University*, vol. 18, no. 2, 2022. [Online]. Available: https://doi.org/10.48036/apims.v18i2.678

[20] M. Roberts, C. A. Gilligan, A. Kleczkowski, N. Hanley, A. E. Whalley, and J. R. Healey, "The Effect of Forest Management Options on Forest Resilience to Pathogens," *Frontiers in Forests and Global Change*, vol. 3, no. 2, 2020. [Online]. Available: https://doi.org/10.3389/ffgc.2020.00007

[21] X. X. Wu, H. Y. Tian, S. Zhou, L. F. Chen, and B. Xu, "Impact of global change on transmission of human infectious diseases," *Science China Earth Sciences*, vol. 57, no. 2, pp. 189–203, 2014. [Online]. Available: https://doi.org/10.1007/s11430-013-4635-0

[22] A. Krämer and M. M. Hossain Khan, "Global Challenges of Infectious Disease Epidemiology," *Modern Infectious Disease Epidemiology*, pp. 23–38, 2009. [Online]. Available: https://doi.org/10.1007/978-0-387-93835-6_2

[23] T. Wu, "The Socioeconomic and Environmental Drivers of the COVID-19 Pandemic: A Review," *Ambio*, vol. 50, 2021. [Online]. Available: https://doi.org/10.1007/s13280-020-01497-4

[24] S. S. Morse, J. A. Mazet, M. Woolhouse, C. R. Parrish, D. Carroll, W. B. Karesh, C. Zambrana-Torrelio, W. I. Lipkin, and P. Daszak, "Prediction and prevention of the next pandemic zoonosis," *The Lancet*, vol. 380, no. 9857, pp. 1956–1965, 2012. [Online]. Available: http://dx.doi.org/10.1016/S0140-6736(12)61684-5

[25] S. Dennis and D. Fisher, "Climate Change and Infectious Diseases: The Next 50 Years," *Annals of the Academy of Medicine, Singapore*, vol. 47, no. 10, pp. 401–404, 2018. [Online]. Available: http://dx.doi.org/10.47102/annals-acadmedsg.V47N10p401

[26] J. O. Stireman, L. A. Dyer, D. H. Janzen, M. S. Singer, J. T. Lill, R. J. Marquis, R. E. Ricklefs, G. L. Gentry, W. Hallwachs, P. D. Coley, J. A. Barone, H. F. Greeney, H. Connahs, P. Barbosa, H. C. Morais, and I. R. Diniz, "Climatic unpredictability and parasitism of caterpillars: Implications of global warming," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 48, pp. 17 384–17 387, 2005. [Online]. Available: https://doi.org/10.1073/pnas.0508839102

[27] N. Ogden and P. Gachon, "Climate change and infectious diseases: What can we expect?" *Canada Communicable Disease Report*, vol. 45, no. 4, pp. 76–80, 2019. [Online]. Available: https://doi.org/10.14745/ccdr.v45i04a01

[28] P. R. Saunders-Hastings and D. Krewski, "Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission." *Pathogens (Basel, Switzerland)*, vol. 5, no. 4, 12 2016. [Online]. Available: https://doi.org/10.3390/pathogens5040066

[29] K. M. Smith, C. C. Machalaba, R. Seifman, Y. Feferholtz, and W. B. Karesh, "Infectious disease and economics: The case for considering multi-sectoral impacts," *One Health*, vol. 7, p. 100080, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235277141830034X

[30] R. N. Thompson and E. Brooks-Pollock, "Detection, Forecasting and Control of Infectious Disease Epidemics: Modelling Outbreaks in Humans, Animals and Plants," *Philosophical Transactions of the Royal Society B Biological Sciences*, vol. 374, 2019. [Online]. Available: https://doi.org/10.1098/rstb.2019.0038

[31] A. Huppert and G. Katriel, "Mathematical Modelling and Prediction in Infectious Disease Epidemiology," *Clinical Microbiology and Infection*, vol. 19, no. 11, 2013. [Online]. Available: https://doi.org/10.1111/1469-0691.12308

[32] N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, C. A. Pearson, B. J. Quilty, A. J. Kucharski, H. Gibbs, S. Clifford, A. Gimma, K. van Zandvoort, J. D. Munday, C. Diamond, W. J. Edmunds, R. M. Houben, J. Hellewell, T. W. Russell, S. Abbott, S. Funk, N. I. Bosse, Y. F. Sun, S. Flasche, A. Rosello, C. I. Jarvis, and R. M. Eggo, "Age-dependent effects in the transmission and control of COVID-19 epidemics," *Nature Medicine*, vol. 26, no. 8, pp. 1205–1211, 2020. [Online]. Available: https://doi.org/10.1038/s41591-020-0962-9

[33] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds, "Social contacts and mixing patterns relevant to the spread of infectious diseases," *PLoS Medicine*, vol. 5, no. 3, pp. 0381–0391, 2008. [Online]. Available: https://doi.org/10.1371/journal.pmed.0050074

[34] M. Barnett, G. Buchak, and C. Yannelis, "Epidemic responses under uncertainty," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 120, no. 2, 1 2023. [Online]. Available: https://doi.org/10.1073/pnas.2208111120

[35] L. Ge, M. C. M. Mourits, A. R. Kristensen, and R. B. M. Huirne, "A modelling approach to support dynamic decision-making in the control of FMD epidemics," *Preventive Veterinary Medicine*, vol. 95, pp. 167–174, 2010. [Online]. Available: https://doi.org/10.1016/j.prevetmed.2010.04.003

[36] M. J. Keeling, M. E. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell, "Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape," *Science*, vol. 294, no. 5543, pp. 813–817, 2001.

[37] N. Ferguson, C. Donnelly, and R. Anderson, "The foot-and-mouth epidemic in great britain: pattern of spread and impact of interventions," *Science*, vol. 292, pp. 1155–1160, 2001. [Online]. Available: https://doi.org/10.1126/science.1061020

[38] S. Riley and N. M. Ferguson, "Smallpox transmission and control: Spatial dynamics in Great Britain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 33, pp. 12 637–12 642, 2006. [Online]. Available: https://doi.org/10.1073/pnas.0510873103

[39] I. M. Hall, J. R. Egan, I. Barrass, R. Gani, and S. Leach, "Comparison of smallpox outbreak control strategies using a spatial metapopulation model," *Epidemiology and Infection*, vol. 135, no. 7, pp. 1133–1144, 2007. [Online]. Available: https://doi.org/doi:10.1017/S0950268806007783

[40] M. Baguelin, S. Flasche, A. Camacho, N. Demiris, E. Miller, and W. J. Edmunds, "Assessing Optimal Target Populations for Influenza Vaccination Programmes: An Evidence Synthesis and Modelling Study," *PLoS Medicine*, vol. 10, no. 10, 2013. [Online]. Available: https://doi.org/10.1371/journal.pmed.1001527

[41] E. Brooks-Pollock, L. Danon, T. Jombart, and L. Pellis, "Modelling that shaped the early covid-19 pandemic response in the uk," *Philosphical Transactions of the Royal Society*, vol. 376, p. 20210001, 2021. [Online]. Available: https://doi.org/10.1098/rstb.2021.0001

[42] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. C. Okell, S. Van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. Gt Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, and A. C. Ghani, "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand," Tech. Rep., 2020. [Online]. Available: https://doi.org/10.25561/77482

[43] T. M. Pham, H. Tahir, J. H. H. M. van de Wijgert, B. R. Van der Roest, P. Ellerbroek, M. J. M. Bonten, M. C. J. Bootsma, and M. E. Kretzschmar, "Interventions to control nosocomial transmission of SARS-CoV-2: a modelling study," *BMC Medicine*, vol. 19, no. 211, pp. 1–16, 2021. [Online]. Available: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-021-02060-y

[44] A. Kleczkowski, A. Hoyle, and P. McMenemy, "One model to rule them all? Modelling approaches across OneHealth for human, animal and plant epidemics," *Philosophical Transactions Of The Royal Society*, vol. 374, 2019. [Online]. Available: https://doi.org/10.1098/rstb.2018.0255

[45] W. O. Kermack, A. G. McKendrick, and G. T. Walker, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 115, no. 772, pp. 700–721, 1927. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118

[46] F. Brauer, P. van den Driessche, and J. Wu, Eds., *Mathematical Epidemiology*. Springer, 2008. [Online]. Available: https://link.springer.com/book/10.1007/978-3-540-78911-6

[47] . J. Mendes, D. T. Haydon, E. McIntosh, N. Hanley, and J. E. Halliday, "Socially vs. Privately Optimal Control of Livestock Diseases: A Case for Integration of Epidemiology and Economics," *Frontiers in Veterinary Science*, vol. 7, no. 11, 2020. [Online]. Available: https://doi.org/10.3389/fvets.2020.558409

[48] R. D. Horan, E. P. Fenichel, C. A. Wolf, and B. M. Gramig, "Managing infectious animal disease systems," *Annual Review of Resource Economics*, vol. 2, no. 1, p. 101–124, 2010. [Online]. Available: https://doi.org/10.1146/annurev.resource.012809.103859

[49] M. Ajelli, B. Gonçalves, D. Balcan, V. Colizza, H. Hu, J. J. Ramasco, S. Merler, and A. Vespignani, "Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models," *BMC Infectious Diseases*, vol. 10, 2010. [Online]. Available: https://doi.org/10.1186/1471-2334-10-190

[50] E. P. Fenichel, C. Castillo-Chavez, M. G. Ceddia, G. Chowell, P. A. G. Parra, G. J. Hickling, G. Holloway, R. Horan, B. Morin, C. Perrings, M. Springborn, L. Velazquez, and C. Villalobos, "Adaptive Human Behavior in Epidemiological Models," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 15, pp. 6306–11, 2011. [Online]. Available: https://doi.org/10.1073/pnas.1011250108

[51] E. P. Fenichel, R. D. Horan, and G. J. Hickling, "Management of infectious wildlife diseases: bridging conventional and bioeconomic approaches," *Ecological Applications*, vol. 20, no. 4, pp. 903–914, 2010. [Online]. Available: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/09-0446.1

[52] L. M. Castro and F. Lechthaler, "The contribution of bio-economic assessments to better informed land-use decision making: An overview," *Ecological Engineering*, vol. 174, 1 2022. [Online]. Available: https://doi.org/10.1016/j.ecoleng.2021.106449

[53] A. Basurto, H. Dawid, P. Harting, J. Hepp, and D. Kohlweyer, "Economic and Epidemic Implications of Virus Containment Policies: Insights from Agent-Based Simulations," *Bielefeld Working Papers in Economics and Management*, no. 05, 2020. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3635329

[54] A. Duarte, S. Walker, A. Metry, R. Wong, J. Panovska-Griffiths, and M. Sculpher, "Jointly Modelling Economics and Epidemiology to Support Public Policy Decisions for the COVID-19 Response: A Review of UK Studies," *PharmacoEconomics*, vol. 39, 2021. [Online]. Available: https://doi.org/10.1007/s40273-021-01045-2

[55] D. Haw, C. Morgenstern, G. Forchini, R. Johnson, P. Doohan, P. C. Smith, and K. Hauck, "Data Needs for Integrated Economic-Epidemiological Models of Pandemic Mitigation Policies," *Epidemics*, vol. 41, p. 100644, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1755436522000846

[56] T. J. Philipson and R. A. Posner, *Private Choices and Public Health*. Cambridge, Mass.: Harvard University Press, 1993.

[57] A. P. Galvani, T. C. Reluga, and G. B. Chapman, "Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 13, pp. 5692–5697, 2007. [Online]. Available: https://doi.org/10.1073/pnas.0606774104

[58] K. M. Thompson and N. D. Badizadegan, "Modeling the Transmission of Measles and Rubella to Support Global Management Policy Analyses and Eradication Investment Cases." *Risk analysis : an official publication of the Society for Risk Analysis*, vol. 37, no. 6, pp. 1109–1131, 6 2017. [Online]. Available: https://doi.org/10.1111/risa.12831

[59] S. Barrett and M. Hoel, "Optimal Disease Eradication," *SSRN Electronic Journal*, vol. 12, no. 5, 2004. [Online]. Available: http://dx.doi.org/10.2139/ssrn.525402

[60] R. N. Thompson, C. A. Gilligan, and N. J. Cunniffe, "Control fast or control smart: When should invading pathogens be controlled?" *PLoS Computational Biology*, vol. 14, no. 2, pp. 1–21, 2018. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1006014

[61] A. M. Bate, G. Jones, A. Kleczkowski, A. MacLeod, R. Naylor, J. Timmis, J. Touza, and P. C. White, "Modelling the impact and control of an infectious disease in a plant nursery with infected plant material inputs," *Ecological Modelling*, vol. 334, pp. 27–43, 8 2016. [Online]. Available: https://doi.org/10.1016/j.ecolmodel.2016.04.013

[62] H. An, S. Lee, and S. J. Cho, "The effects of climate change on pine wilt disease in South Korea: Challenges and prospects," *Forests*, vol. 10, no. 6, 2019. [Online]. Available: https://doi.org/10.3390/f10060486

[63] M. F. Macpherson, A. Kleczkowski, J. R. Healey, C. P. Quine, and N. Hanley, "The effects of invasive pests and pathogens on strategies for forest diversification," *Ecological Modelling*, vol. 350, pp. 87–99, 4 2017. [Online]. Available: https://doi.org/10.1016/j.ecolmodel.2017.02.003

[64] C. Petucco and P. Andrés-Domenech, "Land expectation value and optimal rotation age of maritime pine plantations under multiple risks," *Journal of Forest Economics*, vol. 30, pp. 58–70, 1 2018. [Online]. Available: https://www.nowpublishers.com/article/Details/JFE-0372

[65] N. J. Cunniffe, B. Koskella, C. J. E. Metcalf, S. Parnell, T. R. Gottwald, and C. A. Gilligan, "Thirteen challenges in modelling plant diseases," *Epidemics*, vol. 10, pp. 6–10, 3 2015. [Online]. Available: https://doi.org/10.1016/j.epidem.2014.06.002

[66] I. L. Boyd, P. H. Freer-Smith, C. A. Gilligan, and H. C. J. Godfray, "The Consequence of Tree Pests and Diseases for Ecosystem Services," *Science*, vol. 342, no. 6160, 2013. [Online]. Available: https://doi.org/10.1126/science.1235773

[67] D. Bell, A. Comas-Herrera, D. Henderson, S. Jones, E. Lemmon, M. Moro, S. Murphy, D. O'Reilly, and P. Patrignani, "COVID-19 mortality and long-term care: a UK comparison," LTCcovid, Tech. Rep. August, 2020. [Online]. Available: https://ltccovid.org/wp-content/uploads/2020/08/COVID-19-mortality-in-long-term-care-final-Sat-29-1.pdf

[68] J. Appleby, "NHS sickness absence during the covid-19 pandemic," *BMJ*, vol. 372, 2021. [Online]. Available: http://dx.doi.org/10.1136/bmj.n471

[69] C. E. Coltart, D. Wells, E. Sutherland, and A. Fowler, "National cross-sectional survey of 1.14 million NHS staff SARS-CoV-2 serology tests: a comparison of NHS staff with regional community seroconversion rates," *BMJ Open*, vol. 11, no. 7, 2021. [Online]. Available: http://dx.doi.org/10.1136/bmjopen-2021-049703

[70] J. B. Adams and R. M. Walls, "Supporting the Health Care Workforce During the COVID-19 Global Epidemic," *Jama*, vol. 325, no. 15, 2020. [Online]. Available: https://jamanetwork.com/journals/jama/fullarticle/2763136

[71] E. McTaggart, I. Megiddo, and A. Kleczkowski, "The effect of pests and pathogens on forest harvesting regimes:A bioeconomic model," *Ecological Economics*, vol. 209, no. March, p. 107800, 2023. [Online]. Available: https://doi.org/10.1016/j.ecolecon.2023.107800

[72] M. Garbelotto and P. Gonthier, "Biology, epidemiology, and control of heterobasidion species worldwide," *Annual Review of Phytopathology*, vol. 51, pp. 39–59, 8 2013. [Online]. Available: https://doi.org/10.1146/annurev-phyto-082712-102225

[73] Office For National Statistics, "Impact of coronavirus in care homes in England: 26 May to 19 June 2020," Office For National Statistics, Tech. Rep. July, 2020. [Online]. Available:

https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/impactofcoronavirusincarehomesinenglandvivaldi/26mayto19june2020

[74] T. Shi, C. Robertson, and A. Sheikh, "Effectiveness and safety of coronavirus disease 2019 vaccines," *Current Opinion in Pulmonary Medicine*, vol. 29, no. 3, 2023. [Online]. Available: https://doi.org/10.1097/mcp.0000000000000948

[75] S. Islam, T. Islam, and M. R. Islam, "New Coronavirus Variants are Creating More Challenges to Global Healthcare System: A Brief Report on the Current Knowledge," *Clinical Pathology*, vol. 15, 2022. [Online]. Available: https://doi.org/10.1177/2632010x221075584

[76] M. Mohsin and S. Mahmud, "Omicron SARS-CoV-2 variant of concern: A review on its transmissibility, immune evasion, reinfection, and severity," *Medicine (United States)*, vol. 101, no. 19, 2022. [Online]. Available: https://doi.org/10.1097/md.0000000000029165

[77] X. Trujillo, O. Mendoza-Cano, M. Ríos-Silva, M. Huerta, J. Guzmán-Esquivel, V. Benites-Godínez, A. Lugo-Radillo, J. A. Bricio-Barrios, M. I. Cárdenas-Rojas, E. F. Ríos-Bracamontes, V. M. Ortega-Macías, V. Ruiz-Montes de Oca, and E. Murillo-Zamora, "Predictors of Recurrent Laboratory-Confirmed Symptomatic SARS-CoV-2 Infections in a Cohort of Healthcare Workers," *Vaccines*, vol. 11, no. 3, p. 626, 2023. [Online]. Available: https://doi.org/10.3390/vaccines11030626