Developing acute respiratory infection severity indicators for public health surveillance using primary care computerised medical records

William Elson

Reuben College

January 6, 2026

Contents

List of Figures

List of Tables

Statement of authorship

I certify that the contents of this thesis are entirely my own work. Any contributions and collaborations have been explicitly acknowledged within the text. No material presented in this thesis has been submitted, either in whole or in part, for any other degree or qualification at this or any other university.

Acknowledgements

I owe a huge debt of gratitude to the many people who have supported me throughout my DPhil. Although this has often felt like a personal and, at times, solitary journey, in reality I have been sustained by the guidance, encouragement, and kindness of countless individuals at every stage.

To my supervisory team as a whole, I am incredibly grateful for their collective and varied wisdom, which brought an array of perspectives to my work. To each, in alphabetical order: Professor Richard Hobbs, for your clear and strategic thinking and the helpful doses of reality that kept me on track; Dr Jamie Lopez Bernal, for your ability to tune into the "so what" element amid the noise of tables, figures, and data; Dr Roger Morbey, for your pragmatic approach to statistics that helped overcome countless roadblocks; and finally, Professor Simon de Lusignan, whose enthusiasm and commitment to primary care informatics allowed me to explore the many ways in which this work could be impactful.

I must also mention our research group, the Clinical Informatics and Health Outcomes Research Group (CIHORG), all of whom, past and present, have helped me on this journey. If I do not mention you by name, please take no offence, but know that you have been part of this. Within this team, I would like to make a few specific mentions. Anna Forbes, for the ridiculous number of hours you spent working on my systematic review and for your unflappable moral compass. Gavin Jamie and Rashmi Wimalaratna, for the infinite number of variables curated, second checked, approved, re-evaluated, modified, re second checked, re approved, retired, and re retired. The superhuman Merri Leston, with whom my DPhil journey began and was shared, whose unbelievable efficiency afforded her the ability to abandon me long before I made it over the line. We faced many of the same challenges and came through together, stronger for it.

Rachel Byford and her incredibly dedicated and talented colleagues, including Gunjan Jiwnani and Rosalind Goudie, in the data team, who spent countless hours helping define data cuts and then redefining them when I sheepishly requested yet another essential missing variable. An honourable mention must go to Jess Smylie, with whom I shared the most hilarious day of my DPhil (you know what I am talking about).

In the wider Nuffield Department of Primary Care Health Sciences, a huge thank you to Rafael Perera-Salazar for his ability to transform things shaped like pears into things more uniform. Within the UKHSA syndromic surveillance team, I must also thank Alex Elliot, Dan Todkill, and Sue Smith, whose experience and knowledge of managing large health record based surveillance systems have helped shape many aspects of my work, and without whom a Wednesday morning meeting would not be the same.

Personally, a huge thanks to Adrian Loader, who now knows more about CMR-based surveillance of ARIs than he could ever have hoped or wanted to. The informal and formal

chats we had over the years had a hugely positive impact on many aspects of my thesis.

And most importantly, to my wife Cristina, who endured many long weekends shepherding our twins on walks through Kew Gardens, various London museums, and sometimes even the Peak District, all to give me the peace and quiet I needed to battle through to the end of this thesis, a journey she knows only too well. She made these sacrifices so frequently and so willingly. Where she goes, I follow. A final shout out to Ale and Seb for their stylistic advice on plots, figures, and colour palettes.

In loving memory of Dad, who would have been absolutely delighted that I managed to get the term ‘f-orbitals’ into my thesis.

And to my brother James, who would have thoroughly disapproved of this whole endeavour.

Abstract

Introduction: Acute respiratory infection (ARI) surveillance systems generate essential intelligence to help health authorities protect the public from the consequences of epidemic and pandemic-prone pathogens such as influenza. This Doctor of Philosophy thesis explores how data from primary care computerised medical records (CMRs) can be used to strengthen surveillance of ARIs. Specifically, it describes the development and evaluation of timely population-level severity indicators of ARIs derived from primary care CMRs.

Methods: The thesis consists of four pieces of work: (1) defining an algorithm for the identification of episodes of ARI from the CMR; (2) a systematic review to identify possible markers of severe disease relevant to ARIs in primary care; (3) an assessment of the data quality of these severity markers in the primary care CMR; and (4) a retrospective evaluation of these severity markers to determine their suitability for use in prospective public health surveillance of ARIs.

Key findings: The case detection algorithm provided a unified and flexible approach that increased sensitivity for identifying ARIs and established a suitable cohort for assessing severity. The systematic review identified 30 potential severity markers, comprising seven severe outcomes and 23 more timely predictors of severe outcomes. Severe outcomes included death, hospitalisation, intensive care admission, and complications, while predictors included symptoms, signs, investigations, treatments, and healthcare utilisation markers. The data quality of severity markers varied significantly and was heavily affected by the pandemic. Several predictors showed strong potential as timely severity indicators, with some symptoms, signs, and healthcare utilisation markers demonstrating significant associations with severe outcomes.

Conclusions: This thesis demonstrates that primary care CMR data can be used to create timely severity indicators for ARIs. Future work should focus on piloting severity indicators prospectively and in near real time and improving the recording of severity markers. This could provide a pathway to the implementation of reliable and timely severity indicators for routine public health surveillance.

Lay summary

This lay summary was created following feedback from a patient and public involvement (PPI) workshop, the details of which are described in Appendix A5.

What is this research about?

This work aims to help health authorities identify when and where outbreaks of serious lung infections (such as flu, COVID, or pneumonia) are happening across England. This is particularly important during the winter, when many different bugs are circulating, and during pandemics such as COVID.

Why this is important?

By knowing where and when these increases are happening, support can be given to GP practices and hospitals where it is needed. For example, additional staff or hospital beds could be provided, vaccination programmes expanded, or, in extreme situations, public health measures such as lockdowns considered. This will also help people get the treatment they need and can stop the bugs from spreading to other people.

How we are doing this

We use information written by GPs in patients' medical notes to count how many infections are occurring around the country. We can then look at which infections are serious, and monitor how many serious infections are happening in different parts of the country.

What this means for the public

By improving the way we track serious lung infections, health services can respond more quickly when problems arise. This means better planning and more support where it is most needed. In the long run, it also helps protect people from the worst effects of winter infections and prepares the country to deal with future pandemics.

Academic outputs

The following publications, presentations and posters were a result of the work in this DPhil.

PUBLICATIONS

Main publications

1. Elson W, Forbes A, Jamie G, et al. A systematic review of the markers of severity in acute respiratory infections to inform primary care surveillance. *Influenza Other Respir Viruses*. 2025;19(10):e70172.
2. Elson W, Jamie G, Wimalaratna R, et al. Validation of an acute respiratory infection phenotyping algorithm to support computerised medical record-based respiratory sentinel surveillance, England, 2023. *Euro Surveill*. 2024;29(35):2300682.
3. Leston M, Elson W, Watson C, et al. Representativeness, vaccination uptake, and COVID-19 clinical outcomes 2020–2021 in the UK Oxford–Royal College of General Practitioners Research and Surveillance Network. *JMIR Public Health Surveill*. 2022;8(12):e39141.
4. **Letter** Elson W, Zambon M, de Lusignan S. Integrated respiratory surveillance after the COVID-19 pandemic. *Lancet*. 2022;400(10367):1924–1925.

Indirect contributions to relevant work

1. Whitaker HJ ..., Elson W, Tsang RSM, et al. COVID-19 vaccine effectiveness against hospitalisation and death in clinical risk groups during the Delta period. *J Infect*. 2023;87(4):315–327.
2. Gu X ..., Elson W, Watson C, et al. Postpandemic sentinel surveillance of respiratory diseases in the WHO Mosaic Framework. *JMIR Public Health Surveill*. 2024;10:e52047.
3. Jamie G, Elson W, Kar D, et al. Phenotype execution and modelling architecture to support disease surveillance. *JAMIA Open*. 2024;7(2):ooae034.
4. Hoang U ..., Elson W, Agrawal U, et al. Clinical characteristics of RSV in English primary care. *JMIR Res Protoc*. 2025;14:e60669.

ABSTRACTS AND PRESENTATIONS

1. WONCA Conference, 2022 (**presentation**): *The value of representative disease surveillance cohorts for equitable public health.*
2. UKHSA Conference, 2022 (**abstract**): *Advancements in digital maturity at the Oxford–RCGP RSC in response to COVID-19.*
3. UKHSA Conference, 2023 (**abstract**): *Decoding and encoding COVID and flu: Developing an ARI digital phenotype.*
4. CAARI Conference, 2025 (**presentation**): *Surveillance of work absenteeism in acute respiratory infections.*

Chapter 1

Introduction

1.1 HISTORICAL CONTEXT

In the 17th century, recurrent episodes of the plague had a devastating impact on London society, with some outbreaks causing the death of over 15% of the population [1]. During this period, the Company of Parish Clerks in London reported and disseminated a weekly tally of deaths across 130 parishes [2]. These 'Bills of Mortality' served as an early warning system for plague and helped determine areas of the city to be 'shunned and avoided' [3]. The Bills of Mortality represent an early example of the use of routinely collected data for public health surveillance that is broadly consistent with the modern definition (Table 1.1) [4]. They also highlight the longstanding importance of timely reporting of disease severity to inform public health decision-making, which continues to be relevant today.

Table 1.1: World Health Organization definition of public health surveillance

*The World Health Organization (WHO) defines **public health surveillance** as the systematic ongoing collection, collation and analysis of data for public health purposes and the timely dissemination of public health information for assessment and public health response [5].*

As with the plague in earlier centuries, major epidemics and pandemics of influenza have inflicted a substantial burden on the global population, with the 1918 Spanish influenza pandemic estimated to have caused 50 to 100 million deaths globally [6]. In response to the threat from influenza, global public health surveillance systems have adapted and

innovated. The World Health Organization (WHO) Global Influenza Surveillance and Response System (GISRS) was established in 1952 in response to growing concern about the potential impact of pandemic influenza and has evolved into a critical defence against seasonal viruses and viruses of pandemic potential [7]. Most recently, the COVID-19 pandemic has reinforced the ongoing threat of emerging respiratory viruses and highlighted the need for continued technological innovation to mitigate this risk [8].

A relatively recent advance in respiratory virus surveillance has been the increasing use of computerised medical records (CMRs), which have largely replaced paper-based patient notes by digitally recording patients' interactions with health services [9]. Their strengths for surveillance include their large and representative sample sizes, the breadth and richness of the clinical information captured, and the ability to provide timely, near real-time data feeds. During the COVID-19 pandemic, CMRs were recognised as a valuable resource for monitoring disease trends and informing decision-making [10]. Despite these advantages, challenges remain: because CMRs are collected primarily to support clinical care and administrative processes, one cannot assume the quality of the data is sufficient to meet the specific requirements of surveillance.

This Doctor of Philosophy (DPhil) sets out to strengthen surveillance of acute respiratory infections (ARIs) through the development and evaluation of population-level severity indicators derived from primary care CMRs. This can support data-driven public health interventions and ongoing efforts to prepare for seasonal outbreaks and future pandemics. In doing so, it seeks to leverage the advantages of these data while remaining mindful of their inherent limitations. In this introduction, I outline the current landscape of respiratory virus surveillance, describe the challenges of measuring population-level severitys, and summarise how primary care CMRs may offer a potential solution to these challenges.

1.1.1 Surveillance of acute respiratory infections

The aims of ARI surveillance are multifaceted. They include providing early warning of epidemics and pandemics, monitoring geographical spread, assessing disease severity, and

generating evidence to guide public health interventions [11]. Intelligence from respiratory surveillance systems can inform decisions such as the timing of United Kingdom (UK) primary care antiviral prescribing based on influenza activity, intensive care unit (ICU) capacity planning during seasonal respiratory syncytial virus (RSV) surges, and vaccine strain selection through GISRS [12–14]. During the COVID-19 pandemic, surveillance data also guided lockdowns and vaccine roll-out [15, 16].

To achieve these objectives national public health systems typically employ a range of surveillance operations, with national data then feeding into international programmes such as GISRS. Surveillance activities in the UK are coordinated by the United Kingdom Health Security Agency (UKHSA) and include laboratory-based reporting of respiratory pathogens through the Second Generation Surveillance System (SGSS) and the Respiratory DataMart, as well as secondary care surveillance systems that monitor the incidence of pathogen-confirmed disease among patients admitted to sentinel hospitals. In addition to these components, a major contributor to UK ARI surveillance is primary care CMR-based surveillance, which uses data from general practice consultations to monitor respiratory infection trends across England.

1.1.2 Primary care-based ARI surveillance in England

The Oxford–Royal College of General Practitioners Research and Surveillance Centre (RSC) is one of Europe's longest-running primary care sentinel surveillance systems [17]. Established in 1967, the RSC is a network of English primary care practices that provides weekly reports on a number of clinical conditions, including a range of respiratory syndromes that were originally mapped to the clinical terminology International Classification of Diseases (ICD). From its inception through to the onset of the COVID-19 pandemic, the key respiratory indicator was influenza-like illness (ILI), which served as the cornerstone of national primary care influenza surveillance and was used as a proxy for estimating influenza activity in the community .

In 1992, the RSC introduced seasonal virological sampling, enabling linkage between

clinical ILI data and laboratory-confirmed influenza [17]. This advance allowed timely detection and characterisation of influenza strains circulating in primary care populations. The importance of this system was underscored during the 2009 H1N1 swine influenza pandemic, when clinical reporting from participating practices provided early signals of the arrival and spread of the novel virus in the working-age population [18].

Despite previous successes, ILI proved to be an unreliable indicator of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) activity during the COVID-19 pandemic. Several factors likely contributed to this limitation, including major changes in healthcare access and consultation behaviour, the lack of suitable clinical codes for recording COVID-19 presentations in primary care, and the atypical symptom profile of COVID-19 compared with influenza. These challenges highlighted the need to expand surveillance beyond ILI to capture the broader spectrum of ARIs.

In the aftermath of the COVID-19 pandemic, the GISRS has led efforts to transition toward integrated surveillance of respiratory pathogens with epidemic and pandemic potential [19, 20]. In line with this strategy, national surveillance systems have increasingly shifted focus from ILI to the broader clinical syndrome of ARI, which better reflects the diverse presentations caused by multiple respiratory viruses (Table 1.2). Case definitions used in this DPhil can be seen in Appendix A1.1.

Table 1.2: Definition of acute respiratory infection (ARI) for surveillance

*For the purposes of this DPhil, **acute respiratory infection (ARI)** refers to a surveillance syndrome used to capture the full range of respiratory presentations caused by a variety of pathogens. ARI encompasses more specific syndromes such as influenza-like illness (ILI), upper respiratory tract infection (URTI), and lower respiratory tract infection (LRTI).*

Over the decades, the RSC has substantially expanded in both scope and capability. It currently receives CMR data from more than 18 million patients registered across approximately 1,800 primary care practices in England. Patient-level data are transmitted

directly from IT providers and are securely held on the Oxford Clinical Informatics Digital Hub (ORCHID) servers at the University of Oxford. The data are pseudonymised, making the identification of individuals extremely unlikely. They include all key demographic information and date-stamped clinical events such as diagnoses, symptoms, signs, investigations, treatments, and vaccinations. No free-text data are received. Data from member practices are received in near real time, with a typical lag of two to four days. This enables the production of timely surveillance reports (Figure 1.1 and 1.2) [21].

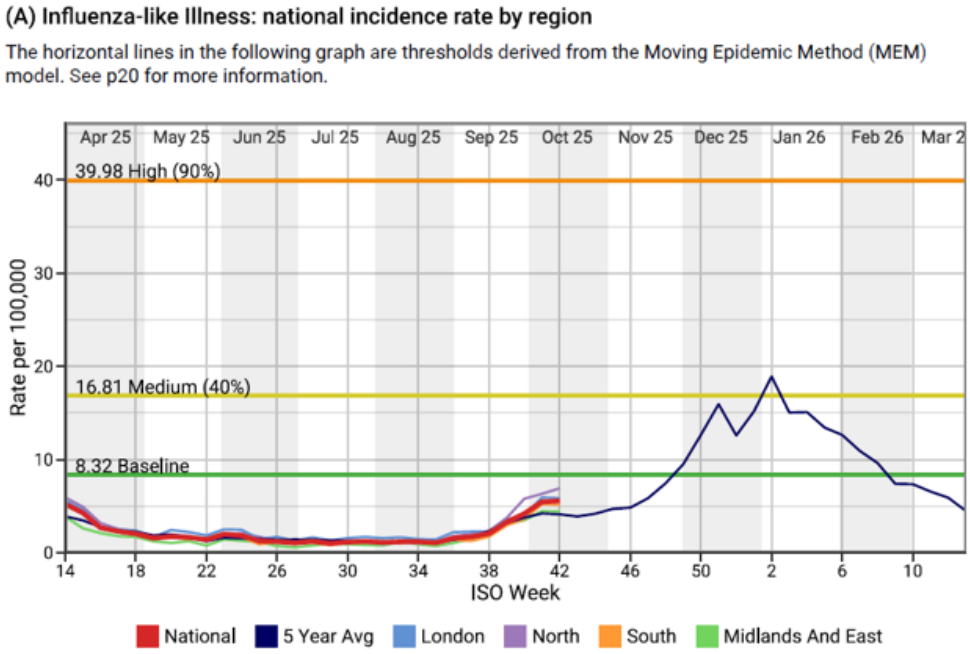


Figure 1.1: Weekly incidence of influenza-like illness (ILI) per 100,000 population, as reported by the Oxford–RCGP Research and Surveillance Centre (RSC) weekly report.

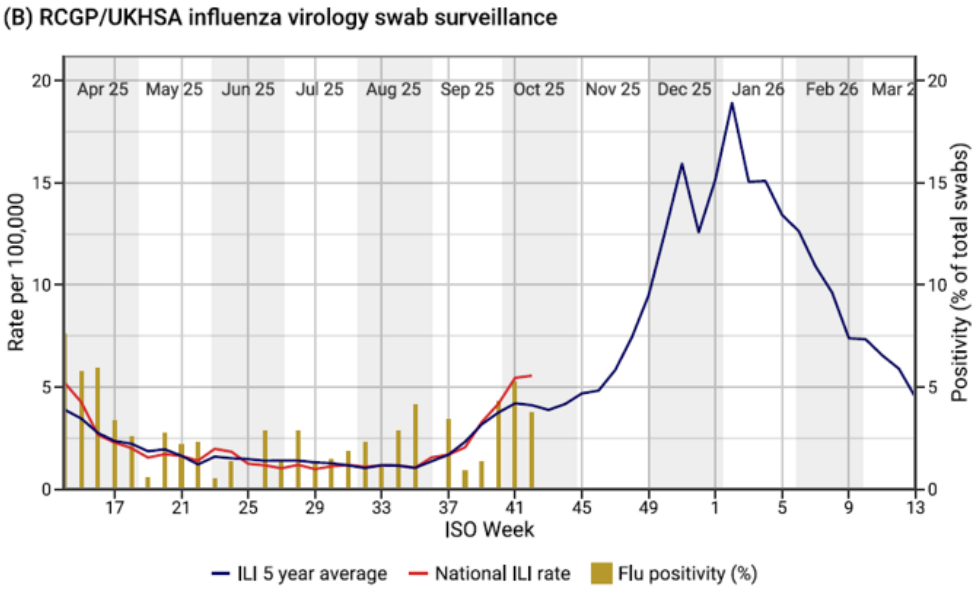


Figure 1.2: Weekly percentage of swabbed samples positive for influenza among patients with influenza-like illness (ILI), from the Oxford–RCGP Research and Surveillance Centre (RSC) weekly report.

Surveillance indicators are the essential quantitative metrics that allow a given surveillance concept to be monitored. For respiratory surveillance, indicators typically include the incidence of ARI and case positivity. Incidence allows tracking of the amount of disease occurring in the surveillance population and case positivity represents the proportion of clinical cases that test positive for a given pathogen, such as influenza or SARS-CoV-2 [22].

Currently, the RSC supplies UKHSA with an ARI-focused weekly report detailing the incidence of several respiratory clinical syndromes, including ILI, and case positivity rates for a number of respiratory viruses, including influenza, SARS-CoV-2, and RSV. Figures 1.1 and 1.2 show examples of ILI incidence and influenza case positivity trends from the RSC weekly surveillance report. Currently the RSC does not report population-level severity.

1.2 POPULATION-LEVEL SEVERITY

The WHO Pandemic Influenza Severity Assessment (PISA) and United States Centers for Disease Control and Prevention (CDC) Pandemic Severity Assessment Framework (PSAF) recognise population-level severity as an essential surveillance indicator for ARIs that can complement incidence estimates [23, 24]. Whereas incidence estimates how much disease is occurring in the community, severity reflects the extent to which a virus is associated with adverse outcomes (Table 1.3). When combined, these measures provide a more complete picture of community threat, as for a single level of incidence the impact of a virus may vary substantially depending on the associated population-level severity. The combination of incidence and population-level severity therefore supports national and regional risk assessment of circulating respiratory pathogens, enabling more targeted and data-driven allocation of resources and public messaging (Figure 1.3) [25, 26].

Table 1.3: Definition of population-level severity

Population-level severity *refers to the seriousness of disease in a population, measuring the extent to which circulating respiratory pathogens are associated with adverse outcomes such as hospitalisation, intensive care admission, or death. In the WHO's PISA framework, this corresponds to the seriousness of disease dimension and complements measures of incidence to estimate overall impact [23].*

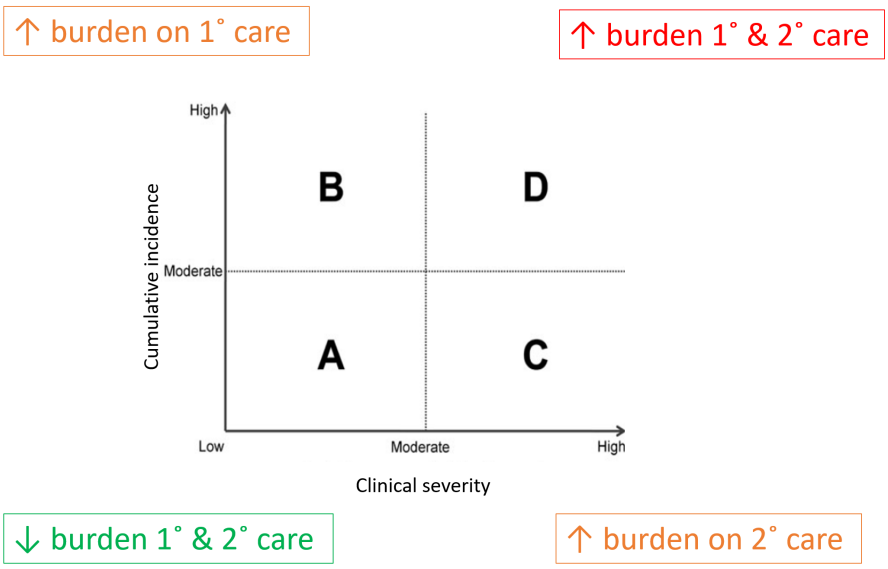


Figure 1.3: Relationship between incidence, clinical severity and its impact on the health care system. X-axis: clinical severity. Y-axis: incidence. Letters A–D reflect different potential scenarios, where D represents high incidence and high clinical severity resulting in the greatest impact on society. The coloured boxes reflect how each scenario could result in differing impacts on the health system. Adapted from Reed et al [24].

In Figure 1.3, four scenarios are illustrated. Scenario A represents low incidence and low severity of disease, where the health service is likely to experience lower levels of strain. In contrast, Scenario D represents the most concerning case, with high incidence and high case severity, which is likely to place significant strain on both primary and secondary care services. Scenarios B and C represent the cases where there is high incidence and low severity and low incidence and high severity, respectively. In reality, the situation is more dynamic, with both incidence and severity varying across a continuum of values.

The WHO and CDC describe the case hospitalisation ratio (CHR) and case fatality ratio (CFR) as key severity indicators [23, 24]. These indicators measure the percentage of cases resulting in hospitalisation or death. A more generic severity indicator can be considered as a simple ratio where the denominator is the number of ARI episodes occurring over a given time period and the numerator represents the number of those episodes with severe disease in that same period. Table 1.4 describes how a generic severity indicator is defined within this DPhil.

Table 1.4: Definitions of severity indicator and severity marker

*A **severity indicator** is a population-level measure that quantifies the proportion of acute respiratory infection (ARI) cases with severe clinical disease. It is typically expressed as a ratio, with the denominator representing the total number of ARI cases and the numerator representing those episodes meeting predefined severity criteria. For a given period p, a generic severity indicator can be written as:*

$$\mathrm{SI}_p = \frac{\text{Number of severe cases in } p}{\text{Total number of cases in } p}$$

In this DPhil, an ***ARI case*** represents an individual with a clinical syndrome consistent with an acute respiratory tract infection, regardless of whether a microbiological diagnosis is known.

*A **severity marker*** is a criterion used to define whether an ARI case is counted as severe and therefore determines the numerator of a severity indicator. In this DPhil severity markers may be classified as:

- **Severe outcomes** are hard outcomes such as hospitalisation or death.
- **Predictors of severe outcomes** are potential early markers of severe disease such as symptoms, signs or treatments. This will be explained in detail in later stages of the DPhil

1.3 THE PROBLEM

Despite the advantages of measuring population-level severity, several challenges limit its accurate and timely reporting, reducing its ability to inform public health interventions during an outbreak [26]. Outcome reporting is often delayed, in part due to the inherent lag between the onset of an ARI and the occurrence of an outcome [27]. This is further compounded by delays in compiling outcome data, such as deaths, and by slow data analysis pipelines [28–30].

Estimates of population-level severity early in an outbreak are often unreliable [31]. This can occur for several reasons. Firstly, at the beginning of an epidemic, small numbers of laboratory-confirmed cases and low death counts can lead to volatility, where minor

fluctuations cause large proportional changes in severity estimates [23]. For this reason, the WHO recommends reporting cumulative CFRs biannually to produce more stable estimates. These are therefore not timely.

Secondly, epidemic phase bias can arise when using outcomes such as hospitalisation or death early in an outbreak. When incidence is rising, severity may be underestimated because many individuals have not yet had sufficient time to experience an outcome. In contrast, when incidence is declining, more individuals will have had the infection for longer, inflating severity estimates [32].

Thirdly, case ascertainment bias can lead to overestimation of severity due to early testing being concentrated among health-seeking populations with clinically more serious disease. During the 2009 swine flu pandemic, such bias was thought to have contributed to initially high estimates of the CFR, which were later shown to be inaccurate [26, 33, 34].

Finally, the absence of accurate baseline estimates of population-level severity makes interpretation of early outbreak data challenging [26].

Table 1.5: Challenges of measuring population-level severity

Inherent lag: Outcomes such as hospitalisation or death occur days or weeks after infection, creating unavoidable delays in estimating severity [27].
Delayed reporting: Administrative processes such as hospital data collation and death registration add further delays to availability of outcome data, especially because reporting is only done on discharge. [26].
Small numbers and volatility: Early in outbreaks, there are low numbers of laboratory-confirmed cases and deaths, so even minor fluctuations can lead to large proportional changes (high volatility) in estimates [23].
Selection bias: Early data are often derived from sicker or hospitalised patients, leading to inflated estimates of severity compared with the general population [33, 34].
Lack of baseline comparators: Without robust seasonal baseline data, it can be difficult to contextualise the severity of novel outbreaks against previous seasons [26].

Continued on next page

Table 1.5: Challenges of measuring population-level severity (continued)

Summary: These challenges contribute to delays in the availability of accurate population-level severity estimates, particularly during the early stages of outbreaks.

1.4 THE POSSIBLE SOLUTION

Using primary care CMRs could provide a more timely and less biased approach to defining severity indicators for ARI. Large primary care datasets, such as those from the RSC, are often representative of the wider population and contain rich information on some severe outcomes and potential predictors of severe outcomes such as symptoms, signs, and treatments that can be used to define severity. They are potentially more timely than hospital-based indicators because patients with ARI typically seek primary health care early in the course of illness, offering an opportunity to identify severe cases sooner. In addition, the large volumes of data available from primary care help to reduce the volatility seen in severity estimates, which is often a problem when using smaller samples or rare outcomes such as deaths or ICU admissions.

Population-level severity indicators derived from primary care CMRs could have several important applications. They could complement existing incidence measures to provide a fuller picture of the impact of circulating respiratory pathogens, helping to identify which seasons or epidemics are associated with more severe disease. Timely severity estimates could support health system preparedness by providing early warning of potential winter pressures and informing the allocation of resources across primary and secondary care. Understanding how unwell people are tending to get when infected can also help with accurate public health messaging. In the context of a pandemic, such indicators could contribute to rapid risk assessment, guide non-pharmaceutical interventions, and support evaluation of the effectiveness of vaccines or treatments.

1.5 KEY DEFINITIONS USED

Here I present the specific definitions of key concepts used in this DPhil. These can also be found in the glossary. Throughout the course of this DPhil these concepts will appear repeatedly.

Acute respiratory infection (ARI)

A surveillance syndrome used to capture the full range of respiratory presentations caused by a variety of pathogens. ARI encompasses more specific syndromes such as influenza-like illness (ILI), upper respiratory tract infection (URTI), lower respiratory tract infection (LRTI) and exacerbations of chronic lung disease (ECLD).

Population-level severity

A measure of the seriousness of disease at the population level, reflecting the extent to which circulating respiratory pathogens are associated with adverse outcomes such as hospitalisation, intensive care admission, or death. In the PISA framework, this corresponds to the seriousness of disease dimension and complements measures of incidence to estimate overall impact.

Surveillance indicator

A quantitative measure derived from surveillance data that describes monitoring of a specific surveillance concept over time. Typically, for ARI surveillance these are incidence, case positivity or population-level severity. Changes in the indicators over time, for example an increase in incidence, can indicate impending public health pressures.

Severity indicator

A type of surveillance indicator that specifically quantifies the proportion of ARI episodes resulting in severe outcomes within a defined period. It is typically expressed as a ratio,

where the denominator represents the total number of ARI episodes and the numerator represents those episodes meeting predefined severity criteria (i.e., severity markers).

Severity marker

Severity markers are the criteria used to decide whether an ARI episode is classified as severe. They include two main types: severe outcomes (e.g., hospitalisation or death) and predictors of severe outcomes (e.g., symptoms, signs, or treatments). Severity markers therefore define the numerator of severity indicators.

Severe outcome

A type of severity marker representing a definitive endpoint of disease, such as hospitalisation or death. Severe outcomes are reliable but are usually less timely because of the lag between symptom onset and the outcomes and delays in the availability of associated data in the primary care CMRs.

Predictor of severe outcome

A type of severity marker representing a characteristic recorded around the time of ARI onset, such as symptoms, signs, or treatments, that may predict subsequent severe outcomes. Predictors are more timely than severe outcomes and can support earlier assessment of severity.

1.6 SUMMARY

The importance of timely indicators of severity during infectious disease outbreaks is as relevant today as it was in 17th-century London. Although technology and methods have advanced, the fundamental need remains unchanged. Delays in reporting and inaccurate early estimates continue to limit the ability of severity measures to prospectively inform public health decision making. Innovations such as the use of primary care CMRs offer an

opportunity to overcome some of these hurdles by improving the accuracy and timeliness of population-level severity reporting for ARIs, though the quality of these data presents a potential limitation that must be carefully considered. This DPhil aims to use primary care CMRs to develop and evaluate population-level severity indicators, thereby enhancing respiratory surveillance and supporting preparedness for emerging epidemic threats.

1.7 AIMS AND OBJECTIVES

The overall aim of this DPhil is to develop and evaluate severity indicators for ARIs, derived from primary care CMRs, in order to strengthen public health surveillance. The work is presented in four main chapters, each addressing a distinct objective.

1.7.1 Chapter 2: Accurate identification of ARI cases

Rationale: In order to develop severity indicators, accurate identification of ARI cases is essential. This forms the basis of the ARI severity indicator denominator. **Aim:** Development and validation of a new ARI case detection algorithm, referred to as the ARI digital phenotyping algorithm. **Methods:** Use of clinical code lists to develop a sharable phenotyping algorithm, with face and internal validation against an existing algorithm.

1.7.2 Chapter 3: Identification of candidate severity markers

Rationale: To construct severity indicators, candidate severity markers must be identified. These severity markers represent the specific criteria used to determine whether a case is severe or not. The number of cases meeting this criterion form the numerator of the severity indicator. **Aim:** To identify severe outcomes or predictors of severe outcomes that could serve as severity markers for ARIs and be used to construct population-level surveillance indicators. **Methods:** A systematic and expert review to identify a list of candidate ARI severity markers.

1.7.3 Chapter 4: Assessment of the data quality of candidate severity markers

Rationale: Primary care CMR data are collected to support patient care; any secondary use of this data should be preceded by an evaluation of data quality to ensure it is fit for the intended secondary purpose. **Aim:** To assess the data quality of candidate severity markers in the primary care CMR. **Methods:** Retrospective cohort study assessing the completeness and temporal stability of severity marker recording.

1.7.4 Chapter 5: Evaluation of severity markers

Rationale: Prior to use in a prospective surveillance system candidate severity markers must be evaluated. **Aim:** To evaluate whether more timely predictors of severity are associated with severe outcomes and could therefore be used in prospective surveillance. **Methods:** A retrospective cohort study assessing the individual-level and weekly aggregate-level association between predictors of severe outcomes and the severe outcomes themselves.

1.7.5 Ethics statement

Ethics approval was obtained from the University of Oxford Central University Research Ethics Committee (CUREC) for all aspects of this work. Ethics Approval Reference: R92694/RE001.

Chapter 2

Identifying cases of acute respiratory infection

2.1 INTRODUCTION

In the context of CMRs, a digital phenotyping algorithm (PhA) is a computer program used to identify individuals with a given characteristic from the raw data [35, 36]. These algorithms have a number of applications including health care quality improvement programs such as the Quality and Outcomes Framework (QoF), observational research and interventional trials [37, 38]. Respiratory surveillance systems need PhAs to identify cases of ARI from the CMR.

For CMR-based ARI surveillance the PhA underpins all principal surveillance objectives. Without it, neither the weekly incidence nor case positivity rates for respiratory viruses can be derived. By extension, it is therefore the foundation of any severity indicators developed during the course of this DPhil. A logical first step in developing severit y indicators is to ensure the PhA on which they are based is fit for purpose; this is the focus of this chapter.

A PhA typically consists of three elements. Firstly, the clinical logic, which defines and documents both the purpose of the algorithm and the reasoning behind its design. Secondly, some mechanism to directly identify cases of interest, typically lists of clinical codes from a given terminology, such as Systematized Nomenclature of Medicine (SNOMED) - Clinical Terms (CT) [39]. Finally, a computer program that implements the logic using these code lists [35]. When run completely, the PhA will retrieve a cohort of the desired cases.

Periodically, PhAs must be reviewed. Over time, surveillance priorities change which can influence clinical logic of a PhA. For example, respiratory surveillance has historically focused on influenza, using ILI incidence as its principal indicator. However, following the COVID-19 pandemic, there has been a shift towards 'integrated surveillance' of a broader range of viruses, including SARS-CoV-2 and RSV [20].

Furthermore, technological advances can also impact on how code lists are defined and implemented. In 2018, for all primary care practices in the UK, there was a mandatory transition from older clinical terminologies (Such as Read) to SNOMED. Additionally, regular updates to the SNOMED hierarchy result in addition and deactivation of codes. Such changes would require a complete re-definition or a comprehensive review of code lists used in a PhA.

In this DPhil I use the infrastructure at the RSC to develop severity indicators. The existing PhA used by the RSC surveillance system, like others, previously focused on ILI, however, during the pandemic it quickly evolved to also support surveillance of SARS-CoV-2 [40]. Nonetheless, an overarching ARI PhA fit for modern integrated surveillance has not been developed.

In this chapter, I describe the development and validation of a new ARI PhA to ensure severity indicators are constructed for a well-defined population of ARI cases. This algorithm will embrace the principle of integrated surveillance and also defines new rule-based SNOMED code lists. In addition to providing the foundation for severity indicator development, the PhA will be integrated into existing RSC workflows and aims to improve the accuracy of ARI surveillance reporting to our partners at UKHSA.

2.2 CHAPTER AIM AND OBJECTIVES

2.2.1 Aim

To develop a new ARI PhA and validate this by comparing it with the existing algorithm used by the RSC.

2.2.2 Objectives

1. **Code list comparison**: Compare the newly developed SNOMED code lists with those from the existing algorithm.

2. **ARI case comparison**: Compare the SNOMED codes that identify ARI cases in the new and existing algorithms, as well as the number of cases identified.

3. **Weekly ARI incidence comparison**: Compare the weekly incidence by age and risk group of ARI computed using the new and existing algorithms.

2.3 METHODS

In the following methods section I start by describing key definitions and some important background concepts. Specifically, I discuss the nature of the working group contributing to the design of the PhA, the difference between intensional and extensional code lists and how this difference relates to the underlying structure of SNOMED. Following this, I briefly describe the existing algorithm and then the new algorithm including its clinical logic and the SNOMED code lists. Finally, I discuss how each objective was addressed.

2.3.1 Key definitions used

In this chapter, I use several technical terms specific to digital phenotyping and clinical terminologies. These are set out in Table 2.1.

Table 2.1: Key definitions used in chapter 4

Term	Definition / Explanation
Digital phenotyping algorithm (DPA)	A computable definition used to identify individuals with a given characteristic from raw primary care computerised medical record (CMR) data. It comprises: (1) clinical logic defining the concept and purpose, (2) clinical code lists representing the concept in the terminology, and (3) executable code implementing these definitions. In this thesis, the DPA identifies cases of ARI from primary care CMRs.
Level	Refers to the hierarchical structure of the ARI DPA's clinical logic: **Level 1**: all ARI cases combined; **Level 2**: major subtypes (URTI, LRTI, ILI, ECLD, ARI–NOS); **Level 3**: specific clinical syndromes (e.g. pharyngitis, bronchitis).
Clinical terminology	A structured vocabulary used to record, classify, and retrieve clinical information in CMRs. In this thesis, the terminology used is SNOMED CT, the mandated coding system in the NHS.
SNOMED CT	A comprehensive polyhierarchical clinical terminology encoding clinical concepts and their logical relationships. Each concept has a unique identifier and may have multiple parents, allowing classification under several hierarchies (for example, upper respiratory tract infection is both an infectious disease and a respiratory condition).
Expression constraint	A logical statement written using the SNOMED Expression Constraint Language (ECL) that defines rules for selecting concepts from the SNOMED hierarchy. Operators such as « (*descendantOrSelfOf*) and `MINUS` (exclude) are used to create dynamic, rule based code lists.

Continued on next page

Table 2.1: Key definitions used in chapter 41

Term	Definition / Explanation
Intensional code list	A code list defined by logical rules or expressions using the SNOMED hierarchy to dynamically include or exclude relevant concepts. For example, it may include all descendants of "upper respiratory infection" except those classified as chronic diseases.
Extensional code list	A code list defined by explicitly enumerating individual SNOMED codes. Unlike intensional lists, extensional lists are static and require manual updating when terminology hierarchies change.

2.3.2 Working group

I developed our new clinical logic and code lists through discussion with clinical informaticians and public health experts actively working in CMR-based surveillance. This group were well placed to inform this process due to their collective experience.

The clinicians included are actively practicing and familiar with the nature of ARI presentations in primary care. They also understand how data is entered into the clinical systems, including the common SNOMED codes in use. As all clinicians had experience in primary care informatics including clinical terminologies they were also trained to effectively use the SNOMED hierarchy.

Public health representatives were individuals currently working at UKHSA in the field of respiratory virus surveillance and included those responsible for interpreting surveillance reports supplied by the RSC. This ensured the clinical logic met UKHSA expectations.

2.3.3 Intensional versus extensional code lists

Code lists used in PhAs can be defined either extensionally or intensionally [41]. In general, an intensional (not to be confused with intentional) definition defines a concept

by rules or logic, whereas, an extensional definition is the thing itself. For example, all odd numbers < 10 would be an intensional definition and $\{1, 3, 5, 7, 9\}$ would the extensional equivalent.

In an extensionally defined code list, individual codes are explicitly enumerated, requiring a review of all possible codes and inclusion or exclusion based on relevance. In contrast, an intensionally defined code list is defined by specifying rules or criteria that leverage the underlying ontological structure of a clinical terminology, such as SNOMED or ICD [41, 42].

The rules used to define an intensional code list depend on the structure of the terminology being used. These rules often take advantage of ontological relationships between codes, particularly the hierarchical structure of a terminology. For example, a rule might specify: *'Return the code for upper respiratory tract infection (URTI) and all its descendant codes'*. In contrast, the extensional version of the same list would require manually selecting the parent code and each descendant individually.

Some evidence suggests that intensional code lists are faster to develop and more accurate [41]. Additionally, they are dynamic and when new codes are added or removed from the hierarchy, they are automatically included or excluded as appropriate. While they require a clear understanding of ontological relationships and computational syntax, intensional code lists are flexible and robust and more resilient to updates in the terminological hierarchy than extensional lists.

2.3.4 SNOMED CT structure

SNOMED is a comprehensive clinical terminology developed by SNOMED International. It is used to encode clinical and administrative health data and is the mandatory clinical terminology used in primary care in the National Health Service (NHS). There are over 360,000 SNOMED codes in the international edition, organised into a polyhierarchy [39].

In a polyhierarchy, codes (except the top-level code) have one or more parents and zero

or more children, unlike standard hierarchies such as ICD, where each code has a single parent. A polyhierarchy better reflects the complexity of healthcare, as concepts can belong to multiple categories. For example, an URTI is both a disease of the respiratory system and an infectious disease.

In SNOMED, the rules used to intensionally define code lists are called expression constraints and the syntax used to represent these rules programmatically is called the Expression Constraint Language (ECL) [43]. To be able to define comprehensive intensional code lists using ECL one must have an understanding of the structure of the polyhierarchy and the types of relationships between concepts.

The most fundamental relationships are hierarchical. These are the direct ancestral relationships between concepts. For example, `descendantOrSelfOf` refers to the concept of URTI and all of its descendants, such as tonsillitis or pharyngitis. More complex hierarchical queries can be defined. For example, all URTIs are included except those concepts that are chronic diseases:

`descendantOrSelfOf URTI` `MINUS descendantOrSelfOf chronic disease`.

This would exclude concepts like chronic conditions not relevant to ARI, such as chronic sinusitis.

In addition to hierarchical relationships, other 'attribute-based' relationships can be referenced when defining intensional code lists [43]. These are not limited to but include the `has active ingredient` attribute that can help define drug code lists. For example, a rule could specify: *'include all drugs that* `has active ingredient` *amoxicillin'*. Other attributes that can also be used to define intensional code lists include `has finding site` or `associated morphology`. Table 2.2 shows some common SNOMED ECL operators used for manipulating the hierarchy.

Table 2.2: SNOMED CT Expression Constraint Language (ECL) operators

Operator / syntax	Description
`descendantOrSelfOf` (`«`)	A code and all its descendants.
`childOrSelfOf` (`«!`)	A code and its immediate children only.
`ancestorOrSelfOf` (`»`)	A code and all its ancestors.
`parentOrSelfOf` (`»!`)	A code and its immediate parents only.
`AND` , `OR`	Logical combination of two sets.
`MINUS`	Excludes a subsequently defined set.

Example:

```
« 54398005 |Acute upper respiratory infection|
MINUS « 3218000 |Mycosis|
```

Where:

- `«` includes the specified code and all its descendants.
- `54398005` is the SNOMED code for "Acute upper respiratory infection".
- `MINUS` excludes concepts matching the next set.
- `« 3218000 |Mycosis|` refers to "Mycosis" and all its descendants.

The RSC defines intensional SNOMED code lists using the hierarchical relationships between concepts and the `MINUS` operator but, at present, only uses attribute-based (`has active ingredient`) relationships for defining drug code lists. Other attribute-based relationships such as `has finding site` or `causative agent` are not available for use in this DPhil due to the current technical limitations of the system.

2.3.5 Existing ARI digital phenotyping algorithm

In the existing weekly surveillance reports, the RSC primarily focused on reporting ILI, consistent with historical public health priorities. While the reports also included the incidence of URTI and lower respiratory tract infection (LRTI), there was no unified

ARI indicator [44]. As a result, three separate PhAs were maintained for ILI, URTI and LRTI. These indicators were created and maintained independently, with no formal documentation of the relationship between code lists. For example, it was unclear whether ILI was a subset of URTI or LRTI, or whether URTI and LRTI shared any common concepts. The code lists used in the old algorithm were originally based on historic ICD classifications but evolved into extensionally defined SNOMED code lists, as SNOMED ECL had not yet been fully implemented by the RSC. The clinical logic underlying the existing algorithm was not formally documented and relied on assumed interpretations based on the indicator names.

2.3.6 New ARI digital phenotyping algorithm

Details of the new phenotyping algorithm including in the appendix of the associated publication [44].

Clinical logic

I defined ARI using a modified version of the 2018 European Union (EU) ARI case definition (Table 2.3) [45]. This was modified to include any symptoms suggestive of an acute respiratory viral infection. This allowed for inclusion of other respiratory infections that may also be caused by relevant viral infections, for example, otitis media and sinusitis. We also aimed to identify only ARI cases that presented to primary care. Therefore I excluded codes that were likely to be recorded from a hospital discharge summary or other non-primary care encounters.

Table 2.3: Comparison of EU and RSC ARI case definitions

Definition source	Criteria

EU ARI case definition	• *Sudden onset of symptoms*
	• *At least one of the following respiratory symptoms:*
	  • Cough
	  • Sore throat
	  • Shortness of breath
	  • Coryza
	• *Clinician's judgement that the illness is due to an infection*

RSC-modified EU ARI definition	• *Sudden onset of symptoms* • *At least one acute respiratory infection symptom* • *Clinician's judgement that the illness is due to an infection*

Note: **EU**: European Union; **RSC**: Oxford–Royal College of General Practitioners Research and Surveillance Centre; **ARI**: acute respiratory infection.

To make the relationship between all the subtypes of ARI clear, I defined a clinical logic that combined all respiratory subtypes into a simple hierarchy (Figure 2.1), where code lists for each indicator at the same level were mutually exclusive. The ARI hierarchy was structured into three levels, with ARI at the top (Level 1). At Level 2, four key indicators are included: ILI, exacerbation of chronic lung disease (ECLD), URTI, and LRTI. Each of these is further subdivided creating 16 Level 3 indicators (Figure 2.1). Since ILI can affect both the upper and lower respiratory tracts, it is included as a separate indicator at Level 2. ECLD, a new indicator, was not part of the previous ARI PhA. Case definitions for Level 2 indicators can be seen in Appendix A1.1.

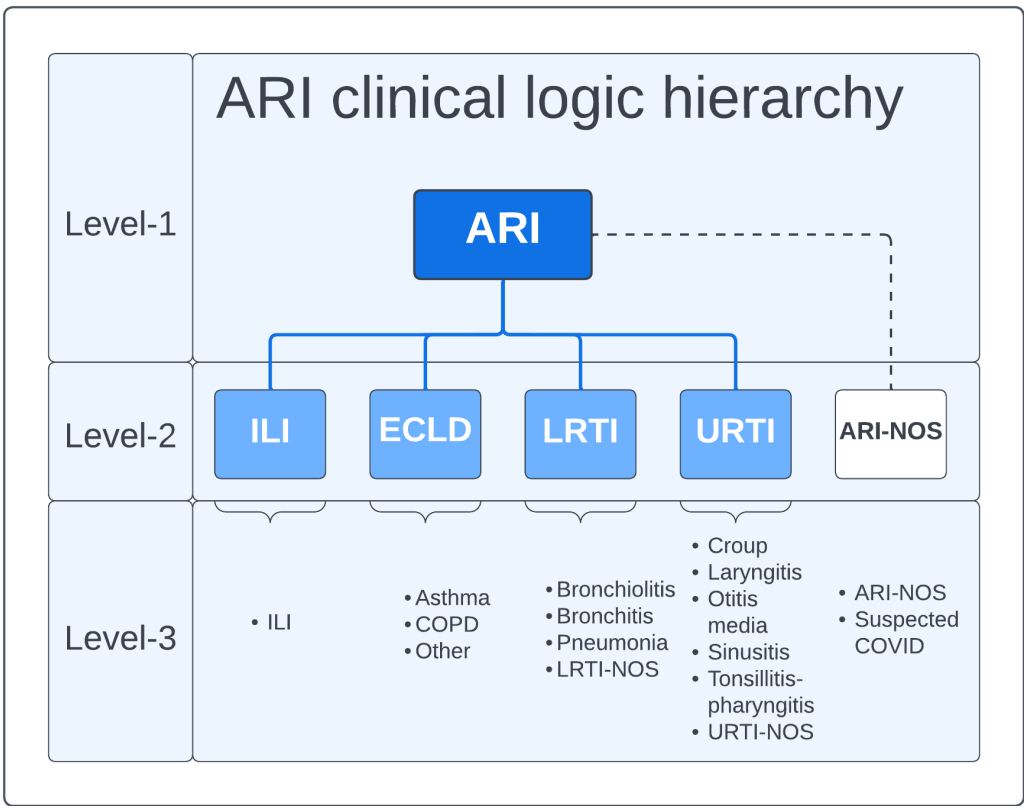


Figure 2.1: ARI phenotype clinical logic: ARI Hierarchy. **ARI**: acute respiratory infection, **ILI**: influenza-like illness, **ECLD**: exacerbation of chronic lung disease, **LRTI**: lower respiratory tract infection, **NOS**: not otherwise specified, **COPD**: chronic obstructive pulmonary disease.

At Level 2, I also included an ARI not otherwise specified (NOS) group. This group encompasses suspected COVID-19 cases and ARI concepts that could not be more specifically classified. The rationale for including suspected COVID-19 in this group was that, following the initial pandemic phase, the use of these specific codes would likely decrease. Clinicians might revert to broader terms like LRTI as confidence in diagnosing COVID-19 as the specific cause of ARI would diminish.

Duplicate case counting is a potential issue in the algorithm for two main reasons. First, a diagnosis might be recorded multiple times by a clinician, such as on consecutive days or through overlapping codes, for example, URTI and tonsillitis. Second, the hierarchical structure could lead to double counting; for instance, an individual could have two Level 2 conditions in a week, but this would still represent a single ARI episode at Level 1.

To address the first issue, we defined new cases of all indicators (at any level) as those recorded at least 28 days after the previous ARI event. This ensured that duplicate codes within the 28-day window were not counted as new cases. To manage the second issue, we restricted individuals to a single instance of any specific indicator. As a result, the total number of Level 1 events does not necessarily equal the sum of Level 2 or Level 3 events.

Code lists

I used SNOMED ECL to define intensional code lists for each of the 16 indicators at Level 3. Code lists could then be combined into Level 2 and the overall Level 1 ARI indicator. I developed code lists using the RSC's in-house 'SNOMED helper tool '. This software allows code list curators to easily navigate the hierarchy and implement ECL rules using a point and click style interface. Code lists were defined by either myself or another clinical informatician. These were then independently checked by a second person, with any discrepancies resolved through discussion, using a third party if required. The software records code list definitions, these are then published on the NHS terminology server. This improves transparency, a known issue when defining code lists [46, 47].

Defining a code list that unambiguously represents a given concept, such as URTI, is challenging as the meaning of a given code may vary depending on the context [48]. For example, the non-specific code 'cough' may be used when reporting an ARI or a chronic cough. Including this code could increase sensitivity of case detection, but would also reduce the specificity of the code list. Striking a balance between sensitivity and specificity is key when building code lists, reinforcing the importance of having a working group with the relevant experience.

In general, only diagnostic codes were included, with relevant symptom codes added where the balance between sensitivity and specificity was deemed acceptable. Codes explicitly mentioning a particular pathogen were considered more likely to arise from microbiological results communicated via hospital discharge summaries, and therefore less likely to reflect the true onset date of a given ARI episode in primary care. Rather than

explicitly excluding pathogen-specific codes from the code lists, the impact of discharge summary-derived codes was minimised at the point of data extraction by excluding ARI cases recorded retrospectively (i.e. where the record entry date was later than the event date). Only ARI cases recorded on the same calendar day as the specified event date were retained, thereby preferentially preserving records most likely to have been generated in real time within primary care, including pathogen-specific codes arising from point-of-care testing, while reducing the inclusion of back-entered discharge summary–related events.

I used a three-step process to define the ECL (Figure 2.2). First, I identified one or, in some cases, several parent codes that best represented the concept of interest, such as, pharyngitis, and included these codes along with all their descendants using the « (`descendantOrSelfOf`) operator. Next, I created a list of generic parent codes and their descendants that represented undesirable codes, such as chronic disease, which were excluded across all 16, Level 3 code lists using the `MINUS` operator. Finally, I made small code list specific modifications that could not be achieved through more generic approaches in the prior two steps.

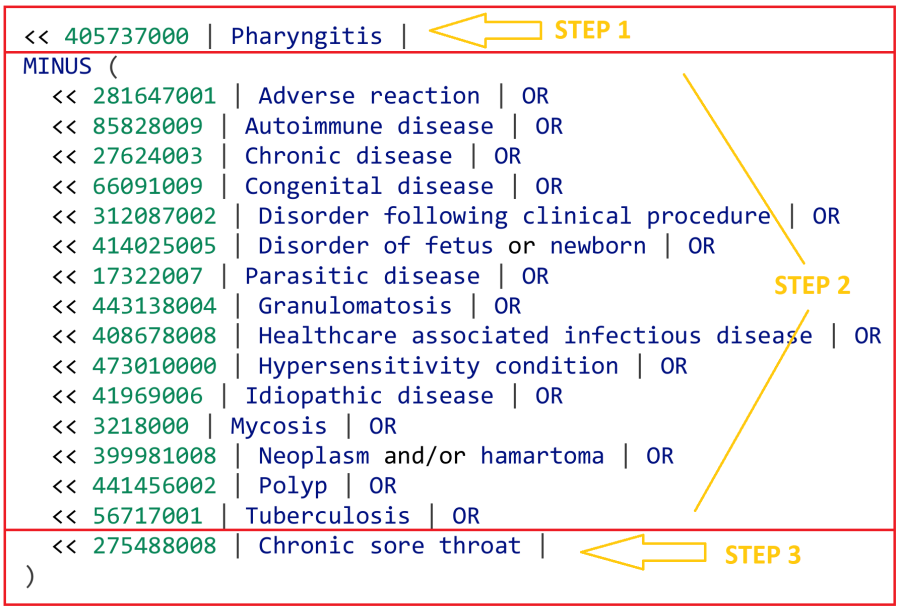


Figure 2.2: Example of Systematized Nomenclature of Medicine Expression Constraint Language for pharyngitis. Most code lists were defined in three steps. Step 1: Include a principal SNOMED parent(s) and descendant codes. Step 2: Exclude higher-level codes to remove irrelevant descendants, such as excluding chronic disease to remove all chronic sinusitis codes. Step 3: Tailor code list by including or excluding specific codes as needed for the specific use case.

Algorithm validation

Objective 1: Code list comparison- I compared all combined codes from the 16 Level 3 intensionally defined code lists with the combination of the old extensional code lists for ILI, LRTI and URTI. To compare these code lists, a set analysis was performed to establish the number of codes present in each code list and the number of intersecting codes.

Objective 2: ARI Case comparison- I used the existing and new PhA to extract ARI cases from the ORCHID for the 2022 to 2023 respiratory surveillance season (International Organization for Standardization (ISO) week 39 2022 to ISO week 38 2023). I then compared codes that identified ARI cases from the CMR using the new algorithm with those codes that identified cases using the existing algorithm. I also counted ARI cases identified by the existing algorithm but no longer identified by the new algorithm and

reported the frequency of codes responsible for these cases. Conversely, I counted ARI cases identified by the new algorithm but not the existing algorithm and codes responsible for these.

Objective 3: ARI Weekly incidence comparison- Finally, to assess the likely impact of the new phenotype on reporting of ARI, I calculated the overall incidence of ARI and the incidence by age band and risk group as cases per 100,000 for the surveillance year 2022/2023. I used 3 age-bands: 0 to 17 years, 18 to 69 years and 70 years and older. Risk groups were defined based on those published in the UK Immunisation Against Infectious Disease Book [49].

An additional algorithm is required to define the incidence of ARI as this must identify both the number of cases but also the denominator population in a given week. Weekly incidence was calculated using an existing algorithm that has two inputs: a list of weekly ARI cases (the output of PhA) and the weekly denominators defined by counting total registered patients in a given week. The list of cases is then converted to a weekly numerator and divided by the weekly denominator and finally multiplied by 100,000. This gives the weekly incidence per 100,000 population.

I calculated the weekly ARI Level 1 and Level 2 indicator incidence using the new and existing algorithm. I compared these and presented time series plots of Level 2 indicators. No comparison was made between the new ECLD or ARI-NOS indicator as no equivalent existed previously. The data extracted for the analysis for objective 3 is not identical to that used for objective 2. The reason for this is that the RSC does not always have reliable denominator data for all practices. Practices with unreliable or absent denominator data were removed from the analysis.

2.4 RESULTS

2.4.1 Objective 1: Code list comparison

The existing ARI code list contained 821 SNOMED codes, compared to the new code list, which included 544, a reduction of 277 codes (33.74%) (Table 2.4). The combined total of unique codes across both lists was 948, of which 417 codes appeared in both lists, 404 were exclusive to the old list, and 127 were exclusive to the new list (Figure 2.3).

Table 2.4: Comparison of existing and new ARI code lists

Level	Codelist	Existing	New	Difference
Level 1	ARI	821	544	−277 (−33.74%)
	URTI	448	206	−242 (−54.02%)
	LRTI	377	243	−134 (−35.54%)
Level 2	ARI-NOS	—	14	14 (N/A)
	ECLD	—	49	49 (N/A)
	ILI	43	49	6 (+6.98%)

Note: **ARI**: acute respiratory infection; **URTI**: upper respiratory tract infection; **LRTI**: lower respiratory tract infection; **ARI-NOS**: ARI not otherwise specified; **ECLD**: exacerbation of chronic lung disease; **ILI**: influenza-like illness.

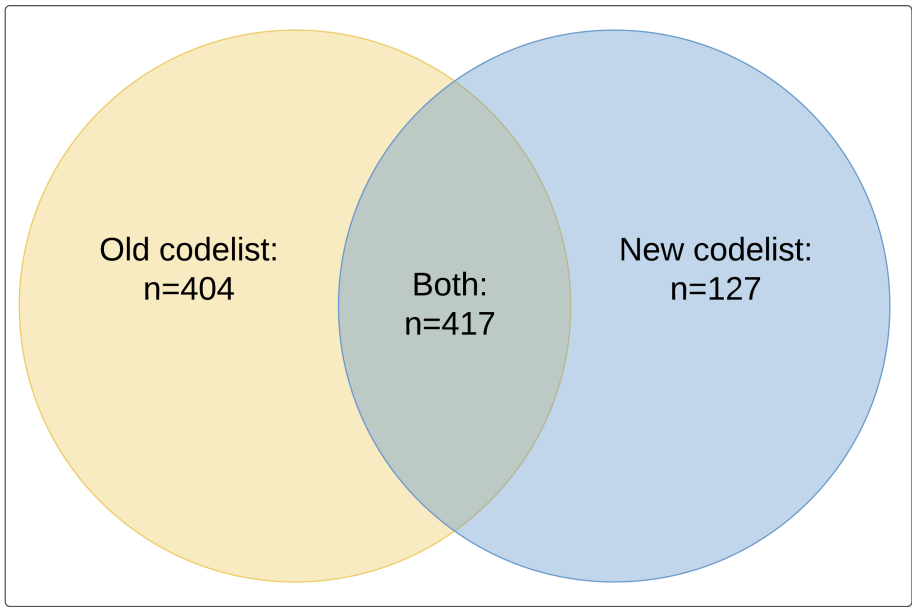


Figure 2.3: Set analysis of new and old Level 1 acute respiratory infection (ARI) code lists. The old ARI code list included 404+417=821 codes and the new code list included 127+417=544 codes. In total, 417 codes appeared in both code lists, 404 only in the old list, and 127 only in the new list.

The 127 codes newly added were distributed across Level 2 in the hierarchy. Because ECLD was newly introduced as a Level 2 subtype, most of the newly added codes were within this group. Although ECLD did not exist as a separate group in the previous code list, four codes that meet the new ECLD definition were already present in the old list but were distributed among URTI or LRTI code lists. These remain in the new ECLD list, resulting in 49 ECLD codes in total, of which 45 are newly added (Table 2.5).

Among the 544 codes in the new code list, 304 (55.88%) were actively used to record ARI cases in the CMR during the study period. Of these 304 codes, 25 (8.22%) accounted for 90.51% of all recorded ARI cases. In the existing code list of 821 codes, 346 (42.14%) were used to record ARI cases. Of these, 16 (4.62%) accounted for 90.55% of all recorded ARI cases, with the remaining 330 codes contributing to the rest.

Table 2.5: Distribution of new ARI codes by Level 2 category

Level 2 codelist	Codes added (n)	Percent
ECLD	45	35.43%
LRTI	42	33.07%
URTI	22	17.32%
ARI-NOS	14	11.02%
ILI	4	3.15%

Note: **ARI**: acute respiratory infection; **ECLD**: exacerbation of chronic lung disease; **LRTI**: lower respiratory tract infection; **URTI**: upper respiratory tract infection; **ARI-NOS**: ARI not otherwise specified; **ILI**: influenza-like illness.

2.4.2 Objective 2: ARI case comparison

I applied the existing and new algorithms to the same patient population to retrieve cases of ARI. The existing algorithm retrieved 2,386,443 cases, while the new algorithm retrieved 3,194,224 cases, an increase of 807,781 cases (33.84%) (Table 2.6). Of the cases identified by the existing algorithm, 52,258 (2.19%) were no longer detected by the new algorithm.

Table 2.6: Number of ARI cases identified by existing and new algorithms

Level	Codelist	Existing	New	% Change
Level 1	ARI	2,386,443	3,194,224	33.84
	URTI	1,647,236	1,862,191	13.05
	LRTI	766,707	987,203	28.73
Level 2	ARI-NOS	—	219,310	—
	ECLD	—	141,482	—
	ILI	47,815	47,812	−0.006

Note: Counts of acute respiratory infection (ARI) cases identified from computerised medical records using the existing and new phenotyping algorithms. Totals at Level 1 are unique individuals and do not equal the sum of Level 2 categories, as individuals may have multiple Level 2 diagnoses. **URTI**: upper respiratory tract infection; **LRTI**: lower respiratory tract infection; **ARI-NOS**: ARI not otherwise specified; **ECLD**: exacerbation of chronic lung disease; **ILI**: influenza-like illness.

Using the new algorithm, ARI cases were most frequently classified as URTI (58.30%), followed by LRTI (30.90%), ARI-NOS (6.87%), ECLD (4.43%), and ILI (1.50%). Compared to the existing algorithm, the new approach identified 28.73% more cases of LRTI, 13.05% more cases of URTI, and a nearly identical number of ILI cases. The three most commonly recorded codes across both algorithms were "Lower respiratory tract infection," "Upper respiratory infection," and "Viral upper respiratory tract infection" (Figure 2.4).

Exclusion reasons varied for the 52,258 cases no longer included by the new algorithm: 43.74% (22,862 cases) represented chronic conditions, 30.04% (15,695 cases) were non-infective conditions, 14.38% (7,515 cases) were recurrent diseases and the remaining cases varied. Figures 2.4 and 2.5 show the codes most frequently associated with cases that are no longer included by the new algorithm (represented by the yellow bars).

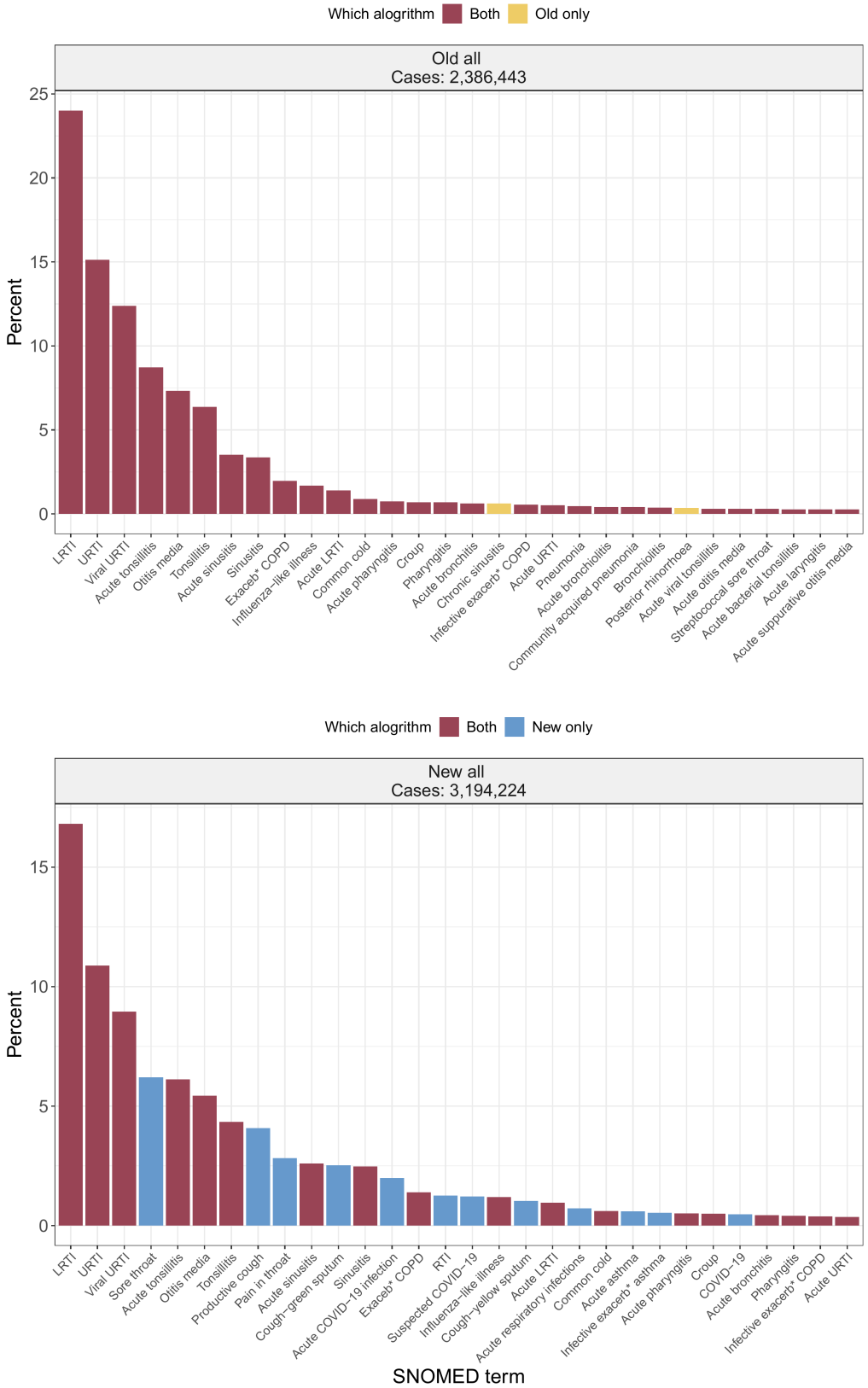


Figure 2.4: SNOMED (Systematized Nomenclature of Medicine) code frequency for acute respiratory infection (ARI) cases by old and new algorithm. **Top:** 30 most frequent codes using the old algorithm. **Bottom:** 30 most frequent codes using the new algorithm. Burgundy bars: cases found by both; yellow: only old; blue: only new algorithm.

Thirteen codes accounted for 91.13% of the 860,039 ARI cases included by the new algorithm but not identified by the old algorithm (Figures 2.4 and 2.5), represented by the blue bars. Of these 860,039 cases, 547,550 (63.67%) were symptomatic codes likely representing an ARI case, 199,299 (23.17%) were ARI-NOS cases, 71,464 (8.31%) were ECLD cases, and 41,726 (4.85%) were attributed to other reasons.

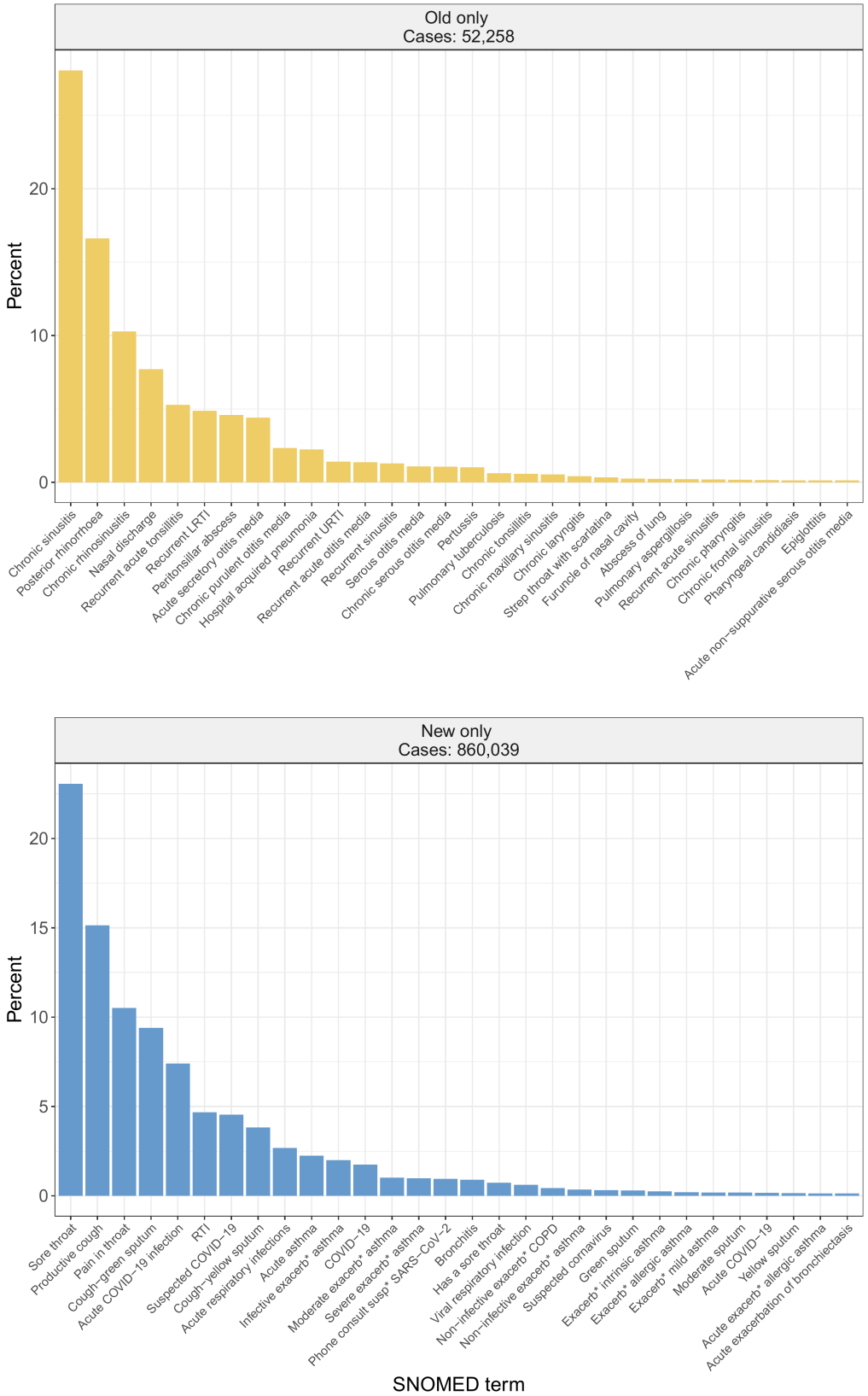


Figure 2.5: SNOMED (Systematized Nomenclature of Medicine) code frequency for acute respiratory infection (ARI) cases uniquely identified by each algorithm. **Top:** 30 most frequent codes found only by the old algorithm. **Bottom:** 30 most frequent codes found only by the new algorithm. Bar colours match those in Figure 2.4.

2.4.3 Objective 3: ARI Weekly incidence comparison

For incidence, practices with unreliable denominator data were excluded, case numbers are therefore less than those in Table 2.6. The existing algorithm identified 1,965,341 cases from practices with reliable denominator data, while the new algorithm retrieved 2,478,473 cases, an increase of 513,132 cases (26.11%) (Table 2.7). The derived median weekly incidence subsequently increased from 205.60 to 258.94 per 100,000 persons, an increase of 25.94 percent. Although the overall incidence increased, the trend over the course of the season was largely comparable (Figure 2.6). Increases in incidence were observed across all age bands and risk group categories for ARI (Level 1), LRTI, and URTI. However, there was no meaningful change in the incidence of ILI across all age bands and risk group categories.

Table 2.7: ARI incidence stratified by risk group and age band, showing percent change in rate (new vs old definitions)

Indicator	Risk group	Age band	Old rate	New rate	% Change (rate)
ARI	Non-risk	0–17	375.95	436.50	16.10%
		18–69	115.15	148.80	29.24%
		70+	101.00	128.05	26.72%
	Risk	0–17	387.90	480.45	23.85%
		18–69	234.85	336.50	43.24%
		70+	254.15	339.55	33.59%
ILI	Non-risk	0–17	1.95	1.95	0.00%
		18–69	3.00	3.00	0.00%
		70+	1.50	1.50	0.00%
	Risk	0–17	2.60	2.60	0.00%
		18–69	4.55	4.55	0.00%
		70+	2.80	2.80	0.00%
LRTI	Non-risk	0–17	54.70	56.55	3.38%
		18–69	30.05	44.80	49.16%
		70+	54.25	76.35	40.77%
	Risk	0–17	74.30	89.30	20.18%
		18–69	116.40	159.35	36.84%
		70+	197.65	232.80	17.79%
URTI	Non-risk	0–17	337.70	381.45	12.95%
		18–69	82.30	98.80	20.02%
		70+	46.60	50.25	7.82%
	Risk	0–17	327.10	369.70	13.03%
		18–69	118.10	137.10	16.09%
		70+	57.30	63.40	10.65%

Note: Incidence are expressed as median weekly cases per 100,000 persons. Percent change reflects relative difference between new and old phenotype definitions within each stratum. **ARI**: acute respiratory infection; **ILI**: influenza-like illness; **LRTI**: lower respiratory tract infection; **URTI**: upper respiratory tract infection. Risk group indicates whether an individual was considered to belong to a clinical risk group, as defined by UKHSA [50]. These groups are operationalised using University of Nottingham, Primary Care Information Services (PRIMIS) business rules, which map SNOMED code lists to specific clinical conditions [51].

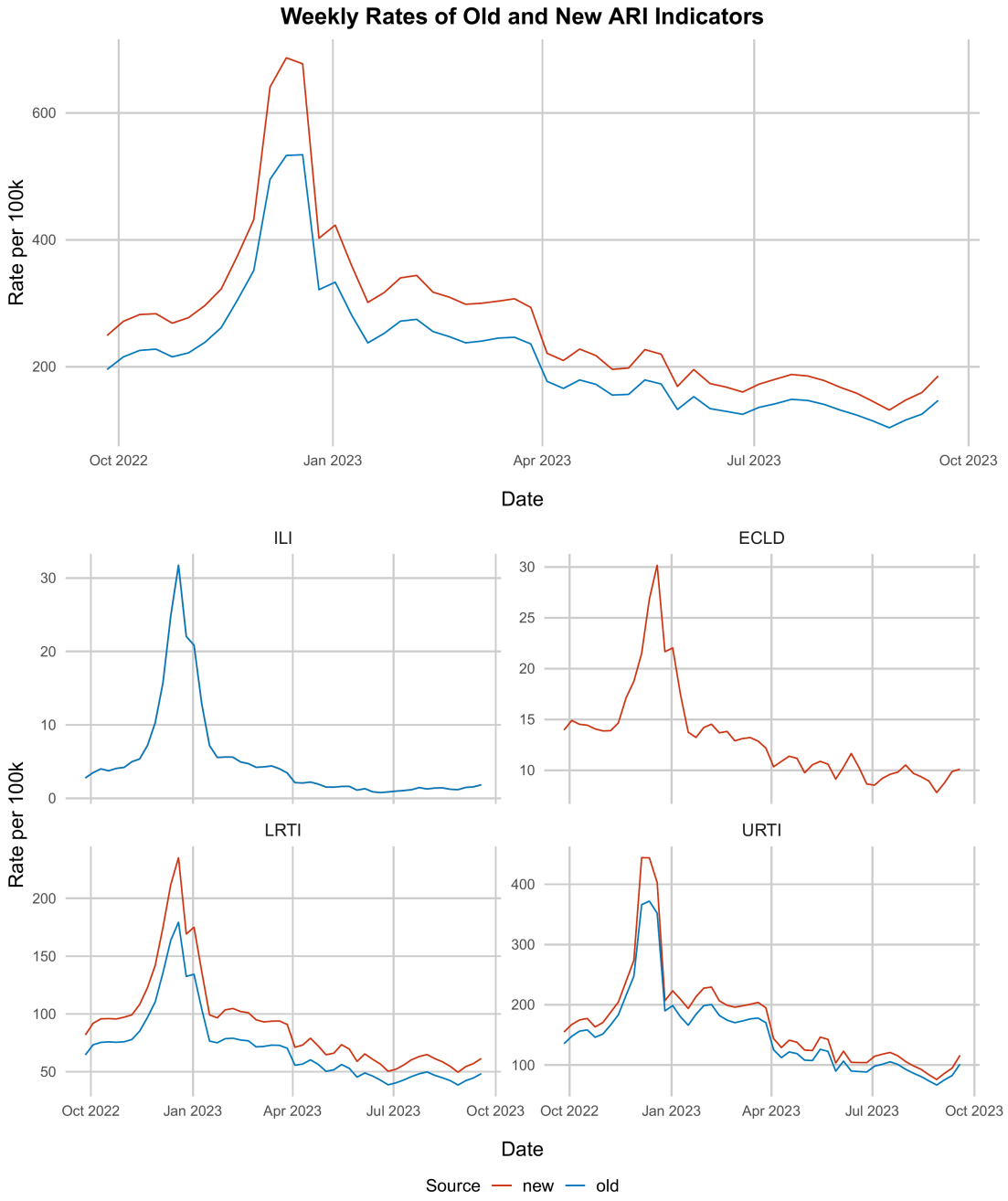


Figure 2.6: Weekly ARI incidence using old and new. Trends of weekly indicator incidence per 100,000 of the population from the old and new indicators. Note, no new indicator is seen for influenza-like illness (ILI) as the incidences are nearly identical. No old indicator is seen for exacerbation of chronic lung disease (ECLD) as there was no old comparator. The y-axes are on different scales to allow adequate comparison of old and new trend lines. **LRTI**: lower respiratory tract infection, **URTI**: upper respiratory tract infection.

2.5 DISCUSSION

2.5.1 Summary of main findings

In this chapter, I developed a new PhA to facilitate accurate detection of cases of ARI from the primary care CMR. The algorithm incorporated a hierarchical clinical logic that accounted for the diverse presentations of ARI, aligning with contemporary integrated surveillance practices. The use of ECL to query SNOMED's polyhierarchy enabled the creation of dynamic code lists and therefore a sustainable algorithm. Face and internal validation demonstrated that the new algorithm increases ARI detection and improves its accuracy. Work in this chapter supports the overall aim of this DPhil by ensuring I develop severity indicators for an appropriate cohort of individuals.

2.5.2 Findings in context

Integrated surveillance

Integrated respiratory surveillance refers to the surveillance of non-influenza respiratory viruses with epidemic and pandemic potential [11]. The trend toward integration has been accelerated by the emergence of new pathogens including severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1) in 2003, Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 and, most dramatically, by SARS-CoV-2 in 2019 [52, 53]. Clinical syndromes used by CMR-based surveillance should be broad enough to account for the varied clinical presentations of these new and evolving pathogens [20, 54].

The hierarchical clinical logic developed here ensures this is the case, through inclusion of a range of upper and lower respiratory tract syndromes. The heterogeneous presentations of novel pathogens would likely be identified through one of the indicators included in the hierarchy. Whilst, atypical presentations of emerging pathogens are documented, for example, anosmia or ageusia with SARS-CoV-2 infection, the PhA is flexible enough to allow rapid addition of new indicators based on relevant evidence [55].

The ability to adapt to novel pathogens is a desirable feature of the PhA. However, 'traditional' pathogens continue to be highly relevant today. Influenza remains a pathogen with epidemic and pandemic potential, and recent bird flu outbreaks and the current emergence of a drifted A(H3N2) strain reinforce the need for continued robust influenza surveillance [56, 57]. The new PhA includes an ILI indicator within the hierarchy that is comparable to previous definitions, thus retaining the ability to analyse historic and future ILI trends. Level 3 indicator, bronchiolitis, is included as it is commonly associated with RSV infection and surveillance of this pathogen will become increasingly important with the roll out of the RSV vaccination programme for pregnant women [58].

Despite the importance of adapting to emerging pathogens, the new algorithm, unlike some other PhAs, intentionally excludes diagnostic codes representing confirmed infections, such as SARS-CoV-2 [59]. Instead, it is designed to detect cases of ARI presenting to primary care with typical clinical syndromes, such as URTI and LRTI. These syndromes may be caused by a range of pathogens. As part of the RSC's standard workflow a subset of practices collect microbiological swabs from patients presenting with an ARI. These samples allow estimation of case positivity rates for specific viruses by ARI subtype, for example ILI. Including virological diagnostic codes in this algorithm could augment RSC virological sampling results but is not appropriate for the initial syndromic case detection.

Use of intensional code lists

Several other studies have reported validation of PhAs for ARIs. A systematic review undertaken by the United States (US) Food and Drug Administration (FDA) in 2021 reported 11 studies that mainly assessed the accuracy of using claims data to define cohorts of patients with ARIs, principally pneumonia [60]. The algorithms reported used the ICD-9 terminology, which unlike SNOMED, is based on a monohierarchy. This limits developers to the use of extensional or very basic intensional code lists. Use of SNOMED ECL has been shown to facilitate development of easily understandable and shareable code lists and is an advantage of this study [61, 62].

Here I took a step wise approach, with rules defining the inclusion of relevant parent codes (and descendants) and exclusion of broad categories, such as chronic diseases. Extensional code lists (a simple enumerated list) on the other hand are not transparent, as examination of the list gives limited clues as to the logic used to define them. Intensional lists are also more resistant to errors caused by the updates to the SNOMED hierarchy. This is because when a new code is added (or removed) from the hierarchy it is automatically included (or excluded) due to inclusion of a parent or other ancestral code. Extensional lists do not have the same capability.

Despite the advantages, the evolving nature of the SNOMED polyhierarchy still presents challenges. For instance, in the tonsillitis/pharyngitis ECL, despite excluding all chronic diseases generically, 'Chronic sore throat' still required manual exclusion, as it was not classified as a descendant of 'Chronic disease' by SNOMED International (Figure 2.2). Furthermore, when new codes are automatically incorporated or discarded from the hierarchy this can still impact on the output of a PhA, therefore ongoing reviews and maintenance of intensional code lists are still required.

Algorithm validation

The updated algorithm significantly increased ARI case detection, predominantly through newly included symptom codes, codes for ECLD and ARI-NOS. In addition to increasing case detection, the algorithm also reduced the identification of inappropriate codes. Most cases excluded by the new algorithm were those attributable to chronic or recurrent disease. These additions likely improve sensitivity but at the cost of reduced specificity, whereas exclusion likely enhanced specificity while slightly lowering sensitivity.

Decisions about whether to include or exclude certain codes were made by the working group. The collective clinical, informatics and public health experience of the working group served to ensure decisions were well considered and collaborative. These decisions were made on the basis that any specific change would, on balance, have a favourable impact overall on the accuracy of the PhA. That is, the benefit of increased sensitivity

outweighed the reduced specificity or vice versa.

Increases in incidence were observed for URTI and LRTI whilst ILI incidence remained almost identical. This reflects the fact that the code list for ILI changed very little compared to the others. This would facilitate continued comparison of current and historical ILI incidence which is favourable for long-term trend analysis. The likely effect of these changes is an improvement in the system's ability to detect low-levels of disease, particularly if rates are aggregated by age band and region.

2.5.3 Translation in practice

The work undertaken in this chapter has been subject to peer-review in a well-respected public health journal [44]. Additionally, the algorithm has now been adopted by the RSC and is used formally to derive weekly incidence figures that are included in reports sent to UKHSA [63]. Extracts of the reports are integrated into the wider UKHSA weekly respiratory surveillance reports [64]. The RSC also supplies weekly ARI and ILI incidence to European Centre for Disease Prevention and Control (ECDC)'s European Respiratory Virus Surveillance Summary (ERVISS). These figures are jointly reported by both ECDC and WHO.

The code lists and logic used in this PhA are available via both the supplementary data of the associated publication and the NHS terminology server where they have been published and are free to access [44, 65]. This approach supports the principles of open science by providing a framework for others working in CMR-based ARI surveillance. Although it is principally of value to those utilising SNOMED, those using other clinical terminologies could derive value in the clinical logic and generic processes that have been applied.

2.5.4 Limitations

The key limitation of this work is the absence of a clear reference standard for defining true ARI cases. For example, when a primary care practitioner records a case of ILI we

have no way of confirming the individual met the case definition criteria. As a result, it is very challenging to measure sensitivity, specificity and accuracy of the algorithm. The best we can do is infer that sensitivity, for example, is likely to have increased.

An ideal validation process would measure the accuracy of coding by clinicians and could be best achieved through prospective studies with primary care practices, capturing detailed symptom data to provide a reference standard. Such approaches have been taken previously and typically involve a medical record review [60]. However these are mostly single centre studies which would allow direct access to detailed records. A review of the RSC CMR associated with each case is feasible in a subset of cases of ARI, but as no free-text data is included, this would be incomplete. This approach is beyond the scope of this DPhil. Given these constraints, the method used here, while not perfect, represents a pragmatic and reasonable compromise.

The exclusion of free-text analysis, which has been implemented successfully in other settings [66], impacts the algorithms' ability to identify relevant ARI cases, thus reducing its sensitivity. At present free text data is not supplied to the RSC, the principal reason for this is the high governance bar to receiving free text data which is potentially more sensitive due to possible inclusion of personal details [48]. Innovations in the use of natural language processing (NLP) algorithms could support not only more accurate coding in primary care, but could also be used to augment terminology-based PhAs.

Primary care CMRs are primarily designed to support patient care rather than for secondary purposes such as surveillance [67]. However, primary care practices are familiar with the importance of high quality coding for secondary purposes, for example, for QoF financial reimbursements. Educational materials for primary care practices that contribute to CMR-based surveillance systems should be provided to ensure coding of ARI cases is optimal. The RSC has a range of materials and a team dedicated to support surveillance practices [44].

2.5.5 Summary

This work has fulfilled its principal objective by supporting accurate and robust detection of ARI cases from the CMR. The cases extracted through application of the PhA will be counted over a given time period (for example, weekly) to derive the denominator of the severity indicators developed in the later stages of this DPhil (Table 1.4). I have also demonstrated the national and international reach of this work through its contribution to surveillance reporting. In the next chapter I shift the focus from defining denominators to defining numerators, and undertake a systematic review to identify candidate severity markers.

Chapter 3

Identification of candidate ARI severity markers

3.1 INTRODUCTION

A severity indicator is a ratio of the number of severe cases (numerator) to the total number of ARI cases (denominator) over a specified period (Table 1.4) [23]. In the previous chapter, I described the development of a PhA to accurately identify ARI cases from the CMR, forming the basis of the denominator. In this chapter, I turn my attention to the numerator.

In order to count severe cases we must define some criteria for what 'severe' means. I refer to these criteria as severity markers (Table 1.4). Typical severity markers include hospitalisation and death. These define the CHR and CFR, which are commonly used severity indicators to assess the impact of an influenza season, epidemic or pandemic [23, 24]. These metrics are typically measured in secondary care-based surveillance [68–70].

Primary care CMR-based severity indicators, as described in chapter 1, could have some advantages over those used in secondary care, including representativeness and timeliness. These data are also rich with additional information such as symptoms, signs, investigations, treatments and other outcomes. While hospitalisation and death data are available and could be used to define CHR and CFR, other parameters recorded at the point of care, such as clinical signs, offer an opportunity to define more timely severity indicators.

In order to test this hypothesis, I needed to identify candidate severity markers relevant in the context of primary care CMR-based surveillance. Factors that determine whether

a given severity marker may be valuable for consideration include: the extent to which it may predict or represent a severe outcome from a clinical perspective; whether the severity marker can be attributed directly to a case of ARI and the likelihood that it is used and recorded in the primary care CMR.

In this chapter, I present the findings of a systematic review aimed at identifying a broad set of candidate ARI severity markers. These are categorised as either severe outcomes, such as complications, hospitalisation or death, or potential predictors of severe outcomes, where predictors are likely to be more timely.

3.2 CHAPTER AIM AND OBJECTIVES

3.2.1 Aim

To identify severe outcomes or predictors of severe outcomes that could serve as severity markers for ARIs and therefore be used to construct severity indicators.

3.2.2 Objectives

1. **Define candidate severity markers**: Identify a list of candidate severity markers for primary care CMR-based ARI surveillance.
2. **Curate severity marker code lists**: Curate a set of SNOMED code lists that can be used to represent these severity markers and facilitate extraction from the CMR.

3.3 METHODS

3.3.1 Overview

This systematic review was registered with PROSPERO (the International Prospective Register of Systematic Reviews: registration number CRD42023460281) and its protocol made publicly available since August 2023 [71]. This was a descriptive, mapping-style systematic review collating and summarising the severity markers reported in studies of

ARI. It did not evaluate intervention effects, compare groups, or perform quantitative pooling. Consequently, no meta-analysis or formal risk-of-bias appraisal was undertaken.

The systematic review was conducted to identify an initial list of candidate severity markers by screening biomedical databases and grey literature for studies and surveillance reports on ARIs. A panel of primary care informatics experts, including myself, then reviewed the candidate list to identify those most suitable for use in primary care CMR-based surveillance. Finally, for each candidate severity marker an intensional SNOMED code list was defined to facilitate retrieval of data from the CMR.

3.3.2 Working group

The working group for this piece of work was led by myself (WE). The systematic review screening, study selection and data collection were performed by myself and a second reviewer (AF) with support from a third reviewer (GJ) where conflicts couldn't be resolved. Both WE and AF are practicing clinicians, I work in primary care and AF in secondary care. This broader experience helped during the screening and study selection process as studies were from a range of settings.

The panel of primary care informatics experts included four individuals (WE, GJ, RW and SdL). Not only were all four practicing primary care clinicians but all have significant experience in primary care informatics and CMR-based ARI surveillance. This included a professor in clinical informatics (SdL) and the author of a beginners guide to SNOMED CT [72] (GJ). These four individuals were well placed to judge the suitability of candidate severity markers for use in CMR-based ARI surveillance.

3.3.3 Eligibility criteria

The systematic review included English-language, cross-sectional or cohort studies published after 1st January 2009 that reported potential severity markers in cohorts of greater than 500 individuals diagnosed with ARI, ILI, severe acute respiratory infection (SARI) or

suspected COVID-19 (Table 3.1). Because multiple definitions exist for these syndromes, each study's own case definition was accepted when determining eligibility. Studies from 2009 were selected to ensure inclusion of studies resulting from the swine flu pandemic of the same year.

This DPhil focuses on primary care, however, studies were included from any setting (community, primary care, hospital or ICU) for two main reasons. Firstly, outcomes measured in other settings could also be recorded in primary care. Secondly, candidate variables were subsequently reviewed by primary care informaticians specifically to assess their potential value for use in primary care CMR-based ARI surveillance.

To determine whether a reported outcome or characteristic would serve as a severity marker reviewers considered the following question:

> *"On balance, might the presence (or absence) of this characteristic in an individual with ARI predict a severe outcome or represent a severe outcome?"*

Table 3.1: Eligibility criteria for included studies

Criteria Category	Details
INCLUSION	
Population	Studies involving individuals with respiratory clinical syndromes including: influenza-like illness (ILI), acute respiratory infection (ARI), severe acute respiratory infection (SARI) or suspected COVID-19. Sample size must exceed 500 participants. Any care setting is eligible (community, primary care, secondary care, intensive care unit (ICU)).
Severity markers	Study must report an outcome that could serve as a marker of clinical severity, including symptoms, signs, complications, treatment, hospitalisation, or death.
Data source	Analysis based on surveillance data or computerised medical record (CMR) data.
Study design	Cross-sectional or cohort studies (prospective or retrospective).
Time period	Articles published on or after 1 January 2009.
EXCLUSION	
Population	Neonatal-only or pregnancy-only cohorts, or populations restricted to specific comorbid groups (e.g. human immunodeficiency virus-positive).
Study design	Randomised controlled trials, case–control studies, reviews or systematic reviews, methodological papers without new data, case reports, or case series.
Article type	Conference proceedings, abstracts, preprints, or letters to the editor.

3.3.4 Information sources

I searched bibliographic databases on the 5th of June 2023 via Ovid: MEDLINE, Embase and Global Health. In each database, free text keywords were paired with controlled vocabulary: Medical Subject Headings (MeSH) in MEDLINE, Emtree in Embase, and Centre for Agriculture and Bioscience International (CABI) thesaurus terms in Global

Health.

In addition, I searched the grey literature using websites from five public health bodies: CDC; the WHO Disease Outbreak News; UKHSA; ECDC; and the Sentiworld sentinel-network repository, screening all retrieved titles and bulletins for relevant human-respiratory surveillance content [73]. Addition of these grey literature searches was an amendment to the original review protocol in order to extend the scope of the search to cover relevant public health resources, as such, these searches were conducted on 15th March 2024.

3.3.5 Search strategy

The search strategy was constructed around three concepts:

1. **Public health surveillance**: This limited results to surveillance-focused records through a combination of surveillance subject headings (e.g., Public Health Surveillance, Sentinel Surveillance) and free-text terms ILI, ARI and SARI.

2. **Respiratory infectious disease**: This combined controlled headings and free-text for influenza and for SARS-CoV-2/COVID-19. The SARS-CoV-2 block was limited to studies occurring after start of the COVID-19 pandemic, ($\geq$2020). Peer-reviewed search terms for SARS-CoV-2 were used.

3. **Clinical severity**: This combined subject-heading terms such as Hospitalisation, Intensive Care Units, Mortality, and Severity of Illness Index with free-text expressions for severity markers. For example, "hospitali?ation rat*," "fatality rat*," "clinical* sever*,". Where * represents any number of characters and ? represents any one character.

Within each concept, search terms and free-text were combined using the OR operator. The AND operator was then used to combine the three main concepts. Further filters

limited the studies to those published in English after 2009 and those involving only human participants. Full search strategies can be seen in Appendix A2.

3.3.6 Study selection

Records were initially imported into Endnote where the automatic deduplication tool was employed [74]. Subsequently, records were uploaded to Rayyan where further deduplication was performed using the inbuilt deduplication tool and a final manual deduplication was done after title and abstract screening [75]. Rayyan then served as the workspace for both the title and abstract screening and the subsequent full-text review.

The titles and abstracts of articles identified during the search process were screened in duplicate by WE and AF and conflicting judgments resolved by discussion between themselves and if necessary GJ. Articles that passed screening went on to a full text review by WE and AF. Duplicate screening of the full text articles was also undertaken.

To reduce subjectivity of severity markers identification and increase methodological transparency WE and AF assessed independently whether an outcome or characteristic reported in a study could serve as a severity marker according to the question in Section 3.3.3.

This question was answered independently for each study and discrepancies were reconciled through discussion with GJ if needed. This process ensured that the identification of severity markers was as transparent as possible.

3.3.7 Data-collection process

A bespoke electronic data-entry form was built in JotForm (an online platform for creating custom data collection forms) [76]. JotForm includes a variety of tools to support accurate data entry, including date selection tools, picking lists, skip logic and validity checks. The form was piloted on ten studies by WE and AF and revised as needed.

Two reviewers (WE, AF) then extracted data independently in JotForm 20 studies at a time.

This allowed amendments of picking lists in the data entry form if new severity markers were identified. Submissions were compiled into a cloud spreadsheet on the JotForm server labelled by reviewer; discrepancies were resolved through discussion between WE and AF, and with GJ if required. The reconciled dataset was exported from JotForm as a CSV file. Study numbers are reported in a PRISMA flow diagram.

3.3.8 Data items

Three groups of variables were extracted from each study (Appendix A2.4).

Study details: citation details, study aim, publication type and date, study period, and geographic scope.

Case data: Respiratory syndrome under study (ILI, ARI, SARI, suspected COVID-19, or other), the case definition applied, the number of cases under study, and whether cases were non-treatment-seeking, treatment-seeking, hospitalised, or in ICU. We grouped participants recruited at a medical facility but not admitted under the label "treatment-seeking" as, in many regions, the distinction between outpatient clinics, emergency departments (EDs), and primary care services is not clear. Non-treatment seeking patients were those who had not attended health care services; typically those identified in the community through surveys.

Severity markers: Severity markers were classified as either (1) a severe outcome itself, such as death, a major complication or hospitalisation (2) a potential predictor of a severe outcome, for example, clinical signs or laboratory findings.

3.3.9 Data synthesis

Extracted outcomes were first tabulated for every included study and grouped into broad groups (e.g., symptoms, vital signs, investigations, complications, treatment, hospitalisation, ICU admission, death). I then produced counts and percentages of studies reporting each group and specific severity marker. For example, the number of studies reporting

any symptom (group) and dyspnoea (severity marker within symptom group). These were also reported by recruitment type (community/primary care, hospitalised, ICU). These data are presented as frequency tables and a frequency chart.

3.3.10 Primary care informatics review

The resulting list of candidate severity markers was subsequently reviewed in two focused meetings by four primary care clinical informaticians as described earlier (WE, GJ, RW, SdL). For each marker the panel reached consensus after discussion, using four criteria:

Severity: The marker's ability to indicate ARI severity. For example, was it felt the presence or absence of the symptom in the context of primary care likely represented a more severe case. For binary variables, such as the presence of a symptom, does the recording of this symptoms indicate the case was more severe. For numeric variables, such as pulse rate, can certain levels of the value indicate a more severe case.

Specificity: The plausibility of linking the marker to the ARI case under surveillance. As there is nothing that explicitly links two or more codes in the RSC CMR data, it is not possible to definitively state that a recorded severity marker is associated with a recorded case of ARI. However, the nature of the severity marker may increase the likelihood that two codes are linked. For example, it is reasonable to assume that a respiratory antibiotic prescribed around a diagnosis of ARI is associated with the case.

The temporal proximity of two codes may also suggest a severity marker and ARI case are linked, this can be defined during the data extract by only including severity markers occurring within a given time from the recording of the ARI case. Therefore, it need not be considered at this stage.

Relevance: Whether primary care clinicians would routinely use the severity marker to characterise an ARI episode. This includes whether they would be contemporaneously recorded during a clinical encounter for ARI or whether they may be recorded in the CMR through administrative mechanisms. For example, an encounter with NHS 111, the free

to call patient advice line, is likely to be recorded through administrative mechanisms.

Recording: The likelihood that the marker is captured in structured CMR data, via the clinical terminology SNOMED. The RSC doesn't extract free text data from the CMR therefore it is important to consider how likely it is that a given severity marker is not only recorded but also encoded via SNOMED.

3.3.11 Code list development

To ensure consistency and accuracy, structured intensional code lists were developed using SNOMED ECL, following the methodologies outlined in Chapter 2. For each severity marker in the final list, our informatics panel determined the most appropriate representation using SNOMED code lists. For example, SNOMED Observable Entities allow the retrieval of measured values associated with a given concept, these were therefore used to define numeric severity markers, such as pulse rate and blood pressure.

3.4 RESULTS

3.4.1 Study selection

Literature searches identified a total of 2,232 unique studies (Figure 3.1). Of these, 2,150 were from database searches and 82 from grey literature All grey literature was put forward for full text review. Following screening of titles and abstracts, 435 database studies were selected for full-text review. Of these, 123 met the eligibility criteria for inclusion. From the grey literature, 3 of the 82 studies met the inclusion criteria. In total, 126 studies were included in the final analysis [4–6,11–133]. The most common reason for exclusion (205/391, 52.4%) was the lack of patient recruitment for any of the specified respiratory clinical syndromes: ILI, ARI, SARI, or suspected COVID-19. The second most common reason was the absence of reporting of a severity marker (70/391, 17.0%).

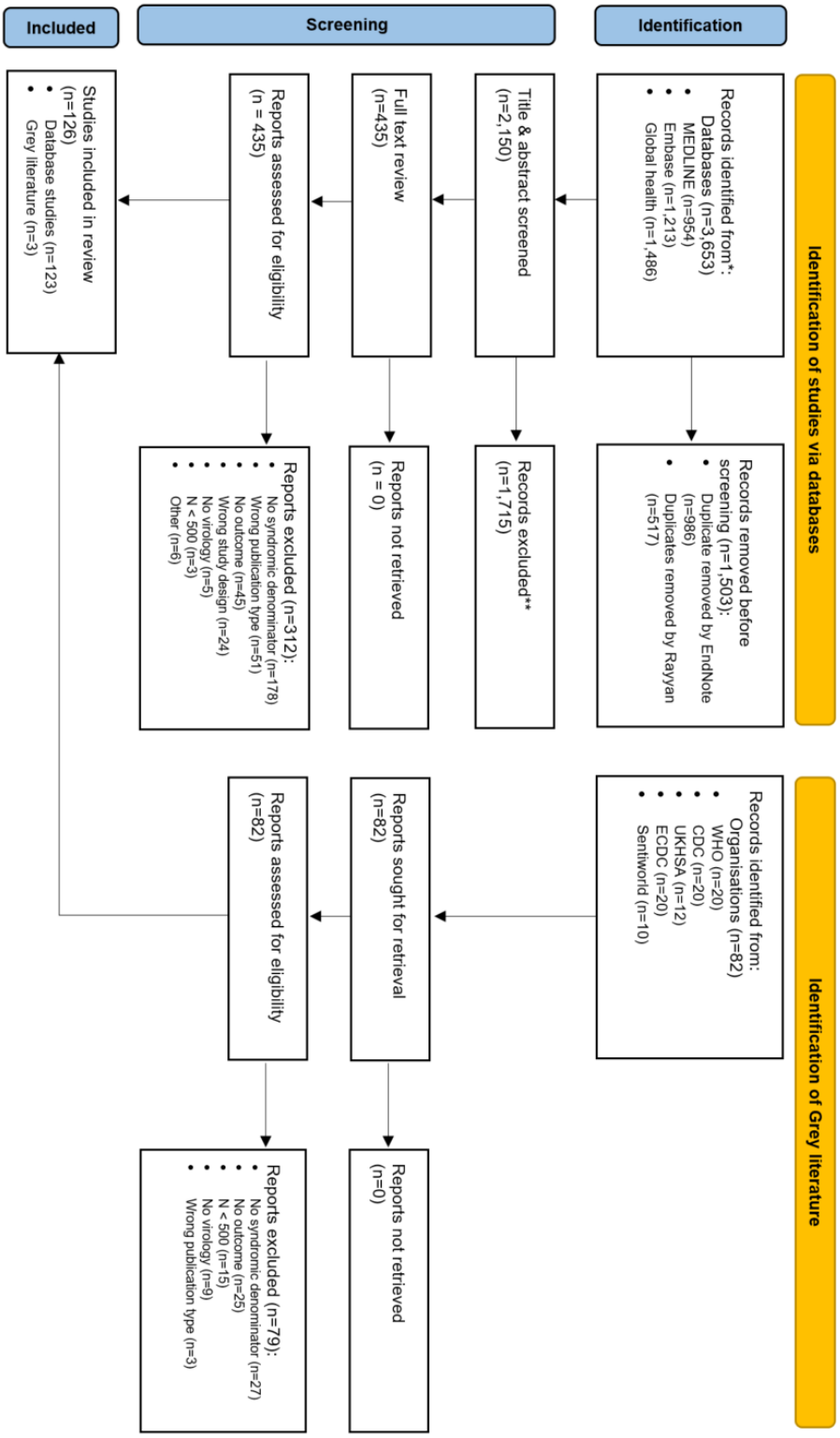


Figure 3.1: PRISMA flow diagram. Adapted from Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

3.4.2 Study characteristics

The included studies varied in nature and focused on a range of ARI-related topics including surveillance, epidemiology, clinical presentation, severity, treatment, outcomes, and disease-burden. Study characteristics are reported in Table 3.2. A full overview of each study is provided in Appendices A2.4.1, A2.4.2 and A2.4.3.

Table 3.2: Summary characteristics of the 126 studies included in the systematic review

Characteristic	Category / Statistic	Value
Publication timeline	Median (range)	Jun 2018 (Oct 2009 – Jul 2023)
Study start	Median (range)	Oct 2010 (Jan 1993 – May 2023)
Study duration	Median (range, days)	763 (7 – 6 208)
Document type	Peer-reviewed articles	118 (94%)
	Surveillance reports	8 (6%)
Geographic scope	Multinational	10 (8%)
	National / sub-national	116 (92%)
World region (255 appearances)	Europe	84 (33%)
	Asia	65 (26%)
	North America	37 (15%)
	Africa	37 (15%)
	South America	25 (10%)
	Oceania	7 (3%)
Income band* (255 appearances)	High	117 (46%)
	Upper–middle	75 (30%)
	Lower–middle	48 (19%)
	Low	14 (6%)
Case types (139 instances)†	SARI	47 (34%)
	ILI	42 (30%)
	ARI	23 (16%)

Table 3.2: continued

Characteristic	Category / Statistic	Value
	Suspected COVID-19	14 (10%)
	MAARI	1 (0.7%)
	Severe pneumonia	1 (0.7%)
	Combined	11 (7.9%)
Case-definition source (139)†	Study-specific	86 (61%)
	WHO	31 (22%)
	Country-specific	10 (7%)
	ECDC	8 (5%)
	Multiple	3 (2%)
	CDC	2 (1.4%)
Recruitment setting (139)†	Hospitalised	72 (52%)
	Treatment-seeking‡	34 (24%)
	Mixed	15 (11%)
	Non-treatment-seeking§	8 (6%)
	ICU	4 (3%)
	Unknown	6 (4%)

Notes: For percentages for world-region, total = 255 as multinational studies contribute >1 country. * Based on World Bank figures from 2024. One country (Venezuela) remained unclassified in 2024 World Bank data [77]. † Thirteen studies reported outcomes in more than one case type, hence 139 reporting instances across the 126 studies. ‡ Treatment-seeking patients included those attending primary care, emergency departments, and other outpatient services.

Most records were peer-reviewed journal articles (118/126, 94%), with peaks in publication around 2009, during the influenza H1N1 (swine flu) pandemic and in 2020 following the start of SARS-CoV-2 pandemic (Appendix A2.5). Of the included studies, 10 (8%) were multinational and the remaining 116 (92.1%) were national or subnational. Data came from 84 countries, most commonly in high-income settings (117/255 country appearances, 46%). Europe was the most commonly represented region (84/255, 33%); however, the

top 3 most represented countries were Mexico (17 studies, 14%), the United States (13, 10%) and Brazil (11, 9%).

3.4.3 Severity markers

Eleven groups of candidate severity markers were identified: 4 groups of severe outcomes and 7 groups of potential predictors of severe outcomes. In total, there were 77 distinct candidate severity markers (Table 3.3) Twenty-one were classified as severe outcomes: 7 complications, 7 hospital-related events, 6 ICU-related events, and death. The remaining 56 were considered potential predictors of severe outcomes. These predictors comprised 20 presenting symptoms, 8 clinical signs, 14 severity scores, 8 investigations, 3 treatments, 2 absenteeism measures, and 1 indicator of health-seeking behaviour.

Table 3.3 shows 77 distinct severity markers identified in the systematic review by group. It also reveals the number of markers subsequently excluded following the primary care informatics review (described in detail in a later section of this chapter) and the number of severity markers included for further use in subsequent chapters of the DPhil.

Table 3.3: Summary of included and excluded severity markers by group

Group	Included	Excluded	Total
All groups (Grand total)	**30**	**47**	**77**
Severe outcomes			
Complications	2	5	7
Hospital	3	4	7
Intensive Care	1	5	6
Death	1	0	1
Subtotal	**7**	**14**	**21**
Predictors of severe outcomes			
Symptom	5	15	20
Health seeking behaviour	1	0	1
Absence	1	1	2
Clinical signs	8	0	8
Clinical scores	2	12	14

Table 3.3 continued

Group	Included	Excluded	Total
Investigation	3	5	8
Treatments	3	0	3
Subtotal	**23**	**33**	**56**

Notes: "Included" refers to markers judged suitable for ARI severity surveillance in UK primary care; "Excluded" were considered unsuitable due to reasons such as lack of specificity, timeliness, or availability in the clinical record. "Total" is the sum of included and excluded markers in each group.

The five most frequently reported severity markers were death (100/139 reporting instances, 72%), ICU admission (62/139, 44%) ventilation (48/139, 35%), cough (47/139, 34%) and dyspnoea (42/139, 35%) (Table 3.2).

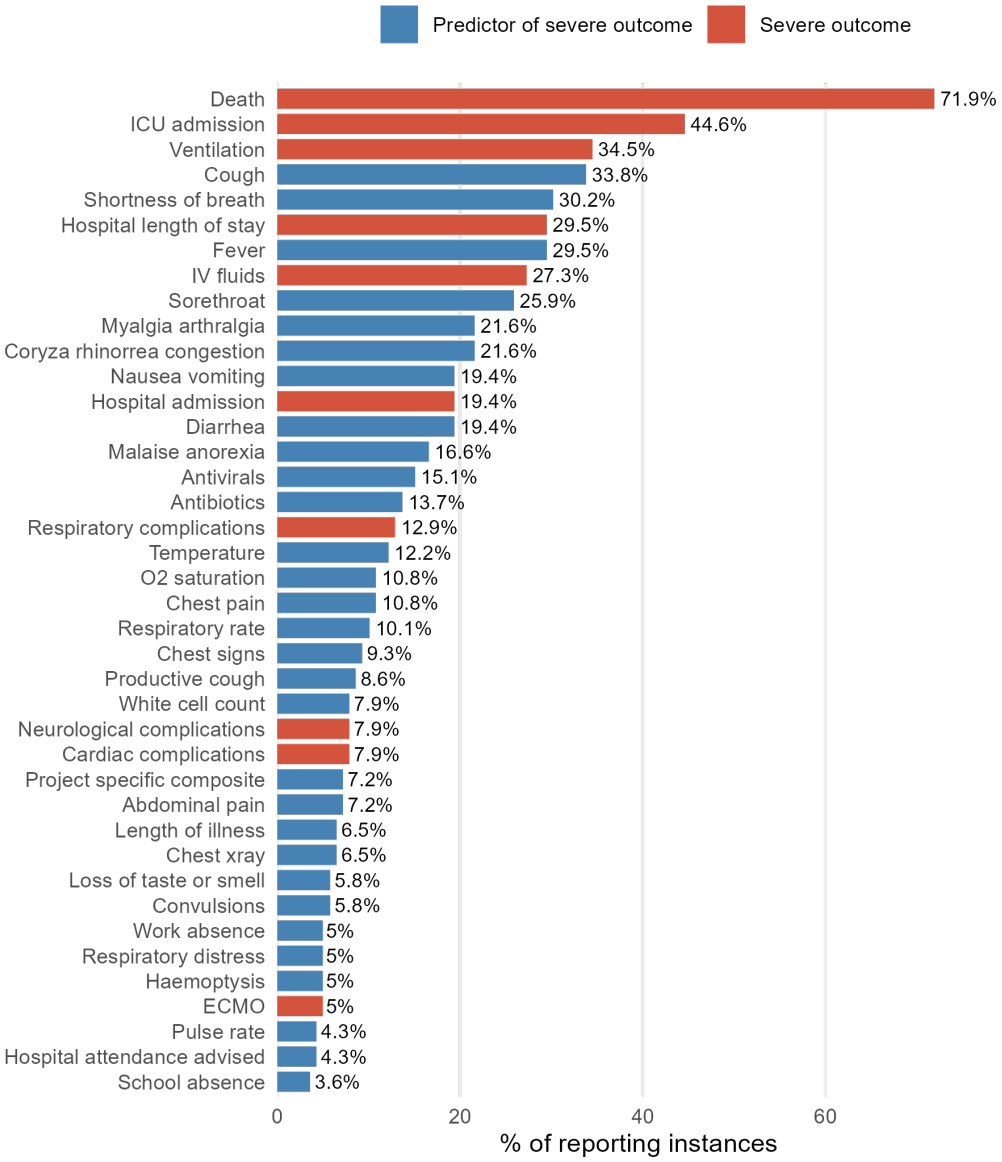


Figure 3.2: Horizontal bars show the top 40 most frequently reported severity markers from the systematic review, ordered by the percentage of studies reporting each marker. Bars are coloured by classification: Severe outcome (red) versus Predictor of severe outcome (blue). Values on bars are exact percentages (% of included studies that reported the marker). **ICU**: intensive care unit; **IV**: intravenous; **ECMO**: extracorporeal membrane oxygenation.

3.4.4 Primary care informatics review

Following the primary care informatics review, 47 of the original 77 candidate markers were deemed unlikely to add value to CMR-based ARI surveillance in primary care. Of the 30 remaining severity markers , 7 were grouped as severe outcomes and 23 as predictors of severe outcomes. Table 3.4 shows all severity markers classified as severe

outcomes and Table 3.5 shows the predictors of severe outcomes identified in the systematic review and whether or not they were included in the final set following the primary care informatics review. The rationale for inclusion or exclusion highlights key considerations that determined whether the marker was included or not.

Whilst there may be pros and cons to inclusion of each marker if included that indicates that on balance it was deemed likely to be valuable. For example, respiratory complications (acute respiratory failure) were included because it was deemed to be a severe outcome and specific enough to relate back to a case of ARI. Whilst not likely to be recorded contemporaneously, it could be recorded following a discharge summary. On the other hand cardiac complications, could occur following an ARI, but this is less specific and harder to directly attribute to the specific ARI case. Another example is the inclusion of a hospital admission or attendance as these are severe and commonly reported outcomes. Furthermore, they are very likely to be recorded (albeit with a delay) following receipt of a hospital discharge summary. Although these are also difficult to directly attribute to an ARI case if they occur in close temporal proximity to the onset of the ARI it would increase the likelihood of there being an association. Hospital length of stay however is not viable as it is rarely if ever recorded in the primary care CMR

Table 3.4: Severity markers: Severe outcomes inclusion status rationale

Severe outcome (n reporting instances from systematic review, %)
COMPLICATIONS
Respiratory complications: INCLUDED (18/139, 13%)
Description: Any acute complication of the respiratory system including respiratory failure and ARDS.
Rationale: Severe outcome. ARI-specific (esp. respiratory failure). Recorded in CMR via hospital discharge summaries, so not timely.
Sepsis: INCLUDED (4/139, 3%)
Description: Severe systemic infection including septic shock and systemic inflammatory response syndrome (SIRS).
Rationale: Severe outcome. Infection-related although not ARI-specific. Recorded in CMR via hospital discharge summaries, so not timely.

Table 3.4: continued

Severe outcome (n, %)
Cardiac complications: EXCLUDED (11/139, 8%) *Description:* Any acute complication of the cardiovascular system including acute coronary syndrome (ACS) and acute heart failure. *Rationale:* Severe, although ACS would likely be recorded in the CMR, less specific for ARI and therefore excluded. Would be recorded from discharge summary and not be timely.
Neurological complications: EXCLUDED (11/139, 8%) *Description:* Acute neurological complications such as transient ischaemic attack (TIA) and cerebrovascular accident (CVA). *Rationale:* Severe, although TIA/CVA would likely be recorded in the CMR but is less specific for ARI and therefore excluded. Would be recorded from discharge summary and not be timely.
Renal complications: EXCLUDED (5/139, 4%) *Description:* Any acute renal complication including acute kidney injury (AKI). *Rationale:* Severe, although AKI would likely be recorded in the CMR, less specific for ARI and therefore excluded. Would be recorded from discharge summary and not be timely.
Organ failure: EXCLUDED (3/139, 3%) *Description:* Any other non-specific organ failure or multiple organ failure. *Rationale:* Severe but non-specific and less likely recorded in the CMR. If recorded not timely.
Haemodynamic shock: EXCLUDED (1/139, 1%) *Description:* Acute circulatory failure. *Rationale:* Severe complication; Non-specific to ARI, therefore excluded. Less likely recorded in the CMR. If re not timely
HOSPITAL
Hospital Admission: INCLUDED (27/139, 19%) *Description:* Emergency hospital admission for an ARI. *Rationale:* Standard epidemiological severe outcome. Admissions are captured via hospital discharge summaries, so reporting is not timely. Included, but in primary care CMRs, both an admission and a diagnosis (e.g., pneumonia) may be recorded, but they are not directly linked, making it difficult to define an ARI-specific admission.
Hospital Attendance: INCLUDED (2/139, 1%) *Description:* Emergency hospital attendance for an ARI.

Table 3.4: continued

Severe outcome (n, %)
Rationale: Standard epidemiological severe outcome. Attendances are captured via emergency department discharge summaries, so reporting is not timely, but may be more timely than an admission for which discharge summaries are often delayed. Included, but again link between reason for attendance and attendance not possible. .
Hospital Attendance Advised: INCLUDED (6/139, 4%) *Description:* Emergency hospital attendance for an ARI advised. *Rationale:* Indicates higher concern by the clinician, as emergency attendance is recommended. More easily attributable to an ARI since the advice is usually recorded at the same time as the ARI consultation in the primary care CMR, so more timely than outcomes captured via hospital discharge summaries. However, such advice is less consistently recorded compared with admissions or attendances.
Length of hospital stay: EXCLUDED (41/139, 29%) *Description:* Duration of time spent in hospital for an ARI. *Rationale:* Not routinely recorded in the primary care CMR therefore excluded.
Oxygen administration: EXCLUDED (4%) *Description:* The administration of oxygen therapy. *Rationale:* Not routinely recorded in the primary care CMR therefore excluded.
Intravenous fluid administration: EXCLUDED (27%) *Description:* The administration of intravenous fluid therapy. *Rationale:* Not routinely recorded in the primary care CMR therefore excluded.
INTENSIVE CARE
Intensive Care Unit Admission (ICU): INCLUDED (62/139, 45%) *Description:* Emergency admission to the ICU for ARI. *Rationale:* Standard epidemiological severe outcome. As for hospital admission. Included but hard to link admission reason to admission event
Mechanical Ventilation: EXCLUDED (48/139, 35%) *Description:* Either non-invasive or invasive ventilator support. *Rationale:* Not routinely recorded in the primary care CMR therefore excluded.
Intensive Care Unit Length of Stay (ICU Length of Stay): EXCLUDED (5/139, 4%) *Description:* Duration of time spent in ICU for an ARI. *Rationale:* Not routinely recorded in the primary care CMR therefore excluded.
Extracorporeal Membrane Oxygenation (ECMO): EXCLUDED (7/139, 7%)

Table 3.4: continued

Severe outcome (n, %)
Description: A life-support measure that provides extracorporeal cardiac and respiratory support. *Rationale:* Not routinely recorded in the primary care CMR therefore excluded.
Inotropic Support: EXCLUDED (2/139, 1%) *Description:* Inotropic support for critically ill patients. *Rationale:* Not routinely recorded in the primary care CMR therefore excluded.
Duration of Ventilation: EXCLUDED (2/139, 1%) *Description:* Duration of time spent on mechanical ventilation for an ARI. *Rationale:* Not routinely recorded in the primary care CMR therefore excluded.
DEATH
Death: INCLUDED (100/139, 72%) *Description:* – *Rationale:* Standard epidemiological severe outcome. Recorded in primary care CMRs therefore included, but attribution to ARI is uncertain as cause of death is not systematically coded. Deaths may be recorded more promptly than hospitalisations, since automatic systems exist in the NHS to capture deaths across all healthcare settings.

Note: Green = Included; Red = Excluded. Within each outcome group, markers are ordered by inclusion status and descending frequency of reporting. Figures represent reporting frequency from the systematic review n (%). **ARI**: acute respiratory infection; **CMR**: computerised medical record; **ICU**: intensive care unit; **ARDS**: acute respiratory distress syndrome.

Table 3.5: Severity markers: *Predictors of severe outcomes* inclusion status rationale

Predictor (n, %)
SYMPTOM
Dyspnoea: INCLUDED (42/139, 30%) *Description:* Subjective shortness of breath. *Rationale:* May indicate compromise of the respiratory system and hypoxia, making it a marker of more severe infection. If recorded, it is usually done at the time of the ARI consultation, making it specific and timely.
Fever: INCLUDED (41/139, 29%)

Table 3.5: continued

Predictor (n, %)
Description: Elevated body temperature. *Rationale:* May indicate a degree of systemic upset and a more widespread response to infection. If recorded, it is usually captured at the time of the ARI consultation, making it specific and timely. Less value as a marker in paediatric patients (where fever is very common) and in older adults (who may not mount a febrile response).
Malaise and Loss of Appetite: INCLUDED (23/139, 17%) *Description:* Generalised tiredness and reduced appetite. *Rationale:* May indicate systemic upset and more severe illness. If recorded, it is usually captured at the time of the ARI consultation, making it timely and specific.
Haemoptysis: INCLUDED (7/139, 5%) *Description:* Coughing up blood. *Rationale:* May indicate more severe respiratory infection due to airway or lung tissue damage and associated inflammation. If recorded, it is usually done at the time of the ARI consultation, making it specific and timely.
Confusion: INCLUDED (3/139, 2%) *Description:* Altered mental state or disorientation. *Rationale:* May indicate systemic upset, especially in very young or older patients. Often linked with hypoxia, sepsis, or shock. If recorded, it is usually captured at the time of the ARI consultation, making it timely and specific.
Cough: EXCLUDED (47/139, 34%) *Description:* Forceful expulsion of air to clear the airways. *Rationale:* Not thought to be a strong enough discriminator of severity therefore excluded.
Sore Throat: EXCLUDED (36/139, 26%) *Description:* Pain or irritation of the throat. *Rationale:* Not thought to be a strong enough discriminator of severity therefore excluded.
Myalgia or Arthralgia: EXCLUDED (30/139, 22%) *Description:* Muscle or joint pain. *Rationale:* Not thought to be a strong enough discriminator of severity therefore excluded.
Nasal Congestion and Rhinorrhoea: EXCLUDED (30/139, 22%) *Description:* Blocked and/or runny nose. *Rationale:* Not thought to be a strong enough discriminator of severity therefore excluded.
Diarrhoea: EXCLUDED (27/139, 19%)

Table 3.5: continued

Predictor (n, %)
Description: Loose or watery stools.
Rationale: Not thought to be a strong enough discriminator of severity therefore excluded.
Nausea and Vomiting: EXCLUDED (27/139, 19%)
Description: Queasiness with or without vomiting.
Rationale: Not thought to be a strong enough discriminator of severity therefore excluded.
Productive Cough: EXCLUDED (12/139, 9%)
Description: Cough producing sputum.
Rationale: Not thought to be a strong enough discriminator of severity therefore excluded.
Chest Pain: EXCLUDED (15/139, 11%)
Description: Pain or discomfort felt in the chest.
Rationale: Not thought to be a strong enough discriminator of severity therefore excluded.
Abdominal Pain: EXCLUDED (10/139, 7%)
Description: Pain localised to the abdomen.
Rationale: Not thought to be a strong enough discriminator of severity therefore excluded.
Loss of Taste or Smell: EXCLUDED (8/139, 6%)
Description: Reduced or absent taste or smell.
Rationale: Recognised indicator of possible SARS-CoV-2 infection, however this doesn't necessarily indicate more severe infection, therefore excluded.
Seizures: EXCLUDED (8/139, 6%)
Description: Febrile convulsions.
Rationale: May be a complication of any febrile illness in child, but may not reflect severity. Although likely recorded less specific therefore excluded.
Ear Pain: EXCLUDED (3/139, 2%)
Description: Earache (otalgia).
Rationale: Not thought to be a strong enough discriminator of severity therefore excluded.
Irritability: EXCLUDED (1/139, 1%)
Description: In a child.
Rationale: Likely marker of severity, but thought to be less likely recorded in the CMR.
Retroocular pain: EXCLUDED (1/139, 1%)
Description: Pain felt behind the eyes.
Rationale: Not thought to be a strong enough discriminator of severity therefore excluded.
Duration of Illness: EXCLUDED (9/139, 6%)

Table 3.5: continued

Predictor (n, %)
Description: Time from symptom onset to presentation. *Rationale:* Not routinely recorded in the primary care CMR therefore excluded.
HEALTH SEEKING BEHAVIOUR
Health seeking behaviour: INCLUDED (3/139, 2%) *Description:* Patient contacts or attends healthcare services. *Rationale:* Includes ambulance encounters or contact with NHS 111 (a free UK advice line). Seeking urgent or unscheduled care may indicate the patient or clinician perceived the illness as more severe. Likely recorded in primary care CMRs. Not ARI-specific, but included on balance.
ABSENCE
Work Absence: INCLUDED (7/139, 5%) *Description:* Time off work due to illness. *Rationale:* Very likely recorded in primary care via electronic fit notes (Med3). Longer absences (e.g., >2 weeks) may reflect greater severity of illness. However, the reason for absence is not always identifiable, and this is only relevant for working-age adults.
School Absence: EXCLUDED (5/139, 4%) *Description:* Time off school due to illness. *Rationale:* Very unlikely recorded as no statutory need to document therefore, excluded.
CLINICAL SIGNS
Body Temperature: INCLUDED (17/139, 12%) *Description:* Measured temperature at presentation. *Rationale:* Fever or abnormal temperature is a recognised marker of systemic upset and can indicate systemic involvement. Likely to be recorded at the time of the ARI event in primary care, therefore timely.
Respiratory Rate: INCLUDED (14/139, 10%) *Description:* Breaths per minute. *Rationale:* An abnormal respiratory rate is a recognised marker of systemic upset and respiratory compromise. Likely to be recorded at the time of the ARI event in primary care, therefore timely.
Chest Examination Findings: INCLUDED (13/139, 9%) *Description:* Auscultatory findings such as wheeze or crackles.

Table 3.5: continued

Predictor (n, %)
Rationale: Findings such as wheeze or crackles may indicate lower respiratory tract involvement and possible complications like pneumonia, indicating greater severity. Likely recorded at time of the ARI consultation in primary care, and therefore timely.
Oxygen Saturation (O_2 Sat): INCLUDED (15/139, 11%)
Description: Peripheral oxygen saturation by pulse oximetry.
Rationale: Oxygen saturation is an indirect measure of hypoxia and a key marker of respiratory compromise and systemic upset. Likely to be recorded at the time of the ARI event in primary care, therefore timely.
Pulse Rate: INCLUDED (6/139, 4%)
Description: Heart beats per minute.
Rationale: Tachycardia or bradycardia is a recognised marker of physiological stress and systemic upset. Likely to be recorded at the time of the ARI event in primary care, therefore timely.
Blood Pressure (BP): INCLUDED (4/139, 3%)
Description: Arterial pressure measured non-invasively.
Rationale: Blood pressure can indicate circulatory compromise in severe ARI. Likely recorded at the time of the ARI event in primary care, therefore timely, but not commonly recorded in children.
Work of Breathing: INCLUDED (7/139, 5%)
Description: Increased effort (e.g., accessory muscle use, grunting).
Rationale: Helps in the assessment of severity as increased effort (e.g., use of accessory muscles, grunting) indicates respiratory distress and possible hypoxia. Likely recorded at the time of the ARI event in primary care, therefore timely.
Cyanosis: INCLUDED (4/139, 3%)
Description: Bluish discolouration of skin or mucosae.
Rationale: Indicates significant hypoxia and is therefore a marker of severe disease. Although relatively uncommon in primary care, if present it is likely to be recorded at the time of the ARI consultation, making it timely.
CLINICAL SCORES
Modified Early Warning Score (MEWS): INCLUDED (1/139, 1%)
Description: Hospital early-warning score based on vital signs (includes MEWS/NEWS/NEWS2/PEWS).

Table 3.5: continued

Predictor (n, %)
Rationale: Captures severity through changes in multiple vital signs and is widely used in hospitals. In the NHS, National Early Warning Score 2 (NEWS2) has replaced MEWS as the standard tool for detecting critical illness. Less commonly recorded in primary care, but included due to its relevance as a cross-sector severity measure.
Glasgow Coma Scale (GCS): INCLUDED (1/139, 1%)
Description: Neurological scale assessing level of consciousness.
Rationale: A neurological scale assessing level of consciousness. Severe reductions in GCS reflect significant systemic or neurological compromise and therefore serve as a strong indicator of severity. Commonly used across the NHS, though less frequently recorded in primary care. Included due to its clear role as a severity marker.
Sequential Organ Failure Assessment score (SOFA): EXCLUDED (3/139, 2%)
Description: Critical care score of organ dysfunction and mortality risk.
Rationale: very unlikely used in primary care therefore excluded.
Simplified Acute Physiology Score II (SAPS II): EXCLUDED (1/139, 1%)
Description: Intensive care physiology-based severity score.
Rationale: very unlikely used in primary care therefore excluded.
Acute Physiology and Chronic Health Evaluation IV (APACHE IV): EXCLUDED (1/139, 1%)
Description: ICU severity score incorporating physiology and comorbidity.
Rationale: very unlikely used in primary care therefore excluded.
Paediatric Risk of Mortality III Score (PRISM III): EXCLUDED (1/139, 1%)
Description: Paediatric intensive care mortality risk score.
Rationale: very unlikely used in primary care therefore excluded.
Barthel Index: EXCLUDED (1/139, 1%)
Description: Activities of daily living functional score.
Rationale: very unlikely used in primary care therefore excluded.
EuroQol EQ-5D: EXCLUDED (1/139, 1%)
Description: Health-related quality-of-life instrument.
Rationale: very unlikely used in primary care therefore excluded.
ViVI Disease Severity Score: EXCLUDED (1/139, 1%)
Description: Influenza-specific clinical severity score.
Rationale: very unlikely used in primary care therefore excluded.

Table 3.5: continued

Predictor (n, %)
Paediatric Chinese Medical Association Score: EXCLUDED (1/139, 1%) *Description:* COVID-specific severity score used in China. *Rationale:* very unlikely used in primary care therefore excluded.
Influenza Symptom Severity Scale (ISS): EXCLUDED (1/139, 1%) *Description:* Patient-reported influenza symptom severity/impact score. *Rationale:* very unlikely used in primary care therefore excluded.
World Health Organization (WHO) severity score: EXCLUDED (2/139, 1%) *Description:* WHO classification of clinical COVID-19 severity. *Rationale:* very unlikely used in primary care therefore excluded.
American Academy Paediatrics Severity Score: EXCLUDED (1/139, 1%) *Description:* Paediatric disease severity scoring tool. *Rationale:* very unlikely used in primary care therefore excluded.
Project-Specific Composite Score: EXCLUDED (10/139, 7%) *Description:* Study-defined combination of markers (e.g., symptoms or hospitalisation+death). *Rationale:* very unlikely used in primary care therefore excluded. However, could be theoretically constructed from a combination of other severity markers, therefore excluded in its own right.
INVESTIGATION
White blood cell count: INCLUDED (11/139, 8%) *Description:* Total circulating leukocytes with differential. *Rationale:* Elevated or abnormal WBC can indicate systemic infection or more severe inflammatory response. If undertaken in primary care, results are recorded in CMRs, but often become available some time after the ARI consultation, reducing timeliness and specificity. However, on balance included.
Inflammatory markers: INCLUDED (5/139, 4%) *Description:* CRP and/or ESR as acute-phase reactants. *Rationale:* Elevated inflammatory markers indicate systemic infection or more severe inflammatory response. If undertaken in primary care, results are recorded in CMRs, but often become available some time after the ARI consultation, reducing timeliness and specificity. However, on balance included.
Chest X-ray: INCLUDED (9/139, 6%) *Description:* Radiograph to assess lung parenchyma and consolidation.

Table 3.5: continued

Predictor (n, %)
Rationale: Findings such as consolidation or parenchymal changes can indicate more severe respiratory disease (e.g., pneumonia). If requested in primary care, results are recorded in CMRs, though they are usually available after the ARI consultation, limiting timeliness and specificity.
Renal function tests: EXCLUDED (3/139, 2%)
Description: Serum urea/creatinine indicators of renal injury.
Rationale: Less specific and commonly ordered for other reasons therefore excluded.
Liver function tests: EXCLUDED (2/139, 1%)
Description: Transaminases and related markers of hepatic injury.
Rationale: Less specific and commonly ordered for other reasons therefore excluded.
Fibrinogen: EXCLUDED (1/139, 1%)
Description: Coagulation protein altered in inflammation/DIC.
Rationale: Very rarely recorded in primary care therefore excluded.
Procalcitonin: EXCLUDED (1/139, 1%)
Description: Biomarker elevated in bacterial infection.
Rationale: Very rarely recorded in primary care therefore excluded.
Arterial blood gas: EXCLUDED (1/139, 1%)
Description: Blood gas assessment of oxygenation/acid–base status.
Rationale: Very rarely recorded in primary care therefore excluded.
TREATMENTS
Antibiotics: INCLUDED (19/139, 14%)
Description: Antibacterial medicines used for suspected bacterial ARI.
Rationale: Prescription of antibiotics often reflects increased clinical concern and may indicate more severe infection. Very well recorded in primary care CMRs and available in a timely manner. Common respiratory antibiotics include amoxicillin, penicillin V, macrolides, doxycycline, co-amoxiclav, and cephalosporins.
Antivirals: INCLUDED (21/139, 15%)
Description: Antiviral medicines (e.g., oseltamivir) for influenza and related infections.
Rationale: Prescription of antivirals may indicate increased severity, depending on the specific drug. Generally well recorded in primary care CMRs and available in a timely manner. SARS-CoV-2 antivirals were excluded as they are rarely prescribed in primary care.
Steroids: INCLUDED (1/139, 1%)

Table 3.5: continued

Predictor (n, %)
Description: Systemic corticosteroids used in selected acute respiratory conditions. *Rationale:* Oral steroids are commonly prescribed in primary care for certain ARIs (e.g., exacerbations of asthma or COPD, and croup). Their use may indicate greater severity. They are highly likely to be recorded in CMRs in a timely manner.

Note: Green = Included; Red = Excluded. Within each predictor group, markers are ordered by inclusion status and descending frequency of reporting. Figures represent reporting frequency n (%). **ARI**: acute respiratory infection; **CMR**: computerised medical record; **CRP**: C-reactive protein; **ESR**: erythrocyte sedimentation rate; **ICU**: intensive care unit; **RR**: respiratory rate; **HR**: heart rate; **CXR**: chest X-ray; **BP**: blood pressure.

3.4.5 Code list development

In total, 36 code lists were developed for the 30 severity markers Table 3.6 and 3.7. Of these 36, 6 were for the severe outcomes and 30 were for the predictors. The number of code lists exceeded the number of severity markers to accommodate 2 specific types of health seeking behaviour deemed to be most relevant in primary care, NHS111 and paramedic (ambulance) encounters. Additionally, 6 antibiotic types were included, amoxicillin, penicillin, macrolides, co-amoxiclav, doxycycline and cephalosporins. Death did not require a code list as it is recorded in a structured (non-free text) field in the CMR.

Of the 36 code lists, 27 were binary (i.e., they represented an event that either occurred or did not occur), and 9 were numeric, representing the recording of a numeric value such as a clinical sign, score, or blood test result.

Table 3.6: Severity markers mapping to code lists: severe Outcomes

Severity marker	**Code list name**	**List type**

Table 3.6: continued

Severity marker	Code list name	List type
SEVERE OUTCOMES		
COMPLICATIONS		
Respiratory failure	AcuteRespiratoryFailure	Binary
Sepsis	Sepsis	Binary
HOSPITAL		
Hospital Admission	HospitalAdmission	Binary
Hospital Attendance	HospitalAttendance	Binary
Hospital Attendance Advised	HospitalAttendanceAdvised	Binary
INTENSIVE CARE		
Intensive Care Unit Admission	ICUAdmission	Binary
DEATH		
Death*	N/A	N/A

*Note: *Death is recorded in a structured field in the clinical medical record (CMR), but not using SNOMED coding*. "Structured field" refers to a dedicated, coded data field as opposed to free text, whereas **SNOMED** (Systematized Nomenclature of Medicine) is a standard clinical terminology for interoperability. In most UK primary care CMR systems, death registration is structured but does not use a SNOMED code.

Table 3.7: Severity markers mapping to code lists: predictors of severe Outcomes

Severity marker	Code list name	List type
PREDICTORS		
SYMPTOM		
Haemoptysis	Haemoptysis	Binary
Shortness of Breath	Dyspnoea	Binary
Fever	FeverSymptom	Binary
Malaise and Loss of Appetite	MalaiseFaitigueAnorexia	Binary
Confusion	Confusion	Binary
HEALTH SEEKING BEHAVIOUR		

Table 3.7: table continued

Severity marker	Code list name	List type
Health seeking behaviour	NHS111*	Binary
	ParamedicEncounter	Binary
ABSENCE		
Work Absence	Med3[†]	Binary
CLINICAL SIGNS		
Body Temperature	Temperature	Numeric
Pulse Rate	PulseRate	Numeric
Respiratory Rate	RespRate	Numeric
Oxygen Saturation (O_2 Sat)	O2Saturation	Numeric
Blood Pressure (BP)	BloodPressure	Numeric
Work of Breathing	RespDistress	Binary
Chest Examination Findings	RespCrackles	Binary
	RespWheeze	Binary
Cyanosis	Cyanosis	Binary
CLINICAL SCORES		
Modified Early Warning Score (MEWS)	NEWS2[‡]	Numeric
Glasgow Coma Scale (GCS)	GCS	Numeric
INVESTIGATION		
White blood cell count	WBC	Numeric
Inflammatory markers	CRP	Numeric
Chest X-ray	CXRRequested	Binary
TREATMENTS		
Antibiotics[§]	Amoxicillin	Binary
	Penicillin	Binary
	Macrolides	Binary
	Co-amoxiclav	Binary
	Doxycycline	Binary
	Cephalosporins	Binary
Antivirals[¶]	Oseltamivir	Binary
Steroids	OralPrednisolone	Binary

*Note:*NHS111 is a free UK healthcare advice line.* †Work Absence is typically captured using electronic fit notes (Med 3). ‡NEWS2 (National Early Warning Score 2) is used as the standard, cross-sector early warning score in the NHS. §Each antibiotic is a separate code list capturing common respiratory antibiotics. ¶Oseltamivir selected as the only antiviral that has been consistently available during the influenza season for many years. SARS-CoV-2 drugs are not commonly prescribed in primary care.

3.5 DISCUSSION

3.5.1 Summary of main findings

In this chapter, I used a systematic review and expert opinion to develop a comprehensive list of candidate severity markers for ARI surveillance. The systematic review analysed 126 studies from 84 countries across various healthcare settings and identified 77 potential markers of severity. These were then reviewed by 4 experienced clinical primary care informaticians who regarded 30 of these 77 as most likely to be of value for CMR-based ARI surveillance. From these 30 markers, 36 code lists were defined, providing a practical way to extract severity data from the RSC CMR and supporting subsequent stages of this DPhil.

The candidate list comprehensively captured severe outcomes and predictors, incorporating studies from diverse settings and healthcare perspectives. Inclusion in the list was based on reviewer discretion rather than strict evidence of an established association between a predictor and a severe outcome. This approach was taken to reflect concepts commonly used in clinical practice, as markers frequently recorded in routine care are more likely to be available for extraction from CMR systems.

Screening of the candidate list by primary care informatics experts identified the most viable markers for use in primary care-based CMR surveillance of ARI. These experts brought both clinical and informatics expertise, providing valuable contextual understanding not only of how data is entered in clinical practice but also of how it is compiled and processed. As a result the final list is likely to represent the most practical and reliable severity markers for further analysis in this DPhil.

3.5.2 Findings in context

Severe outcomes

The severe outcomes included reflect those of a similar or greater severity to hospitalisation. For example, complications, such as sepsis, are highly likely to occur in a hospital setting. These represent hard health outcomes commonly used as endpoints in clinical trials and epidemiological studies and could serve as numerators to derive severity indicators most closely aligned with CHR and CFR [78]. Although, these represent severe outcomes as discussed in the introduction their timeliness is limited as they are recorded only after disease onset, often delayed until a discharge summary is received. Secondly, there is no clear way to explicitly link an ARI episode to an outcome, and a long lag between onset and outcome reduces confidence in their association. To address this, complications more plausibly attributable to ARI were prioritised, such as sepsis and respiratory failure, over less specific concepts like renal failure.

Predictors of severe outcomes

In general, the predictors of severe outcomes selected are those most likely to support timelier reporting of severity indicators. This is because they tend to be recorded contemporaneously during an episode of ARI. For example, symptoms, signs and clinical scores are often used by clinicians to characterise an individual's illness during a primary care encounter. However, further work is required to understand whether these predictors can be used as severity markers:

Firstly, the predictors were selected during the primary care informatics review, as they were considered the most likely to be recorded and used in primary care. However, true recording rates remain unknown. Furthermore, factors beyond severity, such as evolving coding practices, may influence recording rates, potentially leading to increases or decreases in severity indicators, independent of true severity. Before using these markers it will be necessary to measure their recording rates in the primary care CMR and assess how usage changes over time.

Secondly, the extent to which the predictors can actually predict the severe outcomes is at present not known. Many of the predictors, particularly scores, have been shown to predict adverse outcomes in ARI. For example, National Early Warning Score 2 (NEWS2), can identify COVID-19 patients at risk of death or ICU admission within 24 hours of admission to hospital [79, 80]. However, those selected have not been tested as severe outcomes in the context of a CMR-based primary care surveillance system. This work forms the basis of chapter 5 of this DPhil.

Other research

Studies of ARI severity relevant to surveillance are limited, particularly in primary care. This is because, in general, surveillance of severe ARIs is the remit of secondary care surveillance systems [81]. As a result, most studies have focused on severity markers more specific to hospital-based surveillance, for example, use of mechanical ventilation, inotropic support or length of hospital stay [82, 83]. This systematic review, by focusing on primary care surveillance, is a novel contribution to this field.

Many other studies of ARI severity are focused on identifying patient level factors to support clinical risk stratification. This includes development of widely used scoring systems such as NEWS2 and the Sequential Organ Failure Assessment (SOFA) scoring systems [84, 85]. Many of these were identified by this systematic review. The NEWS2 score was included because it is the cross sector NHS critical illness score and could be used to define severity indicators in various settings in the UK. However, any scoring systems are typically validated in a context relevant to their use so here, as with all the severity markers identified, further validation in the context of primary care ARI surveillance is required.

Other considerations

This chapter has focused on identifying disease characteristics or outcomes that can help assess ARI severity. However, it is clear that patient level factors are also incredibly important. This includes age, the patient's socioeconomic status, pre-morbid state and

their vaccination status [86]. Consideration for how these are handled will be required. For example, will severity markers be defined within specific age bands or risk groups.

ARI is a heterogeneous group of conditions and what is considered severe for one ARI subtype may not be considered severe for another. For example, hospital attendance or admission is a relatively common outcome of a URTI particularly in children where as death is not and is not likely to be a suitable severity marker for URTIs [87].

3.5.3 Implications for public health

The severity markers categorised as severe outcomes are those more commonly used to define severity indicators in surveillance, such as CHR and CFR. Although, these indicators are not timely they could support seasonal or annual severity reporting, this could augment RSC annual reports. Identifying potential predictors of adverse outcomes facilitates the development of more timely indicators. Subject to further work, this could reduce the lag in reporting severity and may support near real-time intelligence for public health authorities.

3.5.4 Broader relevance

The work here has been done primarily to meet the needs of this DPhil. However, screening the 77 candidate markers could help generate a list relevant to other research or surveillance settings, broadening the applicability of this work. This is supported by the fact that all severity markers identified for secondary care surveillance in the two aforementioned studies were identified by this systematic review [82, 83]. By extension, those working in other settings, such as, community-based surveillance may also find this work useful [88].

3.5.5 Limitations

The choice to include a severity marker is a subjective decision. Initially, it was down to whether the individual reviewer deemed it suitable for inclusion. The question used to guide the decision on inclusion, the independence of reviewers and use of a third reviewer helped to reduce, but not eliminate, this issue. This could limit the reproducibility of this review.

Furthermore, in identifying potential severity markers, the systematic review did not require a specific predictor to be statistically significantly associated with severe outcomes. While this approach could theoretically limit the predictive value of some markers, it was intended to ensure a comprehensive review and allow for the testing of a broad range of potential markers within the primary care CMR context. Imposing a restriction to include only statistically significant predictors may have introduced bias toward hospital-based predictors, given that most studies were conducted in hospital settings.

The choice of a systematic review, while reasonable, was not the only methodological approach available. A Delphi process could have generated a similar set of severity markers, while also achieving consensus among a broader range of stakeholders [89]. In hindsight, a Delphi approach may have produced greater objectivity in reaching agreement, but would likely have synthesised a narrower range of published evidence. The ideal approach may have been to integrate a Delphi process with the informatics review, thereby combining the breadth of a literature-based synthesis with structured expert consensus.

This DPhil focuses on defining severity markers that are recorded in the primary care CMR. As described, the lack of ability to determine a cause for hospitalisation or a cause for death in the data is a limitation. The effect is that we can only currently assess all-cause hospitalisation or death. If we were able to link these data to other NHS datasets, such as Hospital Episode Statistics (HES), we could then determine a reason for hospitalisation [90]. The same applies to the Office for National Statistics (ONS) death dataset, which contains a cause of death. Linkage is technically possible, but is not within the scope of

this DPhil. Moreover, linkage adds significant technical challenges that would increase the complexity and cost of the surveillance system and reduce its timeliness, as death data and HES data are often only made available up to 2 months after the discharge. Validating the hospitalisation and death data in primary care against the HES and ONS data could be an important piece of work.

3.5.6 Summary

The key implications of this work for the remaining stages of the DPhil are as follows. Firstly, there is potential to develop severity indicators using hard outcomes, though these are likely to be limited in timeliness. Secondly, several potential predictors could form the basis of more timely severity indicators. However, crucially, further work is needed to establish the viability of both approaches. This includes evaluating the data quality of severity markers in the CMR for the relevant ARI subtype and assessing the association between predictors and severe outcomes. The next chapter of my DPhil will examine the recording quality and temporal dynamics of these markers using data from the RSC.

Chapter 4

Data quality assessment of severity markers in primary care computerised medical records

4.1 INTRODUCTION

Through use of a systematic review, Chapter 3 identified thirty potential severity markers for primary care CMR-based ARI surveillance (Table 3.6 and 3.7). In this chapter, I evaluate the data quality of the candidate severity markers in the primary care CMR, assessing the extent to which they are recorded and the stability of recording over time.

This work is necessary because CMR data is primarily collected to support patient care rather than for alternative purposes, such as ARI surveillance [67]. While clinicians are responsible for maintaining high-quality patient records, their focus when entering data into the primary care CMR is on meeting the immediate needs of the patient and their organisation, rather than on the data's potential use beyond that context [91, 92]. Consequently, it cannot be assumed that the data will meet the needs of any given additional purpose [93, 94]. Evaluation of the data quality of candidate severity markers is therefore highly advisable prior to their use for surveillance.

There are a number of frameworks for assessing the quality of CMR data used for secondary purposes [95, 96]. Using expert opinion Kahn et al. identified three data quality 'categories', conformance, completeness and plausibility [96]. Conformance refers to whether a data item meets the technical constraints of a given data model. Completeness relates to whether expected data are present or not and plausibility refers to whether the values present are believable or not. Plausibility includes a temporal component relating

to the believability of changes in recording over time.

While data quality frameworks are valuable, they are context specific and must be tailored to the given use case. Here, I focus on measuring the completeness of recording of numeric severity markers in the context of ARI and the plausibility of observed temporal changes in severity marker recording. Conformance is largely handled internally by the RSC's automated data cleaning processes and is not further considered in this DPhil. Non-temporal plausibility is also handled as part of data processing, for example, non-plausible values such as negative blood pressure values are excluded.

Here, I assess completeness of recording of severity markers because, although the primary care informatics review selected markers that were theoretically used in practice, the extent to which they are actually recorded in the CMR is unknown and needs to be assessed. This is particularly important because this DPhil relies on SNOMED encoded data alone and doesn't include free text. If clinical signs are recorded in only a minority of cases then these are not likely to be of value as ARI severity markers. However, these assertions must be tested empirically.

Assessing the plausibility of temporal changes (or temporal dynamics) in severity marker recording is essential in the context of this DPhil. Surveillance focuses on monitoring changes in indicators over time, such as CHR. However, temporal dynamics may be influenced by a number of non-severity related factors. For example, the introduction of pay for performance schemes such as QoF and the transition to new clinical terminologies [97–100]. Additionally, significant global events like the pandemic are known to have resulted in changes in coding patterns of chronic disease. It is highly likely such events would also influence recording of ARI-related severity markers [101].

4.2 CHAPTER AIM AND OBJECTIVES

4.2.1 Aim

Assess the data quality of severity markers in the primary care CMR.

4.2.2 Objectives

1. **Measure severity marker recording completeness**: Use recording rates to assess the completeness of ARI severity marker recording for ARI and ARI subtypes in the primary care CMR.

2. **Assess the temporal dynamics of severity markers**: Use weekly recording rates over time, to assess the temporal dynamics of recording rates of ARI severity markers in the primary care CMR.

4.3 METHODS

I conducted a retrospective cohort study using data from ORCHID. A detailed description of the RSC and its data infrastructure including ORCHID is presented in chapter 1. [40]. In brief, the data are securely hosted within ORCHID, a large, secure Trusted Research Environment (TRE) that receives and stores routinely collected primary care CMR data from two data providers: Optum Health, formerly Egton Medical Information Systems (EMIS) and Magentus [40]. For this study, only data from practices using Optum Health systems were analysed due to information governance constraints.

The Optum dataset includes information from approximately 1,200 primary care practices across England, covering nearly 20 million patients, of whom around 12 million are currently active. It contains demographic data (including date of death) and SNOMED–encoded clinical information, including diagnoses of acute and chronic conditions, immunisations, prescriptions, investigations, and other relevant clinical concepts. Data encoded in free-text is not made available by data providers to the RSC. Thus, this work does not include any free-text analysis.

4.3.1 Study population

Prior to applying inclusion and exclusion criteria, standard RSC data quality procedures were conducted to remove ambiguous or invalid records. These procedures excluded only

0.05% of patients from the cohort. The procedures included:

1. Only data from the most recently registered practice was included for individuals registered with multiple practices. This includes any data transferred from the patients previous practices and reduces duplication.
2. Individuals without a valid NHS number were not eligible for inclusion.
3. Those registered with an RSC practice but recorded as deceased before the study start date were also excluded.

4.3.2 Eligibility criteria

Inclusion criteria

All individuals passing data quality checks registered with an RSC practice were eligible for inclusion, regardless of age. The study included individuals with a valid ARI episode recorded in the CMR between ISO week 40 of 2008 (29th of September 2008) and ISO week 39 of 2024 (29th September 2024), spanning 16 influenza seasons (2008-2009 to 2023-2024).

Exclusion criteria

To ensure there was sufficient time for outcomes to be observed over the 56-day risk window, I excluded individuals who contributed fewer than 56 days of person-time to the cohort and had no recorded date of death. At the episode level, ARI episodes were excluded if the patient de-registered within 56 days of the episode start date.

This approach could introduce collider selection bias, because eligibility for inclusion then depends on length of registration at the practice. If the factors that influence early de-registration are also related to risk of subsequent ARI or to the chance of a severe outcome, this conditioning can distort the associations of interest.

Retention of those that died within 56 days will have limited the impact of this, but theoretically some unknown association may have remained. However, because de-registration

within a 56-day window was uncommon in this dataset any residual collider bias is likely to be small, although this remains an acknowledged limitation.

Additionally, individuals who had activated their right to opt out of data sharing through the National Data Opt-out (NDOO) or Type-1 GP-based opt out were excluded from the study [102].

4.3.3 Data characteristics and preparation

ARI Episode definition

Episodes of ARI were identified using the PhA developed in Chapter 2, which defines ARI as a hierarchical condition with three levels (Figure 2.1) [44]. In this chapter, Level 2 ARI-NOS is separated into ARI-NOS and suspected COVID to allow closer inspection of the data quality for suspected COVID-19 episodes. To manage duplicate coding, ARI recorded within 28 days is treated as a single episode.

Severity marker definition

Severity markers were identified using the code lists curated in Chapter 3 and categorised as either severe outcomes or predictors of severe outcomes. Severe outcomes refer to less timely 'hard' clinical endpoints such as death, ICU admission, hospital attendance or admission, and serious complications including sepsis and acute respiratory failure. In contrast, predictors of severe outcomes such as symptoms, clinical signs, or scores are less definitive but may be recorded earlier in the episode of ARI.

Variables

The primary dataset used for this analysis consisted of a single table, with each row representing an ARI episode. Variables included unique practice and patient identifiers, episode date variables, ARI Level 2 subtype (e.g., ILI or URTI), age band at the time of the episode, risk group status, and a flag indicating whether or not a given severity marker was recorded (Table 4.1).

Table 4.1: Variables extracted for severity marker data quality assessment

Variable / Group	Description
Unique patient ID	A pseudonymised identifier, distinct from the NHS number.
Unique practice ID	A non-traceable label identifying the practice to which the individual belonged.
Episode date	The earliest date of the given ARI episode.
Episode ISO week	The ISO week of the given ARI episode.
Study period	The study period in which the episode occurred. Study periods are described in the Specific Objectives section below.
ARI subtype	The ARI classification based on its level 2 indicator, e.g., URTI.
Age band	Age bands as used for RSC weekly surveillance reports.
Risk group status	Indicates whether an individual was considered to belong to a clinical risk group, as defined by UKHSA [50]. These groups are operationalised using PRIMIS business rules, which map SNOMED code lists to specific clinical conditions [51].
Severity markers	**Severe outcomes:** A binary flag indicating whether a severe outcome was recorded in the CMR, defined as occurring within −7 to 56 days of the episode start date. **Predictors of severe outcomes:** A binary flag for each predictor recorded within −7 to 14 days of the episode, allowing for minor date inaccuracy in CMR event recording.

Notes: NHS: National Health Service; ARI: acute respiratory infection; ISO: International Organization for Standardization; URTI: upper respiratory tract infection; RSC: Oxford–Royal College of General Practitioners Research and Surveillance Centre; UKHSA: United Kingdom Health Security Agency; PRIMIS: University of Nottingham, Primary Care Information Services; SNOMED: Systematized Nomenclature of Medicine; CMR: computerised medical record.

4.3.4 Specific objectives

Measuring recording rates

Recording rates for each severity marker were calculated. A recording rate is the percentage of ARI episodes in which a given severity marker is recorded in the CMR. These are used to address objectives 1 and 2 as outlined in Table 4.2. Not all severity markers were applicable to all ARI episodes. For example, work absence is not relevant for all age bands. I report overall recording rates for all severity markers regardless of their target population but also calculate rates by ARI subtype and age band to enable more specific assessment within relevant clinical groups.

Table 4.2: Use of recording rates to address study objectives on completeness (Objective 1) and temporal dynamics (Objective 2) of severity marker recording

Subgroup	Use of recording rates	Objective addressed
Overall ARI	The recording rate of all severity markers for the whole study period, focusing on clinical signs and clinical scores.	Objective 1
ARI subtype	The recording rate of all severity markers for each ARI subtype (Figure 2.1) for the whole study period.	Objective 1
Age band	The recording rate of all severity markers for each age band for the whole study period.	Objective 1
ISO week	The recording rate of all severity markers for all ARI episodes by ISO week in the study period. Used to derive time-series plots.	Objective 2
Study periods	The median weekly recording rates and interquartile ranges calculated for the whole study period and separately for each of the four study periods.	Objective 2

Notes: ARI: acute respiratory infection; ISO: International Organization for Standardization;

4.3.5 Episode summary

Prior to presenting results for the main objectives a summary of episode numbers for overall ARI and by ARI subtype and age band are reported.

4.3.6 Completeness

Due to lack of access to free text, data completeness refers to the extent to which a clinical truth is captured in the CMR using SNOMED codes. Completeness is estimated using the recording rates of severity markers; however, these rates do not always accurately reflect data completeness. To aid interpretation, severity markers can be grouped into three contextual categories. Each group illustrates how recording rates relate to completeness. Recognising these distinctions enables a more appropriate interpretation of recording rates:

Context 1: Valid completeness measure

For many numeric severity markers, such as clinical signs, recording rates provide a direct indication of completeness. For example, every individual with an ARI has a pulse rate, whether or not it is recorded. If a pulse rate is not present in the CMR, we can be confident that it was simply not documented, rather than absent in the patient. In these cases, missing data clearly represents incomplete recording (although it may have been recorded in free text).

Context 2: Structurally complete

For some binary markers, such as prescriptions, completeness is inherently high due to the way these data are captured. For example, when a prescription is issued electronically, a code is automatically added to the CMR. Therefore, if a prescription code is not present, it is highly likely that no prescription was issued. While exceptions exist, such as handwritten or private prescriptions, these are uncommon. In this case, recording rates do not add much to our understanding of completeness, but they do offer a relatively accurate reflection of

prescribing activity.

Context 3: Structurally incomplete

Other binary markers, such as symptoms, lack structural mechanisms to ensure consistent recording. Their documentation depends on the judgment of the clinician. If a symptom (for example) is not recorded, it isn't clear whether the patient did not experience it or whether it was simply not documented. This uncertainty means that recording rates cannot be interpreted as a reliable measure of completeness. Some insights can be gained by comparing observed recording rates to expected prevalence from literature or other data sources, but this approach is imprecise. We know that the data is likely to be incomplete but we can't be certain to what extent. Table 4.3 categorises severity markers according to these three contexts.

Table 4.3: Categorisation of severity markers by how recording rates relate to completeness

Severity marker group	Implications for completeness
VALID COMPLETENESS MEASURE	
Clinical Signs	Numeric signs (e.g., pulse rate) can in principle be recorded for all patients regardless of severity. The percentage of ARI episodes with recorded signs directly reflects completeness. Binary signs (e.g., cyanosis, work of breathing, chest signs) provide only partial insight because "recorded" ≈ "present".
Clinical Scores	As for clinical signs.
STRUCTURALLY COMPLETE	
Work Absenteeism	For working-age adults, sickness absence is recorded electronically when a fit note (Med 3) is issued, embedding a SNOMED code in the CMR. The absence of a Med 3 entry suggests no certificate was issued. Older paper Med 3 forms may be missing from the CMR. Prior to 2012, EMIS systems lacked electronic fit notes, so completeness may be limited [103].

Table 4.3: table continued

Severity marker group	Implications for completeness
Prescriptions	When a prescription is issued in primary care, a DM + D (SNOMED CT medicines extension) code is automatically embedded in the record. Absence of a prescription code likely indicates no prescription. Some COVID-19 antivirals are prescribed in secondary care and may not appear in the CMR.
Death	Deaths may be manually coded via SNOMED; if occurring elsewhere, the CMR is automatically updated via the PDS [104]. If not recorded as deceased, a patient is probably alive. PDS updates should occur within 1 day, though delays can happen [105].
STRUCTURALLY INCOMPLETE	
Symptoms	Often recorded during encounters, but documentation is at clinician discretion. Absence may reflect non-occurrence or omission.
Health Seeking	Typically recorded when a summary document is received; inclusion depends on practice workflows. Absence may indicate non-occurrence or failure to record.
Complications	Usually appear after hospital discharge; presence in the CMR depends on documentation workflows. Absence may reflect non-occurrence or omission; e.g., sepsis is known to be incompletely recorded [106].
Hospitalisation	Typically recorded following receipt of a discharge summary. Absence may mean either no admission or lack of documentation.
Intensive Care	As for hospitalisation.
Investigations	Results of primary care ordered tests are usually complete (electronic feeds). Hospital-ordered results often do not appear in the CMR; absence may mean “not ordered” or “done in secondary care”.

Notes: CMR: Computerised Medical Record; EMIS: Egton Medical Information Systems; PDS: Personal Demographics Service; SNOMED: SNOMED CT; DM + D: Dictionary of Medicines and Devices.

Analysis

Recording rates are presented for all severity markers, by category and group. The category of the severity marker refers to whether it was a severe outcome or a predictor of a severe

outcome. The group refers to the type of severity marker. For example, symptoms are a group of severity markers within the category of predictors of severe outcomes.

A detailed assessment of completeness is only undertaken for severity markers where recording rates are a valid measure of completeness (Table 4.3). Completeness is measured for all ARI, ARI subtype and age band. In large datasets such as the RSC CMR, tests of statistical significance may not add much value. This is because even very small differences between groups can produce extremely small p-values, resulting in statistically significant findings that may lack clinical relevance. For this reason, I did not report p-values.

To assess the independent association between data completeness of clinical signs and ARI subtype, age band and risk group status, a multivariable logistic regression was performed. In this analysis, for each individual a random ARI episode was selected to maintain independence of observations. This avoids within-person correlation, which can bias model standard errors.

In this model the dependent variable is a binary flag identifying where temperature or pulse rate or respiratory rate or oxygen saturation or blood pressure were recorded. No missing data is present in this analysis, however if no risk group is recorded the individual is assumed not to be in a risk group, which likely underestimates the true number of individuals in a risk group.

Independent variables included: ARI subtype, age band and risk group status. The model formula is presented in Table 4.4. odds ratio (OR) and 95% confidence interval (CI) are presented. Variance inflation factors were calculated to assess multicollinearity, and model fit was evaluated using McFadden's pseudo R-squared.

Table 4.4: Logistic regression model specification for data completeness

Specification
Model formula:
$$\text{logit}\{P(Y=1)\} = \beta_0 + \sum_{i=1}^{k-1} \beta_{1i}\,(\text{ARI subtype}_i) + \sum_{j=1}^{m-1} \beta_{2j}\,(\text{age band}_j) + \beta_3\,(\text{risk group})\,.$$
Outcome definition:
$Y = 1$ if *any* of the following clinical signs were recorded: temperature, pulse rate, respiratory rate, oxygen saturation, or blood pressure.
Reference categories (for β_0):
Reference levels: ARI subtype = URTI; age band = 15–64 years; risk group = no.
β_{1i} (ARI subtype effects):
Change in log-odds of $Y = 1$ for ARI subtype i, for each of the $k - 1$ non-reference subtypes.
β_{2j} (age band effects):
Change in log-odds of $Y = 1$ for age band j, for each of the $m - 1$ non-reference age bands.
β_3 (risk group effect):
Change in log-odds of $Y = 1$ for individuals in a clinical risk group (vs not in a risk group).

Notes: Model fitted as a multivariable logistic regression on a random ARI episode per individual to avoid within-person correlation. No missing data in predictors; absence of a recorded risk group is treated as not in a risk group (likely underestimates true risk-group prevalence). ARI: acute respiratory infection; URTI: upper respiratory tract infection.

4.3.7 Temporal dynamics

To assess the temporal dynamics of severity markers in the CMR, recording rates for all ARI episodes were calculated by episode week from ISO week 40 of 2008 (29th September 2008) to ISO week 39 of 2024 (29th September 2024). While recording rates cannot be used to measure data completeness for all severity markers, changes in recording rates over time can provide insights into the temporal dynamics of all severity markers. To support

interpretation of temporal dynamics, the study period was divided into four distinct phases centered on key milestones in the COVID-19 pandemic (Figure 4.1). Period 3 was defined first, encompassing the COVID-19 pandemic, from the beginning of the month of the first confirmed UK episode to the end of the respiratory season during which the final legal restrictions were lifted [107, 108]. Everything after Period 3 represents the post-pandemic period (Period 4). The pre-pandemic period was split into two equal phases: the early part (Period 1) and the later part (Period 2).

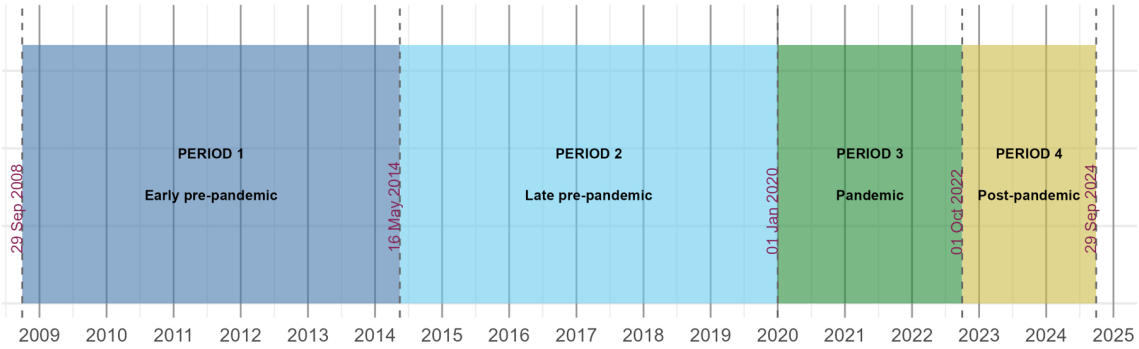


Figure 4.1: Time line showing how periods of the study are defined and how they relate to COVID-19 pandemic.

- **Period 1 (Early pre-pandemic period):** Start: 29 September 2008, End: 16 May 2014.
- **Period 2 (Late pre-pandemic period):** Start: 17 May 2014, End: 31 December 2019.
- **Period 3 (Pandemic period):** Start: 1 January 2020, End: 30 September 2022.
- **Period 4 (Post-pandemic period):** Start: 1 October 2022, End: 29 September 2024.

Comparing recording rates in Period 2 with Period 1 helps assess the magnitude of changes in recording rates occurring prior to the pandemic. Comparison of Period 3 and 2 assesses the magnitude of any changes in recording rates that occurred following the start of the pandemic. Comparison of Period 4 with Period 3 helps evaluate how recording rates have changed in the aftermath of the pandemic. Changes in recording rates may occur gradually rather than in a discrete manner however, the combination of comparing recording rates

in the periods and the plotting of time series allows broad quantification of differences but also a visual assessment of the exact nature of these changes (e.g., gradual or abrupt changes).

Analysis

To allow comparison of recording rates between study periods, median weekly recording rates and interquartile ranges (IQRs) were calculated for all severity markers. Fold change was used to quantify the relative difference in recording rates between two time periods and is defined as the ratio of the recording rate in a comparison period to the recording rate in a reference period (Table 4.5). A fold change greater than 1 indicates an increase, while a fold change less than 1 indicates a decrease, relative to the reference (denominator period in the ratio).

Table 4.5: Fold change of median weekly recording rates

Specification
Definition: $$\text{Fold change} = \frac{\text{mwrr}_{\text{period B}}}{\text{mwrr}_{\text{period A}}}$$
Where: • mwrr = median weekly recording rate. • period A = reference period. • period B = comparison period.

Table 4.5: Fold change of median weekly recording rates (continued)

Specification
Interpretation:
• Fold change > 1: relative *increase* in median weekly recording rates from period A to B.
• Fold change < 1: relative *decrease* in median weekly recording rates from period A to B.
Narrative:
Fold change reflects the *relative change* in median weekly recording rates between two periods (A → B).

Additionally, time series plots of weekly recording rates were generated to visually assess more temporally granular changes in recording rates.

4.4 RESULTS

4.4.1 Episode summary

A total of 8,726,583 unique individuals contributed 25,724,066 episodes of ARI over the study period (Table 4.6). The majority of ARI episodes were URTIs (58.80%), followed by LRTIs (30.52%), ECLD (3.88%), suspected COVID-19 (2.81%), ILI (2.29%) and ARI NOS (1.69%). It is important to note that the age bands are not evenly distributed as they follow the RSC age groupings used in UKHSA surveillance reporting. Most ARI episodes occurred in individuals aged 15 to 64 years (49.65%) followed by those aged over 65 years (18.31%), 1 to 4 year olds (14.66%), 5 to 14 years olds (12.66%) and under 1 year olds (4.73%).

Table 4.6: Summary of ARI episodes recorded, by age band and ARI subtype

Subtype	All ages	<1 yr	1–4 yrs	5–14 yrs	15–64 yrs	65+ yrs
All ARI	25,724,066 (100.00%)	1,217,235 (4.73%)	3,770,203 (14.66%)	3,255,466 (12.66%)	12,771,509 (49.65%)	4,709,653 (18.31%)
URTI	15,124,965 (58.80%)	909,034 (3.53%)	2,993,194 (11.64%)	2,651,556 (10.31%)	7,392,484 (28.74%)	1,178,697 (4.58%)
LRTI	7,852,119 (30.52%)	278,021 (1.08%)	662,091 (2.57%)	400,448 (1.56%)	3,755,049 (14.60%)	2,756,510 (10.72%)
ECLD	999,025 (3.88%)	122 (0.00%)	15,977 (0.06%)	57,822 (0.22%)	499,678 (1.94%)	425,426 (1.65%)
COVID	723,658 (2.81%)	7,868 (0.03%)	16,817 (0.07%)	58,844 (0.23%)	478,105 (1.86%)	162,024 (0.63%)
ILI	589,961 (2.29%)	5,640 (0.02%)	27,215 (0.11%)	45,930 (0.18%)	435,578 (1.69%)	75,598 (0.29%)
ARI NOS	434,338 (1.69%)	16,550 (0.06%)	54,909 (0.21%)	40,866 (0.16%)	210,615 (0.82%)	111,398 (0.43%)

Notes: URTI: Upper respiratory tract infection; LRTI: Lower respiratory tract infection; ARI NOS: ARI not otherwise specified; ECLD: Exacerbations of chronic lung disease; ILI: Influenza-like illness; COVID: Suspected COVID-19. Cells show the number of ARI episodes, with percentages in brackets giving the proportion of all ARI episodes across all ages (N = 25,724,066) represented by each cell (i.e. percentages do not sum to 100% within rows or columns).

4.4.2 Recording rates

Severe outcomes were recorded in 2,032,592 (7.90%) of the 25,724,066 ARI episodes (Table 4.7). Hospital-related outcomes were the most commonly recorded, documented in 7.44% of episodes. Hospital attendance was more frequently recorded than hospital admission (6.56% vs. 1.56%). Death within 56 days was recorded in 0.62% of episodes. Complications were recorded in 0.18% of episodes, with sepsis being the most frequently documented complication (0.16%). ICU admission was rarely recorded, just 0.01% of episodes.

Table 4.7: Recording rates for severe outcomes and predictors of severe outcomes by severity marker group

Severity marker category	Severity marker group	Recording rate n (%)
Severe outcomes 2,032,592 (7.90%)	Hospital	1,915,077 (7.44%)
	Died ≤ 56 days	160,387 (0.62%)
	Complications	45,058 (0.18%)
	ICU admission	2,038 (0.01%)
Any predictor 19,777,697 (76.88%)	Prescription	15,991,549 (62.16%)
	Sign	9,033,534 (35.11%)
	Investigation	1,811,721 (7.04%)
	Symptom	979,188 (3.81%)
	Absenteeism work	928,828 (3.61%)
	Health seeking	466,186 (1.81%)
	Score	12,816 (0.04%)
	Hospital	31,477 (0.12%)

Notes: ICU: intensive care unit.

Predictors of severe outcomes were recorded in 19,777,697 (76.88%) of episodes. Prescriptions for antibiotics, antivirals, or steroids were the most commonly recorded predictor, documented in 62.16% of episodes. Among these, amoxicillin was the most frequently prescribed antibiotic, recorded in 34.88% of episodes. Oseltamivir, the most commonly recorded antiviral, was documented in 0.12% of all ARI episodes and 4.14% of ILI episodes. Clinical signs were the next most common, recorded in 35.11% of episodes- details are presented in the next section. Investigations were recorded in 7.04% of episodes. Symptoms were documented in 3.81% of episodes, with dyspnoea being the most frequently recorded symptom (1.86%). Work absenteeism, indicated by Med 3 fit note recording, was noted in 3.61% of all episodes and in 7.02% of episodes occurring in 15 to 64 year olds. Health-seeking behaviours, including NHS 111 contacts or ambulance encounters, were recorded in 1.81% of episodes. Clinical scores were recorded in just 0.04%, and emergency hospital referrals in 0.12% of episodes.

4.4.3 Completeness

Here I present the completeness analysis for severity markers where recording rates are valid measures of completeness, numeric clinical signs and scores (Table 4.3).

All ARI episodes

Clinical signs were far more commonly recorded than clinical scores (35.11% vs. 0.04%). Temperature was the most commonly recorded sign, with 21.51% of all ARI episodes having a record (Figure 4.2). This was followed by pulse rate (18.38%), oxygen saturation (16.40%), blood pressure (14.24%) and respiratory rate (7.44%). Both Glasgow Coma Scale (GCS) and NEWS2 were very rarely recorded (0.04% and 0.02%).

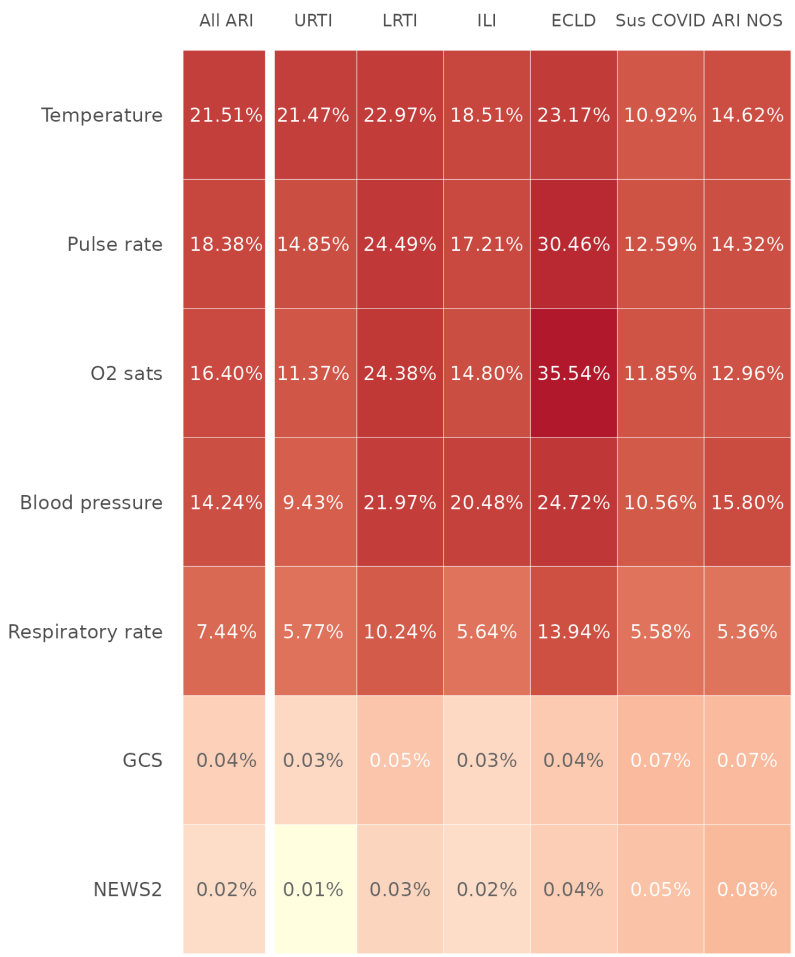


Figure 4.2: Completeness of recording of clinical signs and clinical scores by all ARI and ARI subtype. ARI: Acute respiratory infection; URTI: Upper respiratory tract infection;LRTI: lower respiratory tract infection; ILI: Influenza-like illness, ECLD: Exacerbation of chronic lung disease; Sus COVID: suspected COVID-19; ARI NOS: ARI not otherwise specified; O2 Sats: Peripheral oxygen saturation; GCS: Glasgow Coma Scale; NEWS2: National Early Warning Score 2.

By ARI subtype

The completeness of severity marker recording varied substantially by ARI subtype (Figure 4.2). All clinical signs were more commonly recorded in episodes of ECLD than in any other ARI subtype. This was particularly notable for oxygen saturation, which was recorded in 35.54% of ECLD episodes compared to 24.38% of LRTI episodes, the next most common. Pulse rate was similarly more frequently recorded in ECLD episodes (30.46%) compared to LRTI (24.49%). Completeness of recording of clinical scores was consistently low across all subtypes of ARI.

By age band

There was also variability in the completeness of clinical sign recording by age band (Appendix A3.1). Temperature was more frequently recorded in children under 15 years compared to older age groups, ranging from 24.60% to 26.21% in younger age bands versus 19.30% to 20.42% in older ones. In contrast, pulse rate, oxygen saturation, and blood pressure were more commonly recorded with increasing age. For example, pulse rate was recorded in 13.73% of children under 1 year, compared to 25.13% of those aged 65 years and older. Oxygen saturation was recorded in 5.10% of children under 1 year and 25.07% of the oldest age group, while blood pressure was recorded in only 0.04% of infants but in 29.91% of those aged 65 and above. Respiratory rate showed a different pattern, being most commonly recorded in children under 1 year (11.01%), followed by those aged 1 to 4 years (9.69%) and adults aged 65 years and over (9.10%). Clinical scores were slightly more commonly recorded in older adults, but in less than 0.1% of all episodes regardless of age.

Adjusted analysis

Compared to URTIs, complete recording was more likely in episodes of ECLD, LRTI, and ILI, with adjusted ORs of 1.62 (95% CI: 1.61-1.64), 1.46 (95% CI: 1.46-1.47), and 1.07 (95% CI: 1.06-1.08), respectively (Figure 4.3). In contrast, ARI NOS and suspected COVID were associated with a lower likelihood of complete recording, with ORs of 0.96 (95% CI: 0.95-0.97) and 0.43 (95% CI: 0.42-0.43), respectively. Older age was generally associated with increased odds of complete recording.

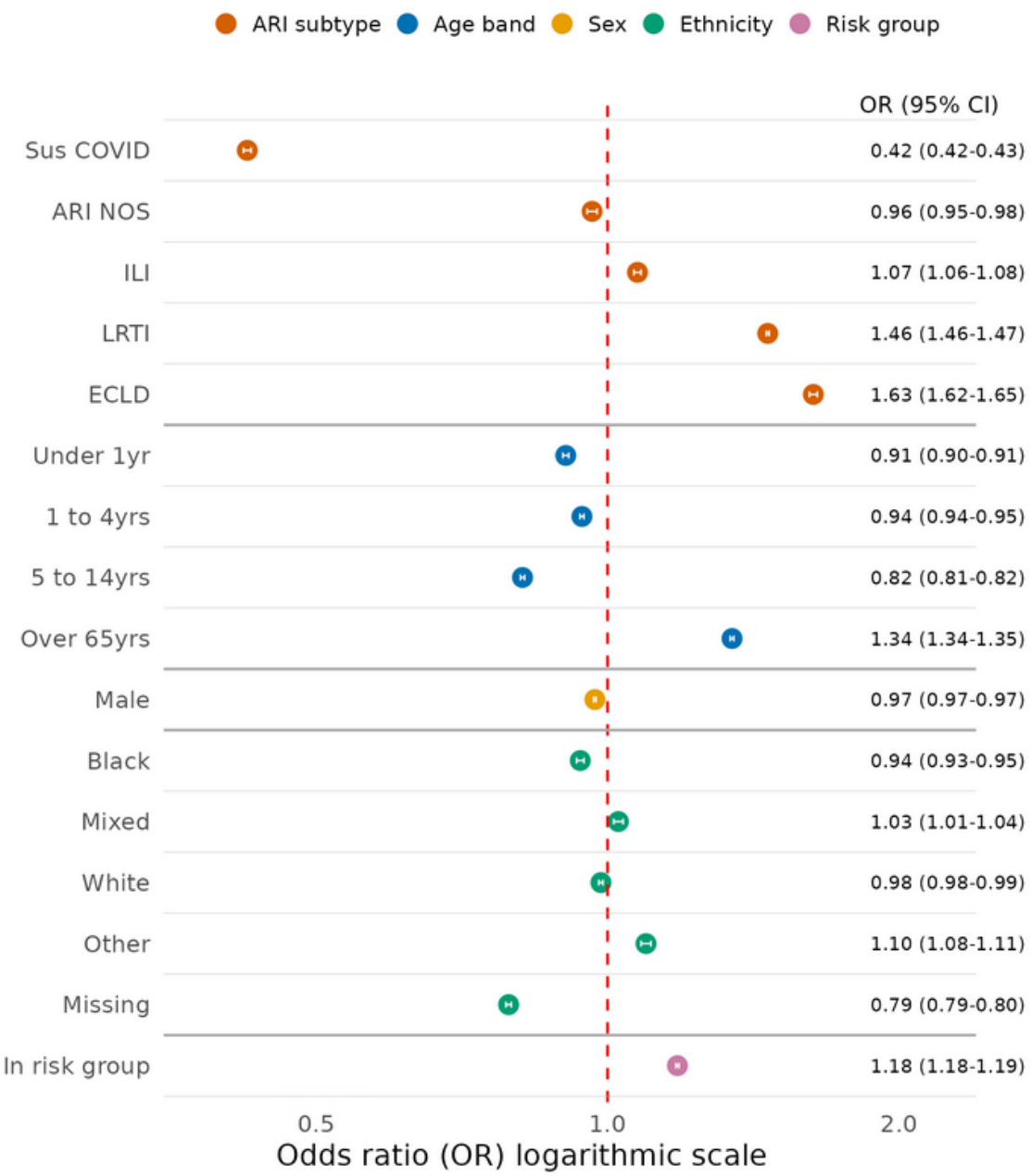


Figure 4.3: Adjusted odds ratios (aORs) and 95% confidence intervals (CIs) from a multivariable logistic regression model assessing the association between completeness of clinical sign recording and acute respiratory infection (ARI) subtype, age band, sex, ethnicity, and clinical risk group status. The dependent variable was the recording of any of the following signs: temperature, pulse rate, respiratory rate, oxygen saturation, or blood pressure. Reference categories: URTI (ARI subtype), age 15–64 years (age band), female (sex), Asian (ethnicity), and not in a clinical risk group (risk group status).

Among acute respiratory infection (ARI) subtypes, individuals with ECLD had the highest odds of having complete recording of clinical signs (aOR 1.63, 95% CI 1.62–1.65), while those with suspected COVID had the lowest (aOR 0.43, 95% CI 0.42–0.43), compared

with those with URTI (Figure 2). In the age group comparisons, people aged over 65 years had the highest odds (aOR 1.24, 95% CI 1.13–1.34), whereas those aged 5 to 14 years had the lowest (aOR 0.82, 95% CI 0.82–0.83), relative to adults aged 15–64 years.

Among those with recorded ethnicity, individuals of Other ethnicity had the highest odds of complete recording (aOR 1.10, 95% CI 1.09–1.12), while those of Black ethnicity had the lowest (aOR 0.94, 95% CI 0.93–0.95), compared to those of Asian ethnicity. People with missing ethnicity also had reduced odds (aOR 0.79, 95% CI 0.79–0.80). Males had slightly lower odds of complete recording than females (aOR 0.97, 95% CI 0.97–0.97). Individuals in a risk group had higher odds of complete recording than those not in a risk group (aOR 1.18, 95% CI 1.18–1.19).

All model variance inflation factors were less than 1.5 suggesting no significant issues with multicollinearity. Model calibration was formally assessed using the Hosmer–Lemeshow test, which indicated a statistically significant lack of fit (χ^2 = 2753.2, df = 8, p < 0.0001), as expected in large datasets. However, the decile-based calibration plot showed close agreement between observed and predicted probabilities, suggesting no meaningful miscalibration (Appendix A3.2).

4.4.4 Temporal dynamics

This section explores recording rates over time to assess the temporal dynamics of severity marker recording. The analysis first focuses on severe outcomes, followed by predictors of severe outcomes. For each group, a representative time series is presented as an illustrative example. For instance, the time series for fever is shown under the symptoms group. These examples are not necessarily reflective of all markers within their respective groups. Recording rates by study for all severity markers within each severity marker group can be seen in Appendix A3.1. Time series plots for a greater range of severity markers than are displayed in the main results can be found in Appendix A3.4.

Severe outcomes

The median weekly recording rates of all severe outcome groups are reported in Table 4.8. A narrative summary of trends over time and between periods and time series figures is provided here for hospitalisation and death-related outcomes.

Table 4.8: Median weekly recording rates by study period for severe outcome groups

Group	All	Period 1	Period 2	Period 3	Period 4
Any	8.52% (6.57–10.03)	5.68% (4.74–6.69)	8.83% (8.31–9.90)	10.73% (9.64–12.30)	10.26% (9.63–11.46)
Hospital	8.10% (6.09–9.58)	5.27% (4.29–6.25)	8.39% (7.85–9.45)	10.08% (9.10–11.42)	9.79% (9.10–10.84)
Died	0.60% (0.52–0.69)	0.52% (0.46–0.59)	0.63% (0.58–0.71)	0.79% (0.64–1.45)	0.64% (0.56–0.84)
Complication	0.21% (0.04–0.33)	0.03% (0.02–0.04)	0.24% (0.09–0.34)	0.34% (0.27–0.39)	0.35% (0.32–0.38)
ICU	0.01% (0.00–0.01)	0.00% (0.00–0.01)	0.01% (0.00–0.01)	0.02% (0.01–0.03)	0.01% (0.01–0.02)

Note: ICU: Intensive Care Unit. Period 1: Early pre-pandemic, Period 2: Late pre-pandemic, Period 3: COVID-19 Pandemic, Period 4: Post-pandemic.

Hospital: The overall median weekly recording rate for hospital outcomes (attendance and admission) was 8.10% (IQR: 6.09–9.58). Within this group, attendance had the highest median weekly rate (7.13%, IQR: 5.37–8.53), followed by hospital admission (1.71%, IQR: 1.13–2.02) and any death (0.60%, IQR: 0.52–0.69). The recording rate of any hospital event (attendance or admission) saw a 1.59-fold increase from Period 1 to Period 2. Time series visualisations show this to be a steady, largely linear increase in both attendance and admission, although there was a small stepwise rise in admissions around 2014 (Figure 4.4).

A further 1.20-fold increase in recording rates occurred between Period 2 and Period 3 (COVID-19 pandemic), with two clear peaks in admissions corresponding to the first and

second SARS-CoV-2 waves in the UK, one in the first half of 2020 and another at the end of 2020 and beginning of 2021. To a lesser extent these peaks are also visible in attendance.

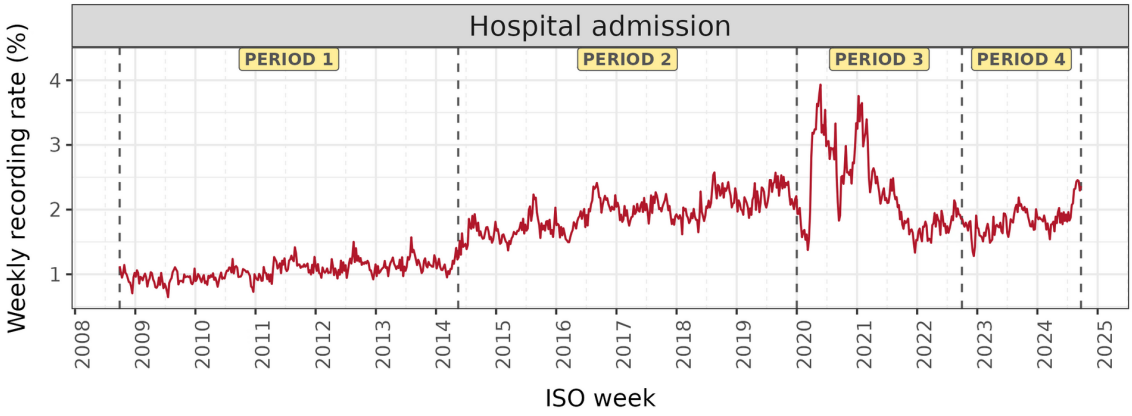


Figure 4.4: Time series of weekly hospital admission recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO: International Organization for Standardization.

Death: The overall median weekly recording rate for death within 56 days was 0.60% (IQR: 0.52–0.69). Between Period 1 and Period 2, the rate increased modestly by 1.20-fold, driven primarily by a subtle but steady upward trend observed across both periods. Additionally, there was a 1.25-fold increase from Period 2 to Period 3 (COVID-19 pandemic) (Figure 4.5). Despite this modest increase, two large spikes in deaths were observed during the first and second COVID-19 waves, with a maximum weekly rate of 5.44% occurring during the first wave, compared to maximum rate of 1.11% outside the pandemic period. In Period 4, recording rates returned to pre-pandemic levels.

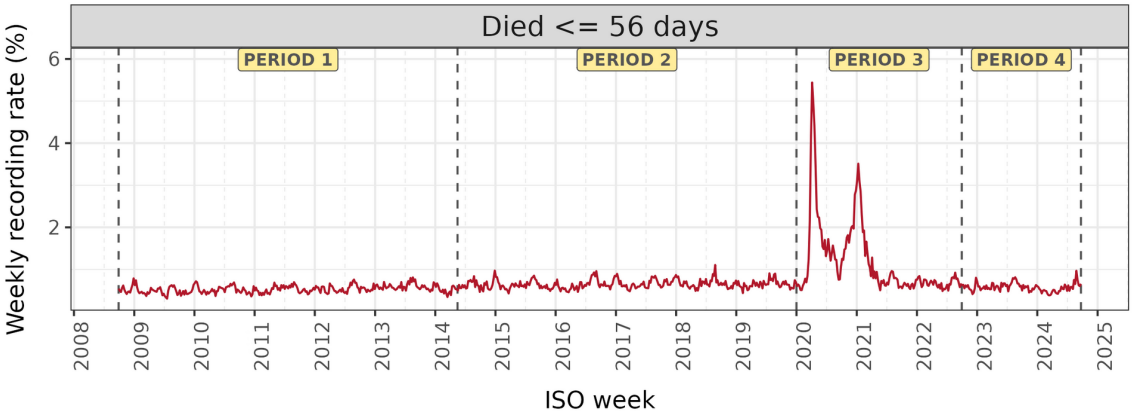


Figure 4.5: Time series of weekly death recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO: International Organization for Standardization.

Table 4.9: Median weekly recording rates by study period for predictors of severe outcomes

Group	All	Period 1	Period 2	Period 3	Period 4
Symptom	4.05 (2.49–5.25)	2.30 (2.06–2.53)	4.31 (3.64–4.85)	7.08 (5.02–8.37)	5.70 (5.21–6.15)
Health seeking	1.15 (0.09–4.63)	0.11 (0.06–0.10)	1.41 (0.78–1.98)	6.18 (4.92–7.44)	4.88 (4.69–5.08)
Absenteeism	3.55 (2.61–4.32)	2.43 (1.94–2.73)	3.91 (3.36–4.15)	5.13 (4.01–5.78)	4.41 (3.88–4.73)
Sign	31.90 (18.75–51.72)	19.00 (16.13–20.96)	47.30 (38.13–56.16)	29.16 (20.53–32.57)	55.53 (51.94–59.84)
Score	0.01 (0.00–0.05)	0.00 (0.00–0.00)	0.01 (0.01–0.02)	0.08 (0.04–0.12)	0.27 (0.24–0.30)
Investigation	7.34 (6.35–8.38)	6.25 (5.78–6.86)	8.03 (7.38–8.73)	7.57 (6.30–8.67)	8.28 (7.67–9.10)
Prescription	62.76 (59.84–64.78)	63.61 (62.30–65.32)	61.59 (59.56–63.66)	56.87 (54.16–60.21)	65.92 (64.72–67.28)
Hospital	0.13 (0.09–0.17)	0.08 (0.06–0.09)	0.14 (0.12–0.17)	0.18 (0.14–0.22)	0.17 (0.16–0.19)

Note: Values are median weekly recording rates with 95% confidence intervals. Subgroup names reflect primary care data across periods.

Symptoms: The overall median weekly recording rate for symptoms (any symptom) was 4.05% (IQR: 2.49–5.25). Within this group, dyspnoea (1.62%, IQR: 1.28–2.72) had the highest median weekly rate, followed by fever (1.45%, IQR: 0.68–1.94) and malaise (0.86%, IQR: 0.51–1.07). There was a 1.90-fold increase in the recording of any symptoms comparing Period 1 to Period 2. Time series figures for fever show this change occurring gradually from around 2013. A further 1.39-fold increase was observed from Period 2 to Period 3 (COVID-19 pandemic). Spikes in malaise recording are visible during both the first and second COVID-19 waves, while for fever the spike was more substantial during the first wave (Figure 4.6). Additionally, a noticeable spike in fever and malaise recording is seen in July 2009, coinciding with the first wave of the H1N1 swine flu pandemic.

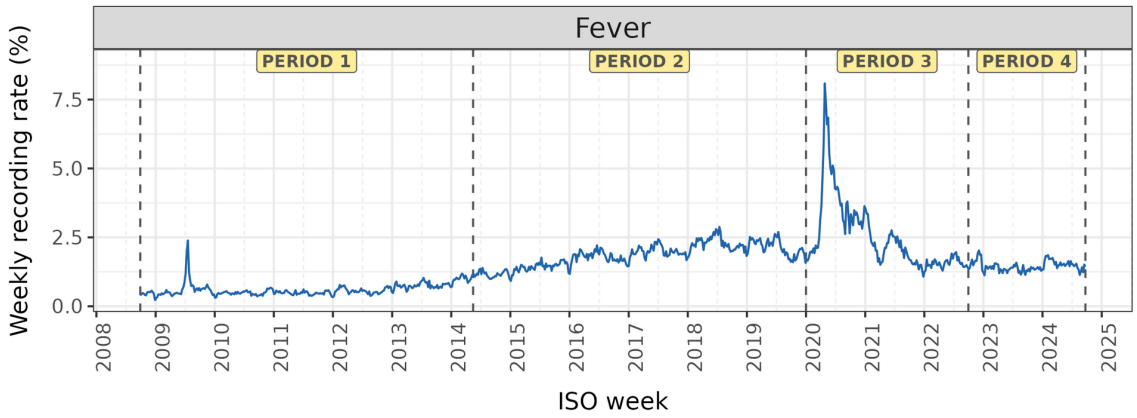


Figure 4.6: Time series of weekly fever (symptom) recording rates within 14 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO: International Organization for Standardization.

Absenteeism: Work absence was the only severity marker within this group, with a median weekly recording rate of 3.55% (IQR: 2.61–4.32). There was a 1.60-fold increase in rates between Period 1 and Period 2. The time series figure shows an upward trend between 2013 and 2015, and a more subtle increase between 2015 and 2019 (Figure 4.7). Additionally, there was a 1.3-fold increase between Periods 2 and 3, and 2 spikes during Period 3 consistent with COVID-19 waves. Absenteeism was highly seasonal with peaks occurring in early January each year. There also appears to be a sharp decline just prior to

the spikes, likely related to Christmas holidays. Recording rates in Period 4 (4.30, IQR: 3.88-5.78) are slightly above pre-pandemic levels (3.72, IQR: 3.36-4.15).

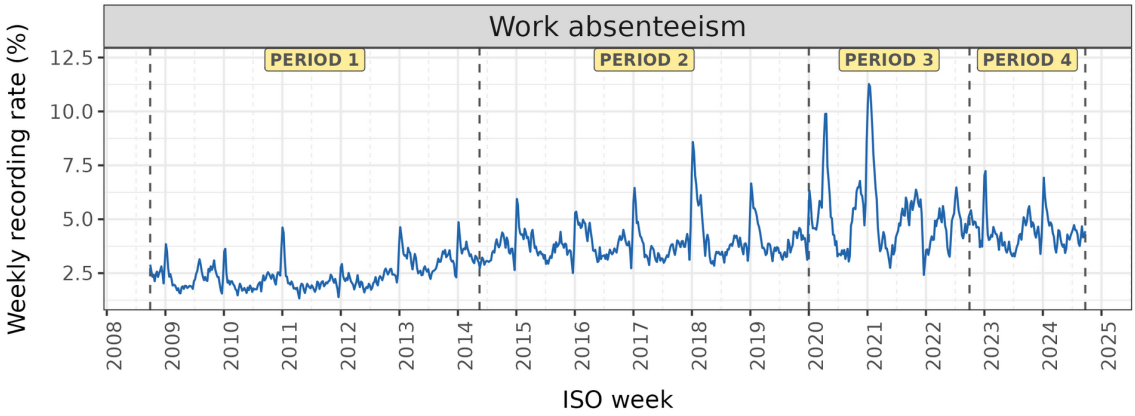


Figure 4.7: Time series of weekly work absenteeism recording rates within 14 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO: International Organization for Standardization.

Signs: The median weekly recording rate for any sign was 31.95% (IQR: 18.76–51.81). Temperature (18.72%, IQR: 6.02–37.91), pulse rate (14.00%, IQR: 5.21–32.34), and blood pressure (13.61%, IQR: 11.56–17.06) had the highest recording rates. There was a substantial increase in overall recording between Period 1 and Period 2 (fold-change 2.77). Time series plots for oxygen saturation show a substantial, initially non-linear but monotonic increase across this period, with a fold-change of 13.33.

A dramatic and sharp decline in the recording of signs (fold-change 0.538) occurred at the onset of the COVID-19 pandemic (Period 3). This was followed by a return to pre-pandemic recording rates over Periods 3 and 4. The median weekly recording rate for any sign in the post-pandemic period (Period 4: 55.45%, IQR: 52.03–59.91) was higher than in the pre-pandemic period (Period 2: 48.58%, IQR: 38.24–56.23). Seasonal patterns were also evident, with narrow peaks in recording of oxygen saturation occurring in early January (Figure 4.8).

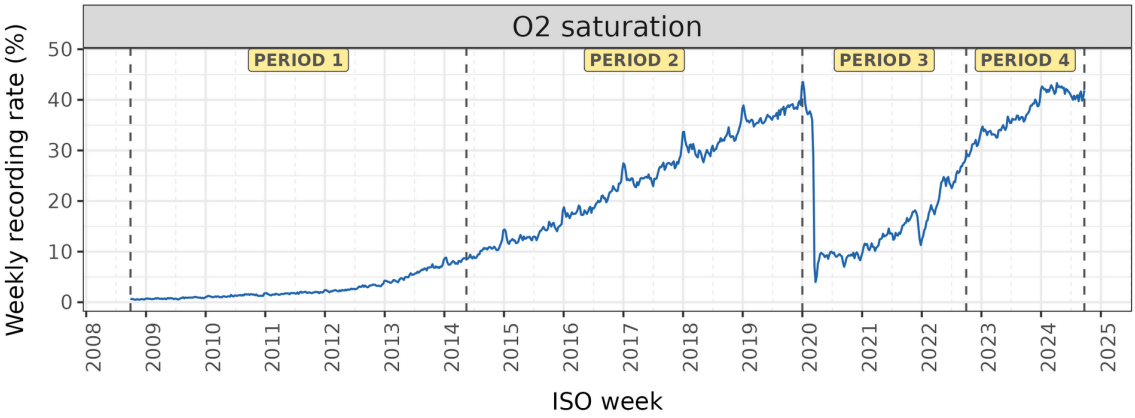


Figure 4.8: Time series of weekly oxygen saturation recording rates within 14 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO: International Organization for Standardization.

Prescriptions: The median weekly recording rate for any prescriptions was by far the highest among all severity marker groups (62.76%, IQR 59.84–64.78). Amoxicillin had the highest median weekly recording rate (34.35%, IQR: 31.86–37.08), followed by penicillin V (11.50%, IQR: 9.27–13.31) and macrolides (9.17%, IQR: 7.90–10.17). While overall prescription recording was relatively stable, notable trends and patterns were evident for individual medications.

Amoxicillin prescribing declined slightly and gradually from around 2014 until just before the pandemic (fold-change 0.91) (Figure 4.9- top panel). More substantial declines were observed for other antibiotics, for example, cephalosporins saw a 0.29-fold change from Period 1 to Period 2 (Figure 4.9- middle panel). Antibiotic prescribing further declined during the COVID-19 pandemic, with amoxicillin showing a fold-change of 0.77 between Periods 2 and 3. A noticeable dip in prescribing of amoxicillin, and to a lesser extent, cephalosporins, was also seen during the H1N1 swine flu pandemic in July 2009.

The most commonly recorded antiviral prescription was oseltamivir, with an overall median weekly rate of 0.01% (IQR: 0.00–0.02). Although this rate is low compared to antibiotics, oseltamivir prescribing showed substantial spikes, most notably during the H1N1 pandemic, reaching 1.92% in July 2009 (Figure 4.9- bottom panel). Smaller sea-

sonal spikes were also observed during certain winters, such as in 2011 and 2018.

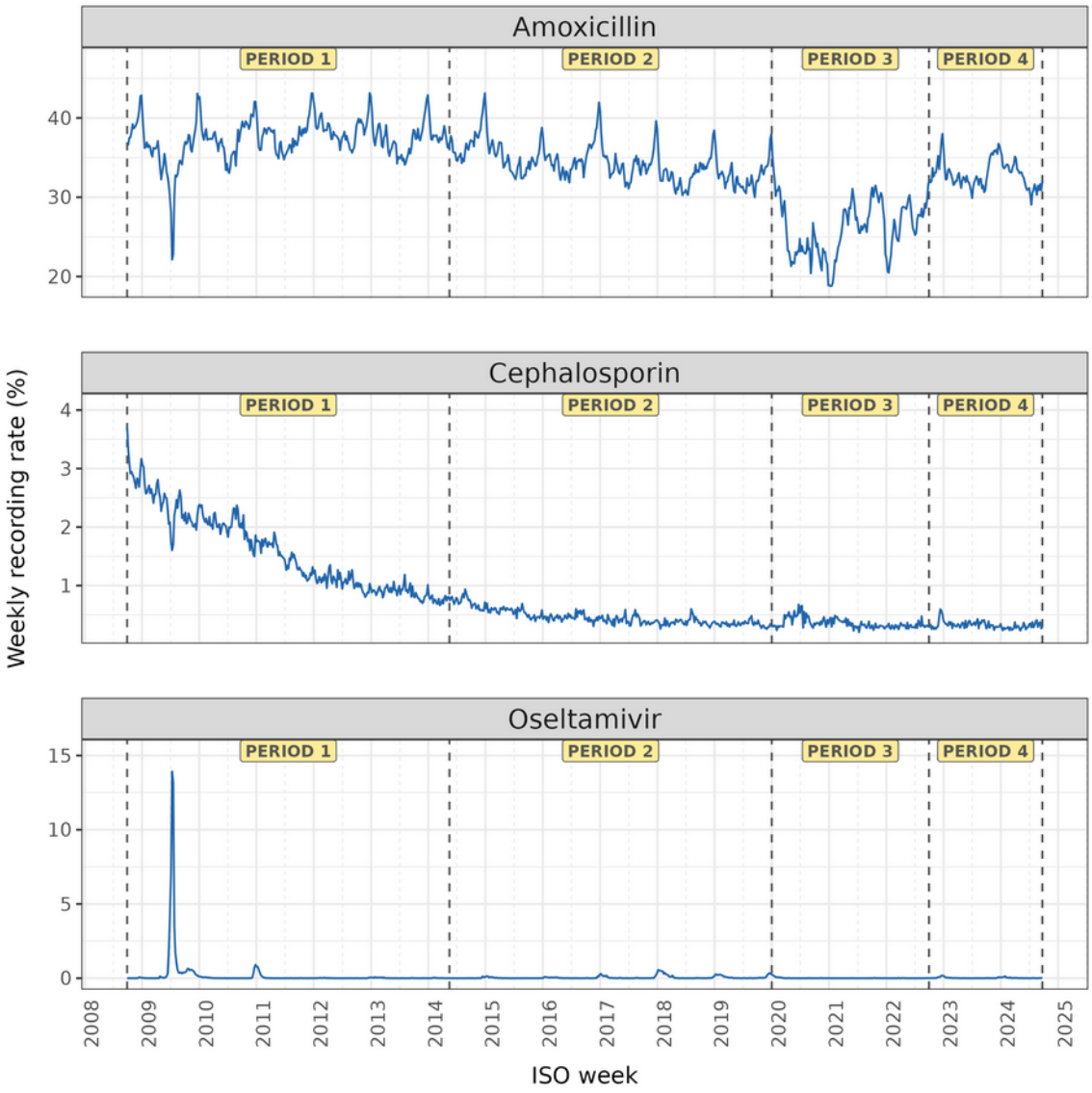


Figure 4.9: Time series of Amoxicillin (top panel), Cephalosporin (middle panel), oseltamivir (bottom panel) recording rates within 14 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO: International Organization for Standardization..

4.5 DISCUSSION

4.5.1 Summary of main findings

To assess the quality of severity marker data, I conducted a comprehensive evaluation of the completeness and temporal dynamics of severity marker recording across 25.7 million

episodes of ARI in 8.7 million individuals. I found that completeness varies substantially, not only by the specific severity marker but also by ARI subtype, age group, and risk status. While completeness cannot always be directly measured from recording rates, contextual information can help infer whether certain markers are structurally complete or incomplete. Recording rates vary over time, influenced by long-term secular trends, as well as abrupt and seasonal changes. In addition, pandemics have been shown to significantly affect the temporal dynamics of severity marker recording.

4.5.2 Completeness

This work highlights that recording of severity markers is incomplete, which is consistent with other studies of primary care data showing incomplete recording of characteristics of acute infections [109, 110]. In my study, the median weekly recording rate for fever was 1.6% in Period 4, which is almost certainly a significant underestimation of the true number of patients reporting fever. This is because fever is very commonly experienced in individuals with an ARI, especially in children.

The overall completeness of recording of any sign was 35.1%. However, completeness varied substantially by ARI subtype and age, indicating that clinicians may selectively record clinical signs in those at higher risk of adverse outcomes or those presenting with more severe disease. This phenomenon has been demonstrated in other studies [111]. An exception to this pattern was suspected COVID-19, which had the lowest expected recording rates of all ARI subtypes compared to URTI (OR 0.43, 95% CI: 0.42–0.43). This likely reflects reduced opportunities to measure clinical signs due to the decline in face-to-face consultations during the COVID-19 pandemic [112].

Differences by sex and ethnicity were small in magnitude, though statistically significant due to the large sample size. For example, the odds ratios for male sex and most ethnic groups (except missing ethnicity) were close to unity, suggesting limited practical impact on completeness. Although, these effect sizes are small it is important to be aware that variability in recording by demographics, including those not considered in this analysis,

could result in biased and inequitable severity indicators. For example, severity indicators in certain populations that are less sensitive to true changes in severity and therefore less effective.

Clinical scores were exceedingly rarely recorded (scores 0.01% of episodes overall and <0.3% of episodes in Period 4). The same applies for antiviral prescriptions (excluding oseltamivir), which in most instances were recorded in <0.01% of episodes. This very likely would limit the value of these individual severity markers for severity indicator development.

A more complete discussion on temporal dynamics is provided below, but first and foremost the completeness of data recording appears to have increased substantially over time. For example, in period 1 of the study, signs were recorded in only 19% of ARI episodes whereas by period 4 this figure reached 56%. These positive trends are universal across the severity marker groups, with the exception of prescriptions. This implies that data quality (with respect to completeness) has substantially improved across the study period. It also implies that the structural mechanisms that drive complete prescription recording have been embedded in the systems for some time. These improvements in data quality strengthen the justification for undertaking the work in this DPhil.

4.5.3 Temporal dynamics

This analysis has highlighted several patterns in recording rates over time, including long-term secular trends as seen in hospital admission and cephalosporin prescribing (Figure 4.8 and 4.9), sudden changes such as those following the onset of the COVID-19 pandemic, seen in most indicators, and seasonal variation as observed in work absenteeism and amoxicillin prescribing (Figures 4.7 and 4.9). The factors influencing these changes are likely heterogeneous, involving both severity and non-severity drivers of change. This study has not attempted to definitively establish the cause of changes over time, but it is possible to suggest external factors that may have contributed.

Severity-driven variation in recording rates

The pandemic led to clear spikes in the recording of hospitalisation and deaths (Figure 4.4 and 4.5), consistent with the timing of the first and second COVID-19 waves in England [113]. Recorded deaths peaked in April 2020 and again in January 2021. This is reassuring as it provides a quick validity check for these data, suggesting they are identifying known changes in disease severity. Other severe outcomes including ICU admission and sepsis also showed peaks in recording rates during the first and second waves of the pandemic. However, these time series were very noisy making identification of more subtle trends challenging.

The COVID-19 waves serve as useful reference points for evaluating severity markers. However, identifying appropriate reference points for severity outside the context of the pandemic is more challenging. Furthermore, due to non-pharmacological interventions, the circulation of other respiratory pathogens declined substantially during the pandemic. As a result, recorded rates of death, hospitalisation, and other markers during this period would largely reflect the severity of suspected COVID-19 episodes, rather than all ARI. This complicates direct comparisons of recording rates between pandemic and non-pandemic periods. While true changes in severity could influence recording rates, it is essential to consider factors unrelated to the occurrence of the event under observation.

Other drivers in variation of recording rates

Long-term trends: Long-term trends may result from gradual shifts in recording rates as a consequence of slowly evolving coding behaviours, clinical guidelines or IT systems [99, 114]. For example, the reduction in recorded cephalosporin prescriptions likely reflects a true decrease in prescribing behaviour, driven by efforts to promote antimicrobial stewardship and the association between cephalosporins and *Clostridium difficile* infection [115, 116] (Figure 4.9). Additionally, the increase in sepsis recording around 2015 to 2016 coincides with a high-profile national campaign highlighting the time-critical need for prompt sepsis treatment, as well as the publication of NICE guidelines on the recognition and management of sepsis [117, 118].

Abrupt changes: Sudden and unexpected events, most notably the onset of the COVID-19 pandemic, also had a considerable effect on recording rates. For example, recording rates for clinical signs saw a sharp decline following the start of the pandemic (0.54-fold change from Period 2 to Period 3). This was likely a consequence of the reduction in face-to-face consultations, driven by rapidly evolving national guidance during the pandemic [112]. Without patients physically present, clinicians had limited ability to measure and record clinical signs.

There was also a significant reduction in recording of antibiotic prescriptions such as amoxicillin at the onset of the pandemic. However, this decline is less likely to be explained by reduced face-to-face contact, as prescriptions can still be issued during telephone consultations. A more likely explanation is the high probability that individuals presenting with ARI symptoms during the pandemic had COVID-19, for which antibiotics were not indicated. Some studies have shown that antibiotic prescribing for URTI increased during the onset of the pandemic [119]. I did not evaluate changes in recording rates over time by ARI subtype to allow direct comparison of these findings.

In contrast recording of symptoms which increased after the onset of the pandemic, this again may reflect the fact that clinicians relied more on the patient's history than on physical examination to determine the severity individuals during remote consultations.

Seasonality: Several severity markers exhibited seasonal patterns. For example, work absenteeism typically showed sharp dips during the Christmas holidays, followed by notable spikes in early January. These fluctuations may limit the reliability of such markers for detecting changes in illness severity during holiday periods. Amoxicillin prescribing also demonstrated clear seasonality, peaking in the winter months, consistent with findings from other studies [120, 121].

4.5.4 Implications for DPhil

The next chapter of my DPhil will aim to explore whether the predictors of severe outcomes could be used to develop timely severity indicators for surveillance. This will involve construction of statistical models to assess the relationship between the predictors and the severe outcomes. Work in this chapter has helped determine aspects of the modeling framework to be used in the next stage. Here, I list the implications of the findings for the remaining parts of the DPhil.

Models will be stratified by ARI subtype and age. This will help to ensure severity markers are most suited to the relevant age bands and ARI subtypes. Also, there is a trade off here as it will reduce the sample size in any given subgroup, the number of episodes in the analysis will still be large. Separate models by risk group status, sex and ethnic group will not be performed as differences were small.

Predictors of severity that are very rarely recorded will no longer be considered. Specifically, antiviral and clinical score recording will be removed from subsequent analysis as recording rates were consistently less than 0.5% which limit the value of these markers.

I will consider the pandemic and non-pandemic periods separately. The striking changes in almost all severity marker recording rates that occurred during the pandemic highlight the huge disruption that occurred to the health system. Separation of study periods into the pandemic and non-pandemic period given these striking differences is logical. A number of non-random differences between these periods are likely to have affected recording of both severe outcomes and predictors.

Firstly, during the pandemic, SARS-CoV-2 was likely to have been the dominant circulating pathogen regardless of the presenting ARI subtype. Therefore during the pandemic severe outcomes and predictors are more a reflection of SARS-CoV-2 infection severity. Outside this time, circulating pathogens are more heterogeneous and include influenza, RSV and other seasonal viruses. Also, as discussed previously, healthcare delivery was massively disrupted, many more remote primary care consultations were taking place previously

which likely affected the documentation of in particular clinical signs and scores.

I will use a composite outcome. The temporal signal in rare outcomes is noisy, which makes detection of changes in severity challenging. Without a clear signal in the outcome, statistical models are not likely to perform well as the relationship will be masked by random variation. This would reduce the chances of detecting meaningful indicators for surveillance. Additionally, from a public health perspective a composite outcome including hospital attendance and admission is likely to more accurately reflect pressures in the health system compared to deaths alone. Separate indicators would perhaps be ideal, but to limit the complexity of the next stages of the DPhil a single composite outcome will be used.

I will focus on more recent data. The long term trends in recording identified could introduce temporal bias into any analysis where the apparent risk of an outcome increases over time due to factors unrelated to a true rise in that outcome. Ideally, I would only use post-pandemic data to construct models as these are most relevant to surveillance today. However this significantly limits the data available for the analysis. As a compromise, I will exclude data from Period 1 from the subsequent analysis.

4.5.5 Broader relevance

There are several important implications of this work outside the context of this DPhil. Firstly, we have seen the huge impact a pandemic can have on data quality in a primary care CMR. Therefore, a pandemic could result in reduced reliability of a surveillance system at a time when public health authorities are most reliant on them. It is essential that we build resilience into our CMR-based surveillance systems such that they are robust during an emerging threat, such as a future pandemic [122, 123].

Secondly, primary care CMR data is widely used for various secondary research purposes. However, this study has shown that the quality of such data can vary significantly over time due to a range of external factors. These temporal changes may introduce unexpected

biases and affect the interpretation of results, particularly in longitudinal studies conducted over extended periods [94]. These are important considerations for future users of CMR data [98, 100].

Furthermore, structurally complete data can be highly valuable for researchers. Knowing that death, prescriptions, and work absenteeism are likely to be consistently and reliably recorded provides confidence that these variables serve as reasonable proxies for the true underlying events. As a result, they can be used in studies with greater confidence than other, less complete and more unstable measures. Moreover, these variables are important in their own right, as mortality, antibiotic prescribing, and the workforce impact of disease are all valuable areas of public health research. Nonetheless, assumptions of structural completeness should be made cautiously, as consistency may vary by setting or over time.

Finally, several key policy documents, including the recent Sudlow Review commissioned by Health Data Research UK, have highlighted the substantial potential of secondary use of CMR data to improve health systems and deliver broader public and societal benefits [67, 124]. Realising this important and ambitious goal will require sustained efforts to improve data quality, driven by both policy interventions and technological innovation. Improvements in data quality should also be as equitable across demographic groups as possible to reduce the risk of introducing bias into models developed directly from CMR data.

4.5.6 Limitations

One of the principal limitations of this study is the lack of a definitive link between ARI episodes and recorded severity markers. Cause of death is not typically available in the CMR data, which means that while a recorded death date confirms that a person has died, it does not establish whether the death was directly related to an ARI. A reasonable assumption is that the closer the event (death or other severity marker) occurs to the onset of the ARI episode the more likely it is related to the ARI episode. As a result, there is a trade-off between specificity and sensitivity when choosing severity marker timeframes.

For example, mortality occurring shortly after the onset of infection is more specific to recent illness but may miss delayed ARI-related deaths, whereas later mortality captures a greater number of deaths but may include more that are unrelated to the ARI episode.

Another limitation is the lack of access to free-text data, which is commonly used to record information in primary care [109, 111]. Currently, access to such data in the NHS is highly regulated due to governance, privacy and ethical concerns. Additionally, analysing this unstructured data requires significant technical expertise. In the future, large language models could play a valuable role in processing this data for surveillance purposes.

I have made inferences about potential drivers of changes in recording rates over time. While these assumptions may appear reasonable, further work would be needed to test them directly. For example, studies specifically assessing the impact of public health campaigns could help validate these interpretations, although such work falls outside the scope of this DPhil. While the exact causes of some temporal changes remain uncertain, I can be confident that many of these shifts are driven by factors unrelated to underlying disease severity. Knowing this is enough information to ensure I proceed with caution.

Finally, further work is needed to validate these indicators, including comparison against established reference standards. For example, ARI-related deaths recorded in primary care could be compared with linked, ONS death data, which includes the certified cause of death. Similarly, ARI-associated hospitalisations could be assessed using the HES dataset. Both ONS and HES data could, in principle, be linked to primary care records using the NHS number, although access to such linkages is currently highly restricted [90, 125].

4.5.7 Summary

This chapter builds on previous work by evaluating the quality of the severity markers identified in Chapter 3. First, the completeness of clinical signs varied according to perceived severity and severity risk, suggesting that sign-based markers may overestimate true severity because they are more frequently recorded in episodes clinicians judge to be

higher risk. Second, changes in recording rates over time appear partly driven by genuine shifts in severity, as shown by the rise in recorded severe outcomes during the pandemic period. This is reassuring, however recording is also shaped by external factors unrelated to severity—including changes in clinician behaviour, health policy, and seasonal pressures, which can obscure interpretation of temporal trends.

Most critically, the findings show that data quality deteriorated sharply during the pandemic, reducing the effectiveness of surveillance when robust, timely information is most needed. Despite these challenges, completeness has improved substantially over the 16-year period. This chapter also highlights key considerations for model development, including the value of composite outcomes, focusing on more recent years, and accounting for ARI subtype and age. Together, these insights inform the next stage of this DPhil, where I now go on to measure the association between predictors and severe outcomes in the following chapter.

Chapter 5

Evaluating candidate severity indicators for use in prospective ARI surveillance

5.1 INTRODUCTION

Previous work for this DPhil involved defining and assessing the data quality of a number of severity markers for ARIs. In this chapter, I build an algorithm to further evaluate the suitability of these severity markers for use in prospective respiratory surveillance of ARIs. The algorithm takes into account lessons from the previous chapter, including separation of the analysis into a non-pandemic and pandemic period, running the analysis stratified by age band, ARI subtype and study period (non-pandemic versus pandemic) and use of a composite outcome. The analysis for each stratum will be in two parts:

1. Individual-level analysis
2. Weekly aggregate analysis

5.1.1 Rationale for methods used

Because predictors are recorded at or around the time of the ARI consultation, they are available in the CMR sooner than severe outcomes. If they can be shown to be reliably associated with subsequent severe outcomes, then timely severity indicators could be developed using those predictors most strongly linked to severe outcomes.

From a surveillance perspective, an aggregate analysis is essential because it best reflects the intended use case. In this context, 'aggregate' refers to grouping data by week to

calculate the proportion of severe cases over time (Table 1.4). The intention in monitoring these trends over time is to identify increases in severity that may indicate a potential public health threat thereby supporting public health decision making. The overall purpose of the aggregate level analysis is to establish whether temporal changes in the predictor are similar to temporal trends in the outcome. For example, do trends in proportion of ARI episodes with abnormal clinical signs follow trends in the proportion of ARI episodes with the composite outcome recorded?

An aggregate analysis is essential although not sufficient for evaluation of candidate severity markers. This is because of the risk of the ecological fallacy, where an aggregate level analysis identifies spurious relationships in the data that do not exist at the individual level [126]. This typically occurs as a consequence of confounding. For example, more people wear a coat in winter and more people die, but death is not associated with wearing a coat. For this reason, I also undertake an individual-level analysis to further support the case for a plausible association between the severity marker and the outcome. Whilst such an approach cannot totally eliminate the possibility of spurious relationships being identified, it provides a more robust methodological framework for identifying potential severity indicators and reduces the risk of the ecological fallacy.

Finally, I will use the results of the individual and aggregate analyses to rank the severity indicators within each stratum in both non-pandemic and pandemic periods, aiming to identify aggregate-level associations that are also clinically and epidemiologically plausible. The top ranked candidate severity markers can then be included in a future prospective pilot. The best approach for further prospective evaluation and subsequent implementation within a real world surveillance system is a principal subject of discussion in the final chapter of this DPhil.

5.2 CHAPTER AIMS AND OBJECTIVES

5.2.1 Aims

For both the non-pandemic and pandemic period, measure the relationship between severity markers and the composite outcome of complications, hospitalisation or death by 56 days, across strata of ARI subtype and age band, at both the individual-level and weekly aggregate-level.

5.2.2 Objectives

1. **Individual-level analysis:** Estimate the ORs and 95% CIs to quantify the association between each severity marker and the outcome, for each combination of ARI subtype and age band.

2. **Weekly aggregate analysis:**

   a: Cross correlation. Estimate the correlation coefficient (CC) and 95% CIs to quantify the association between weekly aggregate predictors of severity and the weekly aggregate outcome for each combination of ARI subtype and age band.

   b: signal-to-noise ratio (SNR). Estimate the SNR of each aggregate predictor to assess how much true aggregate-level variation is distinguishable from random noise.

3. **Indicator prioritisation:** Based on a combination of the individual and weekly aggregate analyses, rank severity markers for each combination of ARI subtype and age band.

5.3 METHODS

5.3.1 Study population

The data sources, study participants, and inclusion and exclusion criteria used in this chapter are described in detail in Section 4.3. Participants with episodes of recorded ARI of any subtype and age were included in the analysis.

This chapter excludes episodes of ARI that occurred in Study Period 1 (the early pre-pandemic period Figure 4.1) to ensure that models are more relevant to the present day context, as discussed in Section 4.5.4. Therefore, all episodes occurring between ISO week 20 of 2014 and ISO week 39 of 2024 were included.

5.3.2 Analysis overview

The analysis pipeline used for this evaluation includes five individual steps and generates four evaluation metrics (Figure 5.1).

Step 1 describes the definition of the outcome. **Step 2** explains the definition of the analysis strata. **Step 3** describes how the severity markers are configured for the analysis. **Step 4a** is the individual-level analysis, which produces the first metric (OR), whereas **Step 4b** is the weekly aggregate analysis, which produces the second and third metric (CC, SNR). **Step 5** describes how severity markers are ranked based on a combination of the ORs and CCs. Metric 4 is a geometric mean of the OR and CC. The remainder of the methods details each of the steps and metrics highlighted in Figure 5.1.

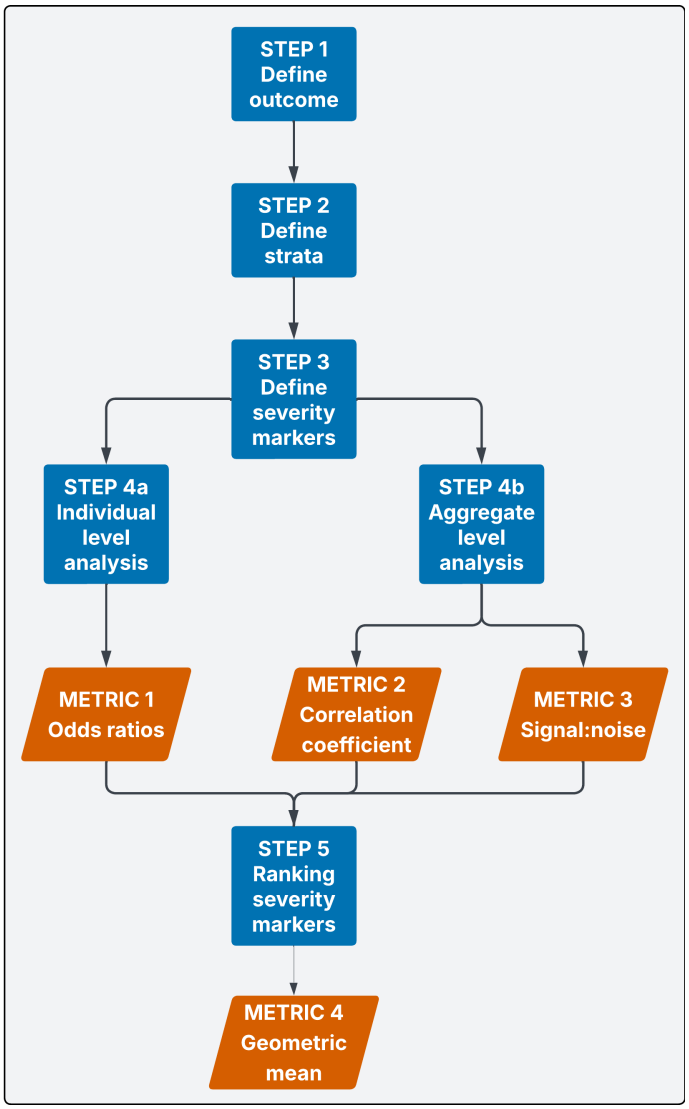


Figure 5.1: Overview of the analysis workflow. Blue boxes indicate **steps**: defining the outcome, strata, and severity markers (Steps 1–3), conducting individual- and aggregate-level analyses (Steps 4a–4b), and combining results (Step 5). Orange boxes indicate **metrics**: **Odds ratios** (Metric 1), **Correlation coefficient** (Metric 2), **Signal-to-noise ratios** (Metric 3), and the final **geometric mean** (Metric 4).

5.3.3 Step 1: Defining the outcome

The binary outcome for this analysis was a composite of hospital attendance or admission, ICU admission, complication (sepsis or acute respiratory failure- derived from discharge summaries), or death within 56 days of the index ARI episode. The rationale for this is set out in Section 4.5.4. These outcomes are treated as 'all-cause' because, when using the primary care CMR, it is not possible to definitively link an ARI episode to a specific outcome. The primary care informatics review within the systematic review

chapter attempted to increase the specificity of outcomes by focusing them around ARI. For example, respiratory failure was included as a logical complication of ARI. However, for deaths and hospitalisations, there are no data that can systematically confirm that the event was ARI-related.

The temporal proximity of the outcome to the ARI index case affects the likelihood that the case and outcome are linked. For example, if an index case of ARI and a hospitalisation are recorded in the CMR on the same day, it is reasonable to assume they are associated. The probability of a causal link decreases as the time between the ARI episode and the outcome increases. In this analysis, a 56-day window was chosen to capture delayed outcomes, as recommended by the WHO [127]. The trade-off is the potential inclusion of outcomes unrelated to the index ARI episode.

Despite these limitations, all-cause mortality may be the preferred endpoint in studies because it is unambiguous, captures the indirect and direct effects of an ARI, and allows comparability across settings [128]. For example, in an elderly person with cardiovascular risk factors, an ARI may precipitate a stroke, myocardial infarction, or pulmonary embolism leading to hospitalisation or death [129, 130]. Such outcomes might not be identified if only disease-specific mortality were used.

5.3.4 Step 2: Defining the strata

Due to significant disruption in the operation of health services and the impact that this had on the quality of data during the SARS-CoV-2 pandemic, this period was assessed separately. The pandemic analysis therefore included only ARI episodes occurring in Period 3 of the study (Figure 4.1).

Non-pandemic analysis

During the non-pandemic period, the range of circulating viruses would have been heterogeneous and included all seasonal winter viruses including, for example, influenza, RSV and rhinovirus. As the range of clinical syndromes caused by these viruses is broad it

makes sense to analyse this period by individual ARI subtype. Therefore, for the non-pandemic analysis, the following ARI subtypes were considered: URTI, ILI, and LRTI combined with ECLD (LRTI-ECLD). LRTI and ECLD were combined to limit the total number of strata and thereby reduce analytical complexity, while ARI NOS was excluded as reporting severity for an ill-defined group of syndromes has limited practical value. To further simplify reporting, the five original age bands were merged into three: <15 years (combining <1 year, 1–4 years, and 5–14 years), 15–64 years, and ≥65 years. ***There were therefore 9 strata in the non-pandemic analysis: three age bands by three ARI subtypes*** (Table 5.1).

Table 5.1: Non-pandemic analysis strata (9 strata)

Stratum	ARI subtype	Age band
1	URTI	<15 years
2	ILI	<15 years
3	LRTI OR ECLD	<15 years
4	URTI	15–64 years
5	ILI	15–64 years
6	LRTI OR ECLD	15–64 years
7	URTI	≥65 years
8	ILI	≥65 years
9	LRTI OR ECLD	≥65 years

Pandemic analysis

In contrast, during the pandemic period, all ARI subtypes were classified as suspected COVID-19, whereas in early sections of the DPhil suspected-COVID-19 was considered as a separate indicator (Figure 2.1) , since SARS-CoV-2 was the predominant circulating virus in primary care, regardless of clinical presentation [17]. The overall ARI indicator thus provides a proxy for a SARS-CoV-2-specific indicator. In addition, because severe outcomes were rare in children, only two age groups were considered in this analysis: 15–64 years and ≥65 years [131]. ***There were therefore 2 strata in the pandemic analysis: suspected COVID-19 in 15–64 year olds and suspected COVID-19 in those ≥65 years*** (Table 5.2).

Table 5.2: Pandemic analysis strata (2 strata)

Stratum	ARI subtype	Age band
1	Suspected COVID-19	15–64 years
2	Suspected COVID-19	≥65 years

5.3.5 Step 3: Defining severity markers

The final list of severity markers included in the analysis is presented in Table 5.3. These markers represent the timely predictors of severe outcomes identified in the systematic review, excluding scores, which were removed due to infrequent recording. As a reminder, 30 severity markers were identified in the systematic review and 36 code lists were developed. From the 36 code lists, I created a total of 43 binary severity 'flags' (Table 3.7), where a flag represents the presence or absence of a given severity marker recorded in the CMR. More than a single severity marker per predictor occurs, as a number of cut offs are used for numeric clinical signs and additionally I created a number of composite severity markers (Table 5.3).

Table 5.3: Severity markers by category and group

Severity marker	Details
Symptoms: 6 markers	
Shortness of breath	
Haemoptysis	
Fever	
Malaise / Anorexia / Fatigue	
Confusion	
Any symptom	Composite indicator (any symptom of the above five)
Health seeking behaviour: 3 markers	
Use of NHS direct (NHS 111)	
Ambulance encounter	
Hospital attendance advised	
Absenteeism: 1 marker	
Work absence	
Clinical signs (NEWS2/PEWS Flags): 18 markers	
Respiratory rate	Level 1 Flag; Level 2 Flag; Level 3 Flag
Oxygen saturation	Level 1 Flag; Level 2 Flag; Level 3 Flag
Systolic blood pressure	Level 1 Flag; Level 2 Flag; Level 3 Flag
Pulse rate	Level 1 Flag; Level 2 Flag; Level 3 Flag
Temperature	Level 1 Flag; Level 2 Flag; Level 3 Flag
Work of breathing	
Chest signs	
Cyanosis	
Composite sign flags	
Any-sign (Flag 1)	'True' if *any* sign met a Level 1 threshold
Any-sign (Flag 2)	'True' if *any* sign met a Level 2 threshold
Any-sign (Flag 3)	'True' if *any* sign met a Level 3 threshold
Subtotal (Composite signs): 3 markers	
Investigations: 3 markers	
White cell count	Abnormal: $>$ 12,000 or $<$ 4,000 cells/mm^3
C-reactive protein	Recorded
Chest radiography	Chest X-ray request recorded
Prescriptions: 9 markers	

Continued on next page

Table 5.3: Severity markers by category and group (continued)

Severity marker	Details
Amoxicillin	Antibiotic
Doxycycline	Antibiotic
Macrolide	Antibiotic
Co-amoxiclav	Antibiotic
Penicillin	Antibiotic
Cephalosporin	Antibiotic
Any antibiotic	Composite indicator (any of the above antibiotics; excludes antivirals)
Oseltamivir	Antiviral
Prednisolone	Steroid
Overall total: 43 severity markers	

Notes: SBP: systolic blood pressure; CRP: C-reactive protein. Level-specific sign flags follow National Early Warning Score 2 (NEWS2) (adults) and Paediatric Early Warning System (PEWS) (children); endpoints are inclusive. Work of breathing, chest signs, and cyanosis are recorded as yes/no (no levels). **Any symptom** is 'true' if any listed symptom is recorded. **Any antibiotic** is 'true' if any listed antibiotic prescription is recorded.

Included numeric severity markers were converted to binary flags using clinically informed cut offs. For clinical signs, cut offs followed those used in NEWS2 for adults and PEWS for children (Figure 5.2) [84, 132]. For ARI episodes in younger age groups, the PEWS was used, which defines cut offs for 0 to 11 months, 1 to 4 years, and 5 to 12 years. Accordingly, cut offs for children younger than 1 year used the PEWS 0 to 11 months thresholds; those for 1 to 4 year olds used the PEWS 1 to 4 years thresholds; and those for 5 to 14 year olds used the PEWS 5 to 12 years thresholds. As only 3 age bands are used in this analysis, flags were determined before the age bands were combined to ensure the accurate cut offs were used.

For each sign, I created three level-specific flags corresponding to the scoring bands (levels 1, 2, and 3) in the NEWS2 and PEWS scoring charts. A flag was marked 'true' when the recorded value fell within the range for that level, or when the criteria for a more severe flag were met. Otherwise, it was marked 'false' (Figure 5.2).

- *Example (adult respiratory rate, NEWS2):*
  - Flag 3 'True' (**most severe**) : ≤8 or ≥25 breaths per minute
  - Flag 2 'True': 21–24 breaths per minute (or when Flag 3 is 'True')
  - Flag 1 'True' (**least severe**): 9–11 breaths per minute (or when Flag 2 or Flag 3 are 'True')

Physiological parameter	Score 3	2	1	0	1	2	3
Respiration rate (per minute)	≤8		9–11	12–20		21–24	≥25
SpO_2 Scale 1 (%)	≤91	92–93	94–95	≥96			
SpO_2 Scale 2 (%)	≤83	84–85	86–87	88–92 ≥93 on air	93–94 on oxygen	95–96 on oxygen	≥97 on oxygen
Air or oxygen?		Oxygen		Air			
Systolic blood pressure (mmHg)	≤90	91–100	101–110	111–219			≥220
Pulse (per minute)	≤40		41–50	51–90	91–110	111–130	≥131
Consciousness				Alert			CVPU
Temperature (°C)	≤35.0		35.1–36.0	36.1–38.0	38.1–39.0	≥39.1	

Figure 5.2: NEWS2 physiological parameter scoring chart. *Reproduced from* the National Early Warning Score (NEWS) 2 chart, Royal College of Physicians (2017).

In addition to flags for each individual sign, I also defined an 'any-sign' flag for each severity level. This flag was set to 'true' if any one of the signs met the threshold for that level or for any more severe level. For example, the level-2 'any-sign' flag for adults was true if any of the following were recorded: respiratory rate 21–24 breaths per minute, oxygen saturation 92–93%, systolic blood pressure 91–100 mmHg, pulse 111–130 per minute, pulse 111–130 per minute, or temperature ≥39.1°C, or if a level-3 criterion was met for any sign. Tables 5.4 and 5.5 provide a detailed reference for individual and any-sign flag criteria.

Table 5.4: Severity marker cut offs (by age band)

Age band(s)	Flag 1	Flag 2	Flag 3
Systolic blood pressure (mmHg)			
< 1 yr	< 70 or > 90	< 60 or > 100	< 50 or > 110
1–4 yrs	< 80 or > 100	< 60 or > 120	< 50 or > 130
5–14 yrs	< 90 or > 110	< 80 or > 120	< 70 or > 130
≥15 yrs	< 111	< 101	< 91 or > 219
Pulse rate (beats/min)			
< 1 yr	< 110 or > 150	< 90 or > 170	< 80 or > 180
1–4 yrs	< 90 or > 140	< 70 or > 150	< 60 or > 170
5–14 yrs	< 80 or > 120	< 70 or > 140	< 60 or > 160
15–64 yrs	< 51 or > 90	> 110	< 41
65+ yrs	< 51 or > 90	> 110	< 41 or > 130
Respiratory rate (breaths/min)			
< 1 yr	< 30 or > 40	< 20 or > 60	< 10 or > 70
1–4 yrs	—	< 20 or > 50	< 10 or > 60
5–14 yrs	< 20 or > 25	< 15 or > 40	< 10 or > 50
≥15 yrs	< 12	> 20	< 9 or > 24
O_2 saturation (%)			
< 15 yrs	< 95	—	< 92
≥15 yrs	< 96	< 94	< 92
Temperature (°C)			
< 15 yrs	< 36.0 or > 38.0	—	—
≥15 yrs	< 36.1 or > 38.0	> 39.0	< 35.1
White cell count (10^9/L)			
All ages	< 4.0 or > 12	> 30	—
C-reactive protein (mg/L)			
All ages	> 20	> 100	—

Notes: cut offs used to convert numeric quantities into binary variables. Flag 1 represents the least severe level: scoring 1 in the NEWS2 or PEWS scoring system. Flag 3 represents the most severe scoring 3.

Table 5.5: Definition of composite severity flags

Composite Flag	Constituent flags
Any-sign (Flag 1)	RR Flag 1 **OR** O_2 sats Flag 1 **OR** SBP Flag 1 **OR** PR Flag 1 **OR** Temp Flag 1.
Any-sign (Flag 2)	RR Flag 2 **OR** O_2 sats Flag 2 **OR** SBP Flag 2 **OR** PR Flag 2 **OR** Temp Flag 2.
Any-sign (Flag 3)	RR Flag 3 **OR** O_2 sats Flag 3 **OR** SBP Flag 3 **OR** PR Flag 3 **OR** Temp Flag 3.
Any antibiotic	Amoxicillin **OR** Doxycycline **OR** Macrolide **OR** Co-amoxiclav **OR** Penicillin **OR** Cephalosporin.
Any symptom	Dyspnoea **OR** Fever **OR** Haemoptysis **OR** Malaise **OR** Confusion.

Notes: Definition of composite severity markers used. Where a given Flag is not used in a certain age group this is ignored. For example, there is no Oxygen saturation Flag 2 for <15 year olds, therefore this is ignored. SBP: systolic blood pressure; PR: pulse rate; RR: respiratory rate; O_2 sats: peripheral oxygen saturation; Temp: temperature; WBC: white blood cell count (10^9/L); CRP: C-reactive protein (mg/L).

A single cut off for white blood cell counts was used and based on the Systemic Inflammatory Response Syndrome (SIRS) sepsis definition (Table 5.3) [133]. For C-reactive protein (CRP), a cut off of 20 mg/L is used which aligns with the National Institute for Health and Care Excellence (NICE) clinical guidelines for management of suspected ARIs [134].

5.3.6 Step 4a: Individual-level analysis

The individual-level data included a row per episode of ARI. Columns included: ARI subtype, age band, the episode start date and the ISO year and week of the episode. A column indicating whether an episode occurred in the non-pandemic or pandemic period was also included. The outcome column was included as a binary flag indicating whether the outcome occurred. A column for each severity marker flag was also included as

outlined above.

The pandemic and non-pandemic periods were considered separately and thus the data was partitioned into two. To estimate the association between binary severity markers and outcomes within each strata, ORs were calculated to quantify the strength of the relationship between severity markers and outcomes. The unadjusted ORs were calculated directly from 2 x 2 contingency tables. This is computationally more efficient than using logistic regression that would return identical results (Table 5.3). A continuity correction of 0.5 was applied to all cells to prevent division by zero. ***The OR is the first metric used to evaluate the severity marker performance*** (Figure 5.1).

Confidence intervals were calculated using Woolf's approximation, which, like the odds ratios, is derived directly from the contingency tables and is computationally efficient [135, 136]. To account for multiple testing, a Bonferroni correction was applied separately within the pandemic and non-pandemic analyses, controlling the family-wise error rate at $\alpha = 0.05$ in each [137]. Results are presented in forest plots.

Table 5.6: Odds ratios and confidence intervals: individual analysis

Odds ratios and confidence intervals

Contingency table

	Outcome present	Outcome absent
Marker present	TP	FP
Marker absent	FN	TN

Odds ratios

Continued on next page

Table 5.6: Odds ratios and confidence intervals: individual analysis (continued)

Odds ratios and confidence intervals

$$\mathrm{OR} = \frac{(\mathrm{TP} + 0.5)(\mathrm{TN} + 0.5)}{(\mathrm{FP} + 0.5)(\mathrm{FN} + 0.5)}$$

Standard errors (Woolf)

$$\mathrm{SE}(\log \mathrm{OR}) = \sqrt{\frac{1}{\mathrm{TP} + 0.5} + \frac{1}{\mathrm{FP} + 0.5} + \frac{1}{\mathrm{FN} + 0.5} + \frac{1}{\mathrm{TN} + 0.5}}$$

Bonferroni adjustment

$$Z = \Phi^{-1}\left(1 - \frac{\alpha}{2\, n_{\mathrm{tests}}}\right)$$

Calculated separately for the pandemic and non pandemic analyses with $n_{\mathrm{tests}} = (\#\mathrm{markers}) \times (\#\mathrm{age\ bands}) \times (\#\mathrm{ARI\ subtypes})$.

$$\mathrm{CI}_{\mathrm{lower}} = \exp(\log \mathrm{OR} - Z \times \mathrm{SE}(\log \mathrm{OR}))$$

$$\mathrm{CI}_{\mathrm{upper}} = \exp(\log \mathrm{OR} + Z \times \mathrm{SE}(\log \mathrm{OR}))$$

Notes: TP: True positives; FP: False positives; TN: True negatives; FN: False negatives; OR: Odds ratio; SE(log OR): Standard error of the log odds ratio; $\mathrm{CI}_{\mathrm{lower}}$: Lower confidence interval bound; $\mathrm{CI}_{\mathrm{upper}}$: Upper confidence interval bound; Z: Critical value from the standard normal distribution using the Bonferroni adjusted α; Φ^{-1}: Inverse standard normal distribution function (Z score); α: Familywise error rate (for example 0.05 for 95% confidence intervals); n_{tests}: Number of tests in the Bonferroni family (markers × age bands × ARI subtypes).

5.3.7 Step 4b: Weekly aggregate analysis

The purpose of this analysis was to assess whether changes in the weekly aggregate predictors over time were similar to changes in the weekly aggregate outcomes. As the weekly aggregate analysis is more complex than the individual analysis I have laid out a summary of the specific sub-steps in Figure 5.3.

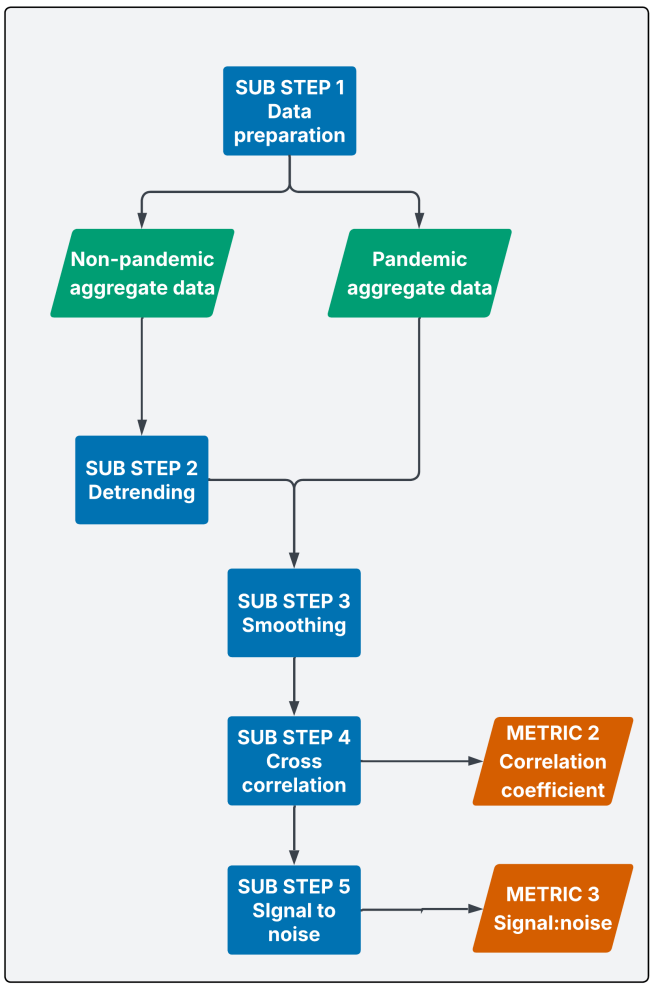


Figure 5.3: Overview of the analysis workflow. This figure covers the detail of Step 4b in the overall analysis plan (Figure 5.1). Green boxes indicate data partitions by study period. Blue boxes indicate steps in the analysis. Orange boxes indicate metrics

Data preparation

For all predictors, data were aggregated at the weekly level within each study stratum. This aggregation transformed the data from individual level data to weekly-level data. In the individual data, each row represented a single ARI case with binary flags indicating

the presence of the outcome and predictors. The resulting weekly-level data contained one row per study week, with columns indicating the proportion of individuals for whom each predictor was present.

This weekly data meets the criteria for a generic severity indicator outlined in Table 1.4. This results in a severity indicator for each of the 43 flags and the composite outcome. These can then be compared to establish their aggregate-level relationship. To ensure weeks with no episodes were not missed, a Cartesian grid of all study ISO weeks for combinations of ARI subtype and age bands (as described above) was created and merged with the aggregated dataset.

Detrending

The previous chapter identified long term trends in the data. These were likely due to practitioner recording behaviours rather than true changes in severity. Long term trends can result in spurious correlations between two time series. For example, a shared long term positive trend in the outcome and a severity indicator could result in a positive correlation coefficient that doesn't reflect a meaningful relationship between the two. These trends do not exist in all indicators and the exclusion of Period 1 from the analysis will limit the impact of trends. However, a methodology for removing trends is required to limit identification of spurious correlations.

I used locally estimated scatterplot smoothing (LOESS) to remove long-term trends (Figure 5.4) [138, 139]. LOESS is a localised, weighted, non-parametric regression that fits a smooth curve by performing weighted linear regression in a neighbourhood around each time point with weights that decrease with distance. The span parameter α controls the proportion of data used in each local fit and therefore takes a value of between 0 and 1 (exclusive of 0). A larger α produces a smoother curve (approaching linear). I set $\alpha = 0.6$, which uses the nearest 60% of points for each local fit, allowing some flexibility to capture non-linear trends while removing long-term drift. This value was chosen based on visual inspection of different levels for alpha. The fitted LOESS curve is then subtracted from

the original data to obtain detrended values (Figure 5.4).

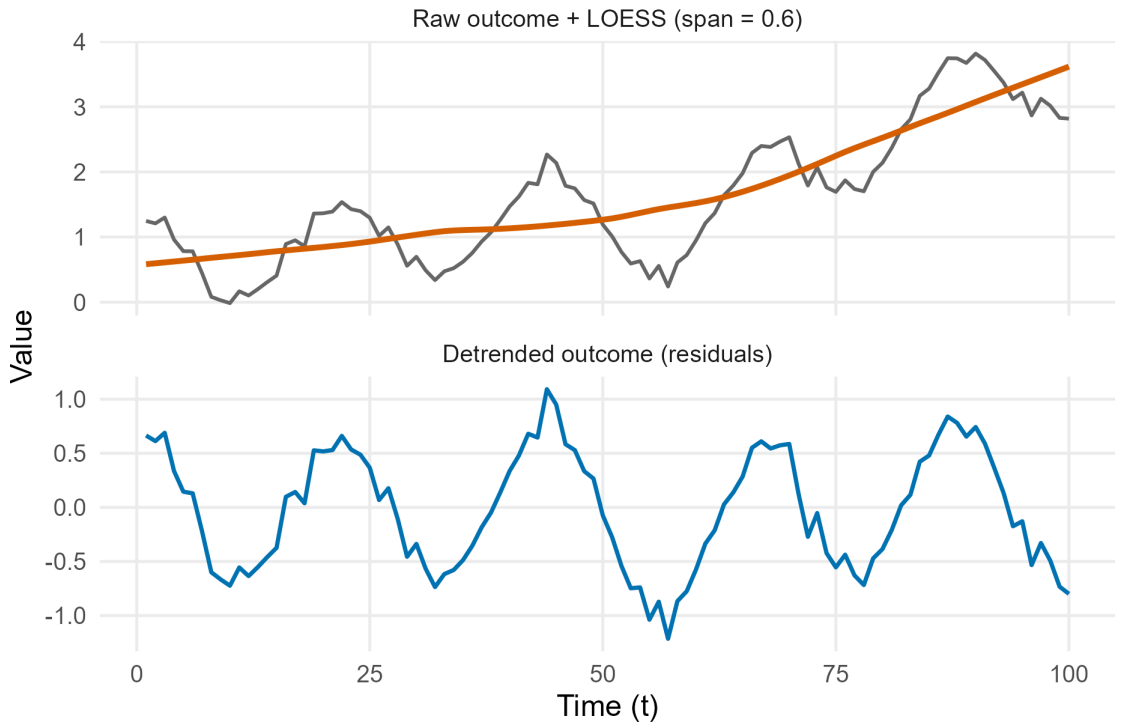


Figure 5.4: Locally estimated scatterplot smoothing (LOESS) detrending using synthetic data. **Top:** Example of raw data for a generic severity indicator in dark grey and a fitted LOESS line with an alpha (span) of 0.6. overlaid in orange. **Bottom:** detrended data, obtained by subtracting the LOESS fit from the raw data. Note the scale of detrended data is now centred around zero

For the non-pandemic analysis, I detrended all weekly aggregated predictor and outcome data using LOESS, irrespective of whether long-term trends were visually apparent. The rationale for this is that in the absence of a trend the LOESS will produce a model without a trend, and when subtracted results in a time series of the same shape to the original. In contrast, I did not perform detrending for the pandemic analysis. Due to the short length of the pandemic period (Period 3), trends were more likely to represent true epidemiological signals.

Smoothing

As identified in the previous chapter, recording rates were highly variable and indicator signals often noisy (Figure 5.5) [139]. To reduce this noise before assessing associations between severity indicators and the outcome, a trailing four-week rolling mean was applied

to all time series. This approach is commonly used to reduce the noise in epidemiological analyses and is easy to communicate. Specifically, each weekly value was replaced with the average of that week and the preceding three weeks.

I calculated a rolling four-week average as

$$\tilde{y}_t = \frac{1}{4} \sum_{i=0}^{3} y_{t-i},$$

where y_t is the value in ISO week t, and $\tilde{y}_t$ is the smoothed series.

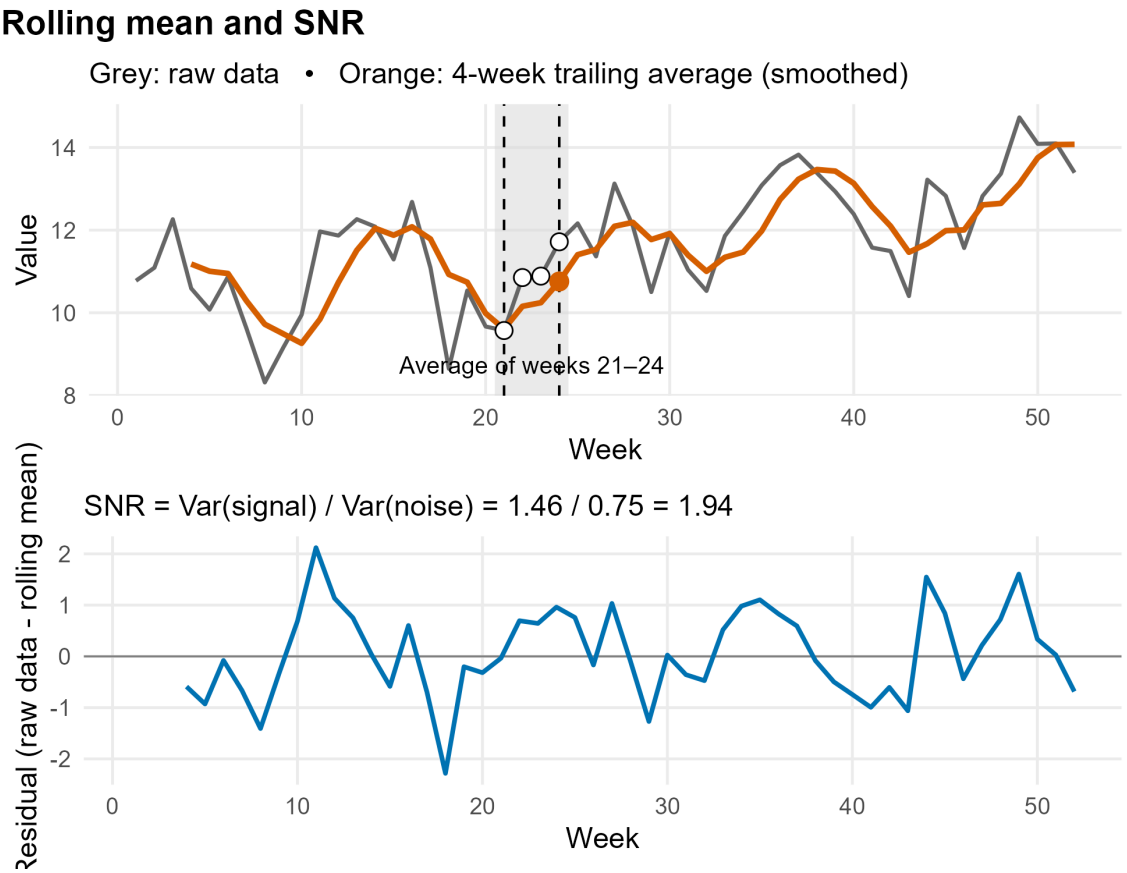


Figure 5.5: Top panel: Synthetic example of raw data for a generic severity indicator in dark grey with the 4-week rolling mean overlaid in orange. The shaded band highlights the window used to compute the rolling value at week t_{24}, averaging weeks t_{21}–t_{24}; the orange point marks the resulting mean. **Bottom panel:** Residuals, or noise, obtained by subtracting the rolling mean from the raw data. The signal-to-noise ratio (SNR) is calculated as the variance of the smoothed series divided by the variance of the residuals.

Cross correlation analysis

To measure the similarity between predictor and outcome time series, I used a cross correlation analysis. This calculates the Pearson's correlation coefficient between the two time series at varying lags. In this analysis these two time series are the predictor and the outcome time series (Appendix A4.1). Figure 5.6 is a synthetic example illustrating the correlation between two time series, the top panel shows where two time series have a strong correlation and the bottom shows a situation where there is no discernible correlation.

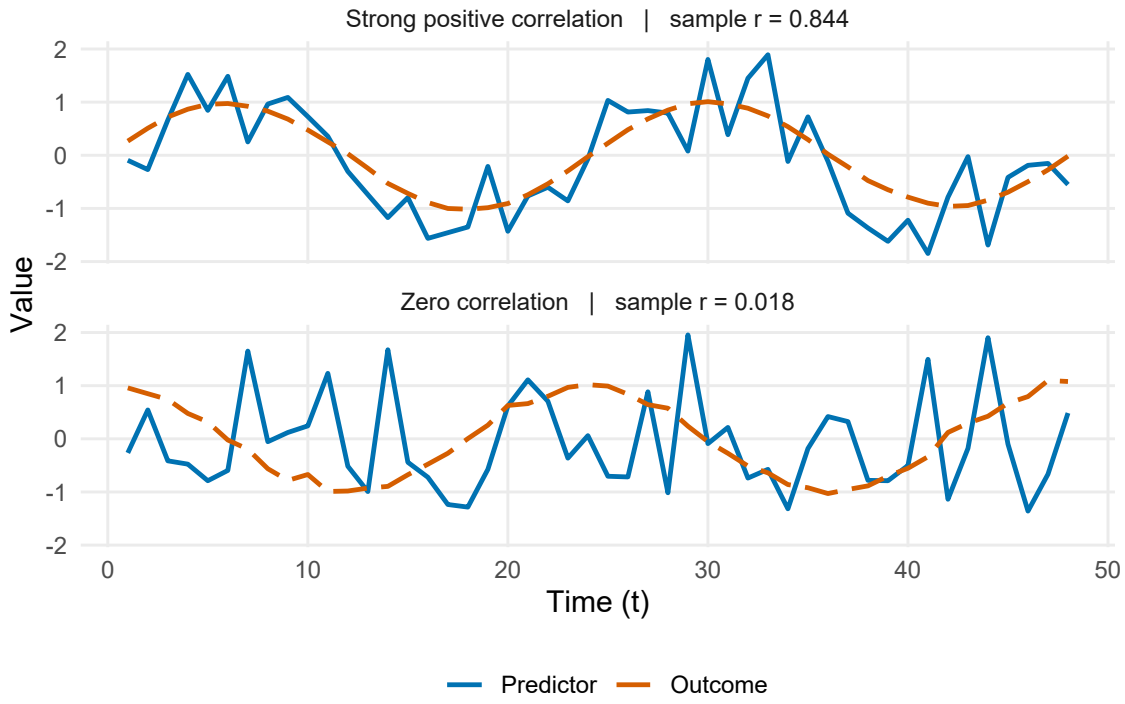


Figure 5.6: Examples of correlated and uncorrelated time series using synthetic data. **Top:** Strong positive correlation. **Bottom:** No substantial correlation. Predictors are shown in blue and the 'true' outcome signal in orange. Each panel reports the sample Pearson correlation r above the figure panel.

Cross-correlation identifies lagged relationships by computing Pearson's correlation coefficient across a range of lags [139]. This is done by shifting the predictor series forward (and/or backward) in one unit steps relative to the outcome series and recomputing the correlation at each step. This is traditionally used to quantify the association between two time series and ascertain at what lag this association is maximised. An example where we

might expect a lead-lag relationship could be the incidence of cases of ARI derived from a primary care population compared to the incidence of ARI in secondary care, as primary care attendances traditionally peak before hospitalisations [140].

However, in this DPhil, because the outcome and predictors are derived from the same population and indicators are based on the date of the index ARI case no substantial lead-lag relationship would be expected. Here, I restricted the cross-correlation analysis to lags from -2 to +2 (±2 weeks) to allow a small amount of flexibility and account for some noise, but did not attempt to define the lead-lag relationship as is traditional with cross correlation.

Confidence intervals for the CCs were defined using block bootstrap resampling [141]. This method repeatedly samples short continuous sections ('blocks') of the time series and re-runs the correlation analysis for each sample, at the maximum lag defined whilst calculating the confidence interval. This block-based approach preserves the ordered structure of the time series data, resulting in more appropriate CIs. The resulting set of estimates defines a bootstrap sampling distribution for the correlation coefficient. The standard deviation of this distribution provides an estimate of the standard error, from which confidence intervals are derived.

Signal to noise ratio

The strength of the underlying signal relative to random noise is called the SNR [142]. One common definition expresses this as the ratio of the signal variance to the noise variance:

$$\text{SNR} = \frac{\text{Var}(s_t)}{\text{Var}(n_t)}$$

where $\text{Var}(s_t)$ and $\text{Var}(n_t)$ are the variances of the signal and noise components, respectively. The SNR in this form can be derived from the rolling four-weekly mean. This is achieved by dividing the variance of the model (orange line, top panel, Figure 5.5), which represents the signal, by the variance of the residuals of the model (blue line, bottom panel,

Figure 5.5), which represents the noise. An SNR of greater than 1 indicates the signal is stronger than the noise and less than 1 indicates the noise is stronger than the signal. Figure 5.7 demonstrates how SNR can vary. ***The SNR is the third metric and is used to filter out noisy indicators, see detailed section in Step 5, in the following section, and Figure 5.1***.

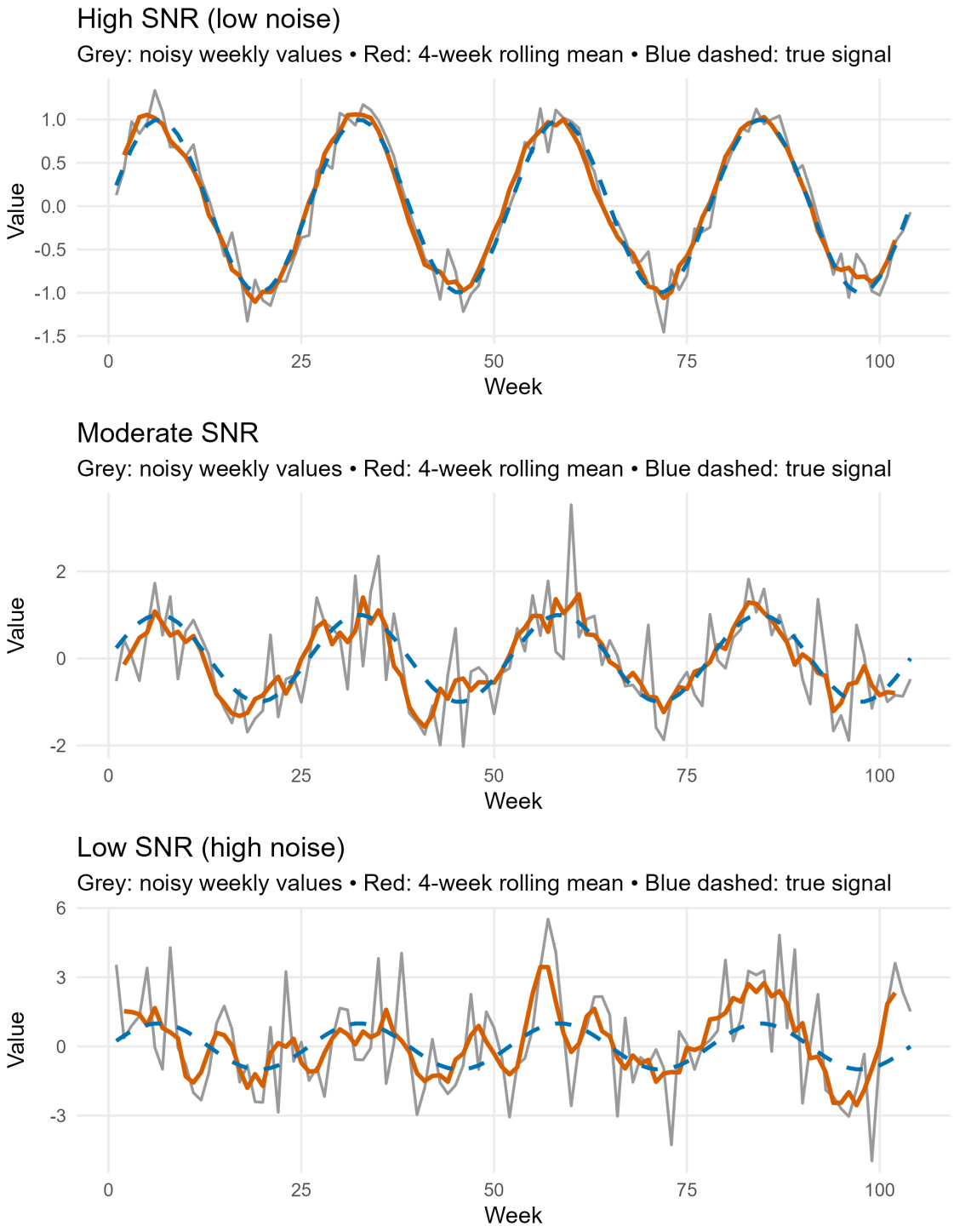


Figure 5.7: Example of variable signal-to-noise ratio (SNR) levels using synthetic data. Each panel shows an example with different degrees of noise. Example of raw data in dark grey, 4-week rolling mean overlaid in orange and 'true' underlying signal in dashed blue. **Top panel:** High SNR: minimal noise in the system. **Middle panel:** Moderate SNR, some noise but a signal is apparent. **Bottom panel:** Low SNR, high noise level and signal is almost lost to noise.

5.3.8 Step 5: Ranking indicators

Metrics 1 and 2 from the individual and aggregate analyses are then used to rank severity indicators. However, prior to ranking indicators a simple set of filters are applied using Metrics 1, 2 and 3 to exclude least relevant severity markers. Severity markers were included if they met the following criteria:

1. **Patient-level association:** indicators with a positive odds ratio (>0) and a statistically significant association at the patient level were retained.

2. **Aggregate-level association:** indicators with a positive and statistically significant correlation coefficient of moderate or greater strength at the selected lag were retained. I defined the cut off as 0.4, a commonly cited value for a moderate strength correlation [143].

3. **Signal quality:** indicators with a signal-to-noise ratio (SNR) greater than 1 were retained.

The OR and CC were then combined using the geometric mean (GM) (Table 5.7). Prior to calculating the GM, both OR and CC were rescaled to lie between 1 and 10. Specifically, the scaled value x_{scaled} was defined using min–max rescaling as

$$x_{\text{scaled}} = 1 + 9\,\frac{x - \min(x)}{\max(x) - \min(x)},$$

where $\min(x)$ and $\max(x)$ denote the minimum and maximum observed values of that metric in the sample. Rescaling is necessary because the CC is bound between -1 and 1, whereas the OR can theoretically extend from 0 to infinity. Without this step, large odds ratios would have a dominant contribution to the GM. The SNR is used to filter out noisy signals and is not a component of the GM, as it is not a reflection of association. This approach gives equal weight to the OR and CC, which is reasonable as there is no supporting evidence favouring the weighting of these. Furthermore, the GM is preferred

over a simple arithmetic mean as it favours a more balanced contribution from the OR and CC.

Table 5.7: Geometric mean definition

Composite score
$$\mathrm{GM}_i = \left(\tilde{\mathrm{OR}}_i \times \tilde{\mathrm{CC}}_i \right)^{\frac{1}{2}}$$
Notes
Where $\tilde{\mathrm{OR}}_i$ denotes the rescaled odds ratio (OR), and $\tilde{\mathrm{CC}}_i$ the rescaled cross-correlation coefficient (CC) for severity indicator i. Both metrics were rescaled to the range $[1, 10]$ to ensure comparability. The geometric mean (GM) is used so that an indicator must perform consistently across both metrics to achieve a high score, while poor performance in either dimension reduces the composite.

5.3.9 Sample size

A sample size calculation was not performed for this analysis as I have used all data available to me for the study period. This is a large representative sample derived from over 25.7 million episodes of ARI, and there is no mechanism to collect further data.

5.3.10 Missing data

As described in the previous chapter, the completeness of recording of severity marker data was highly variable. However, for binary severity markers it was not possible to state definitively whether data were truly missing or whether the marker was simply not recorded (Table 4.2). In this analysis, if a binary severity marker was not recorded, I assumed that it was not present.

For numeric severity markers, I took the same approach for the patient-level analysis,

assuming that if a severity marker was not recorded, a given threshold was not breached. However, for the weekly aggregate analysis, I took a different approach. Because the proportion of ARI episodes for which signs were recorded varied over time, I only calculated the rate of abnormal signs for those ARI episodes in which signs were recorded. In other words, episodes with missing values for numeric severity markers were not included in the calculation of those specific severity indicators. This helped to reduce the impact of differences in completeness of clinical sign recording over time.

5.4 RESULTS

5.4.1 Presentation of results

I present the results of the non-pandemic analysis first, followed by the pandemic analysis. The following results are displayed in each of these 2 sections:

1. Summary of population characteristics
2. Composite outcome and predictor reporting frequencies
3. Individual-level analysis forest plots of odds ratios
4. Weekly aggregate-level analysis forest plots of correlation coefficients
5. Ranked indicators by strata
6. Example of time series of top ranked indicators

As there are a large number of potential results to display, for items 2, 3, 4 and 6 in the above list, I only present the results for a single example stratum in both non-pandemic and pandemic periods. For the non-pandemic period I present these results for LRTI-ECLD combined in over 65 year olds and in the pandemic period I present results for suspected COVID-19 in over 65 year olds. For the remaining items I present results relevant to all strata. All results not presented in the main text are included in the Appendix.

To ensure that an overview of the whole analysis is presented, I combine the results of both analyses to compare the performance of all indicators in all strata in both the pandemic and non-pandemic periods.

5.4.2 Non-pandemic analysis

Study population

The non-pandemic analysis included 12,210,422 episodes of ARI: 7,385,365 URTI, 4,576,310 LRTI-ECLD combined and 248,747 episodes of ILI. The mean age of participants at the time of the episodes was 34.95 years and episodes occurred more commonly in women (57.9%). There were 1,690,783 episodes of LRTI-ECLD combined in over 65 year olds. Other demographic characteristics of individuals included in the study are presented in Table 5.8. Full analyses for all other strata including forest plots are included in Appendix A4.2.

Table 5.8: Baseline characteristics of ARI episodes, stratified by ARI subtype

Characteristic	Overall	URTI	LRTI-ECLD	ILI
n	12,210,422	7,385,365	4,576,310	248,747
Rate per week[a]	30,678	18,552	11,499	625
Age band (%)				
Under 15yrs	4,063,668 (33.3%)	3,321,570 (45.0%)	715,044 (15.6%)	27,054 (10.9%)
15 to 64yrs	5,857,349 (48.0%)	3,499,920 (47.4%)	2,170,483 (47.4%)	186,946 (75.2%)
Over 65yrs	2,289,405 (18.7%)	563,875 (7.6%)	1,690,783 (36.9%)	34,747 (14.0%)
Age years (mean (SD))				
Mean (SD)	34.95 (27.47)	25.07 (23.12)	50.52 (26.93)	41.57 (20.69)
Sex (%)				
Male	5,145,313 (42.1%)	3,058,063 (41.4%)	1,981,743 (43.3%)	105,507 (42.4%)
Ethnicity (%)				

Continued on next page

Table 5.8: Baseline characteristics (continued)

Characteristic	Overall	URTI	LRTI-ECLD	ILI
Asian	1,261,659 (10.3%)	876,845 (11.9%)	350,318 (7.7%)	34,496 (13.9%)
Black	398,051 (3.3%)	273,522 (3.7%)	112,089 (2.4%)	12,440 (5.0%)
Mixed	245,248 (2.0%)	182,080 (2.5%)	58,509 (1.3%)	4,659 (1.9%)
White	8,970,360 (73.5%)	5,138,055 (69.6%)	3,658,920 (80.0%)	173,385 (69.7%)
Other	204,657 (1.7%)	151,990 (2.1%)	47,772 (1.0%)	4,895 (2.0%)
Missing	1,130,447 (9.3%)	762,873 (10.3%)	348,702 (7.6%)	18,872 (7.6%)
Risk group (%)				
Risk group	4,388,383 (35.9%)	1,586,148 (21.5%)	2,724,378 (59.5%)	77,857 (31.3%)
Risk group (%)				
Risk group	906,919 (7.4%)	120,394 (1.6%)	778,089 (17.0%)	8,436 (3.4%)

Notes: URTI: upper respiratory tract infection; LRTI-ECLD: lower respiratory tract infection and exacerbations of chronic lung disease combined; ILI: influenza-like illness.
Frequency and percentages is shown for categorical variables and means standard deviations are shown for numeric variables.

[a] Average weekly rate calculated over 2,786 days (approximately 398 weeks).

Outcome and severity marker recording: LRTI-ECLD in over 65 year olds

Of the 1,690,783 included episodes of LRTI-ECLD in over 65 year olds, 257,073 (15.2%) had a recorded composite outcome (Table 5.9). The most commonly recorded predictors by far were prescriptions. The three most frequently recorded were: any antibiotic prescription (1,390,330; 82.2%), prescription of amoxicillin (863,964; 51.1%), and prescription of doxycycline (424,434; 25.1%). Prescriptions of oseltamivir (1,073; 0.1%),

cephalosporins (17,293; 1.0%) and penicillin (6,151; 0.4%) were less commonly recorded.

Table 5.9: Predictor recording: non-pandemic, LRTI-ECLD, 65+yrs

Parameter	**n (%)**
Episodes	
Episodes n	1,690,783
Outcome n (%)	257,073 (15.2%)
- Hospital attendance	168,003 (9.9%)
- Hospital admission	90,220 (5.1%)
- ICU admission	287 (0.0%)
- Complication	14,292 (0.8%)
- Died	62,613 (3.7%)
Predictors (ordered by % desc.)	
(1) Any antibiotic	1,390,330 (82.2%)
(2) Amoxicillin	863,964 (51.1%)
(3) Doxycycline	424,434 (25.1%)
(4) Any-sign Flag 1	424,156 (25.1%)
(5) Prednisilone	414,210 (24.5%)
(6) Macrolide	224,097 (13.3%)
(7) O_2 sats Flag 1	203,660 (12.0%)
(8) PR Flag 1	156,610 (9.3%)
(9) Any symptom	153,124 (9.1%)
(10) Any-sign Flag 2	152,731 (9.0%)
(11) Dyspnoea	123,354 (7.3%)
(12) Chest X-ray request	111,349 (6.6%)
(13) O_2 sats Flag 2	80,695 (4.8%)
(14) Chest signs	78,658 (4.7%)
(15) Temp Flag 1	66,097 (3.9%)
(16) Co-amoxiclav	62,128 (3.7%)
(17) BP Flag 1	60,969 (3.6%)
(18) Any-sign Flag 3	53,301 (3.2%)
(19) RR Flag 1	50,421 (3.0%)
(20) RR Flag 2	49,916 (2.9%)
(21) CRP	34,262 (2.0%)
(22) O_2 sats Flag 3	33,025 (2.0%)
(23) WBC	31,984 (1.9%)

Continued on next page

Table 5.9: Predictor recording: Non-pandemic, LRTI-ECLD, 65+yrs (continued)

Parameter	n (%)
(24) Ambulance	24,903 (1.5%)
(25) NHS 111	24,640 (1.5%)
(26) Malaise	19,579 (1.2%)
(27) PR Flag 2	19,180 (1.1%)
(28) BP Flag 2	18,089 (1.1%)
(29) RR Flag 3	17,517 (1.0%)
(30) Cefalosporin	17,293 (1.0%)
(31) Work absence	11,729 (0.7%)
(32) Fever	7,702 (0.5%)
(33) Confusion	7,120 (0.4%)
(34) Penicillin	6,151 (0.4%)
(35) BP Flag 3	4,396 (0.3%)
(36) Hospital referral	3,776 (0.2%)
(37) Haemoptysis	2,983 (0.2%)
(38) Temp Flag 2	2,326 (0.1%)
(39) PR Flag 3	2,223 (0.1%)
(40) Temp Flag 3	1,510 (0.1%)
(41) Oseltamivir	1,073 (0.1%)
(42) Work of breathing	298 (0.0%)
(43) Cyanosis	162 (0.0%)

Grouped sign flags were also common: any sign Flag 1 (424,156; 25.1%), any sign Flag 2 (152,731; 9.0%), and any sign Flag 3 (53,301; 3.2%). Among specific signs, the most frequently present were oxygen saturation Flag 1, which represents episodes with saturations < 95% (203,660; 12.0%), pulse rate Flag 1, which represents episodes with pulse rate <51 or >90 (156,610; 9.3%), and chest signs (78,658; 4.7%). Some signs: work of breathing (298; 0.0%), and cyanosis (162; 0.0%) were almost never recorded.

The most frequently recorded symptoms were dyspnoea (123,354; 7.3%), malaise (19,579; 1.2%), and fever (7,702; 0.5%). Other symptoms were rarely recorded: haemoptysis (2,983; 0.2%) and confusion (7,120; 0.4%).

Table 5.9 presents these results in full in descending order of frequency.

Individual-level analysis: LRTI-ECLD in over 65 year olds

All predictors had a significant association with the composite outcome except for a prescription of penicillin (Figure 5.8). Thirty-seven severity markers were positively associated with the outcome and 6 were negatively associated with the outcome. There were a wide range of ORs from 0.46 (any antibiotic prescription) to 11.18 (hospital referral).

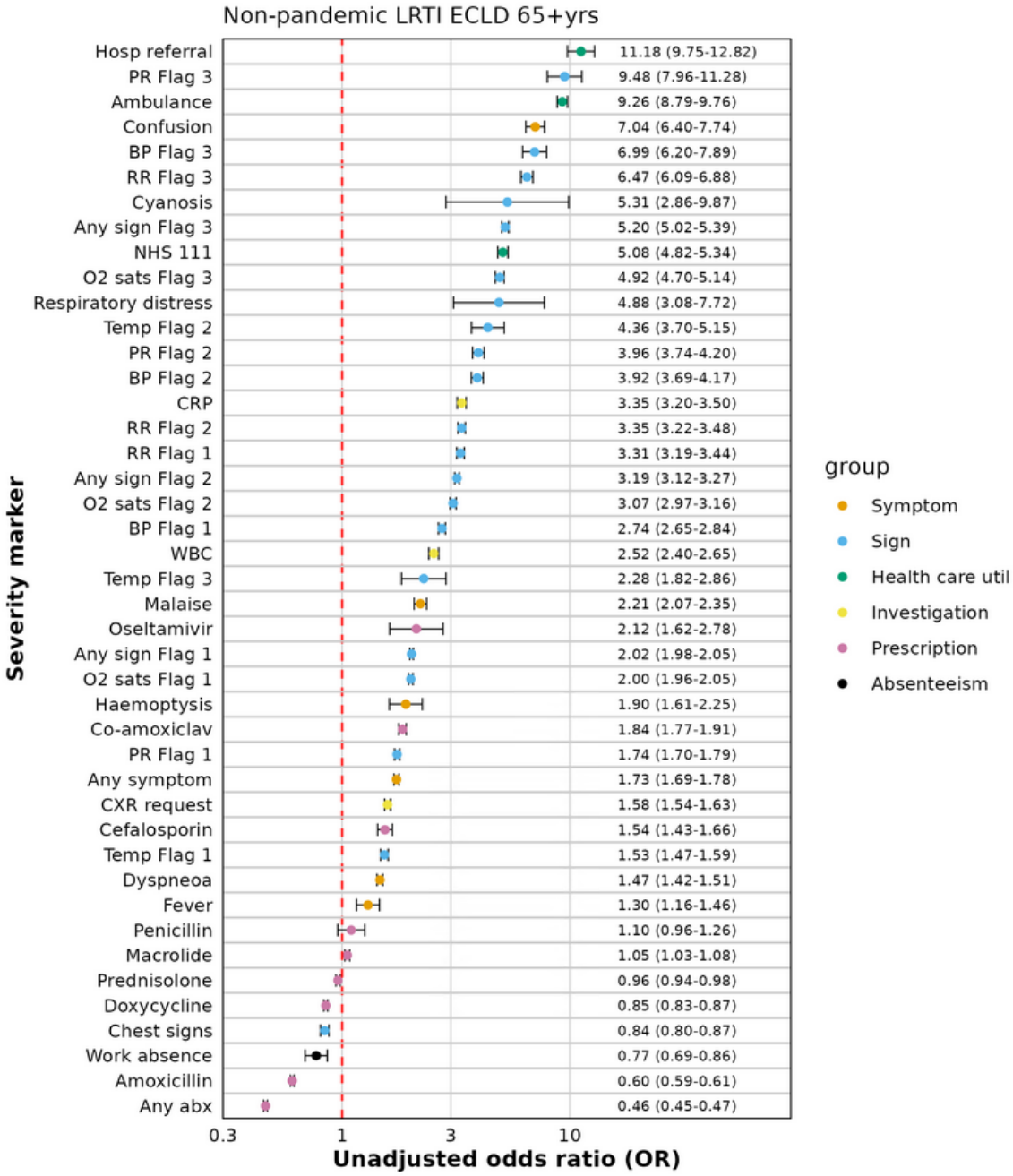


Figure 5.8: Odds ratios (OR) and 95% confidence intervals (CI) for the association between 43 severity markers and the composite outcome, for LRTI-ECLD cases in over 65 year olds. Markers with OR < 1 indicate a negative association, while those with OR > 1 indicate a positive association. Severity marker groupings are defined by colour as seen in the legend. Health care util: health care utilisation, Hosp: hospital, PR: pulse rate, BP: blood pressure, RR: respiratory rate, O2 sats: oxygen saturation, NHS 111: National Health Service 111 (telephone advice line), CRP: C-reactive protein, WBC: white blood cell, Temp: temperature, CXR: chest X-ray, Abx: antibiotic.

Those with the strongest positive associations included hospital referral (OR 11.18, CI 9.75–12.82), PR Flag 3 (OR 9.48, CI 7.96–11.28), ambulance encounter (OR 9.26, CI

8.79–9.76), confusion (OR 7.04, CI 6.40–7.74), BP Flag 3 (OR 6.99, CI 6.20–7.89), RR Flag 3 (OR 6.47, CI 6.09–6.88), cyanosis (OR 5.31, CI 2.86–9.87), and any sign Flag 3 (OR 5.20, CI 5.02–5.39).

Numeric clinical sign flags showed progressively stronger associations with the outcome from Flag 1 to Flag 3. For example, oxygen saturation Flag 1 was modestly associated with the outcome (OR 2.00, CI 1.96–2.05), while oxygen saturation Flag 2 showed a stronger association (OR 3.07, CI 2.97–3.16), and oxygen saturation Flag 3 the strongest (OR 4.92, CI 4.70–5.14).

Temperature was the only sign where flags were not sequentially associated with the outcome, as Flag 3 was less strongly associated than Flag 2. Specifically, temperature Flag 1 was modestly associated with the outcome (OR 1.53, CI 1.47–1.59), Flag 2 showed a stronger association (OR 4.36, CI 3.70–5.15), but Flag 3 was weaker (OR 2.28, CI 1.82–2.86). Flag 3 corresponded to a temperature $< 35.1°C$ (Table 5.4).

Six of the 43 severity markers were negatively associated with the outcome: any antibiotic (OR 0.46, CI 0.45–0.47), amoxicillin (OR 0.60, CI 0.59–0.61), work absence (OR 0.76, CI 0.68–0.85), chest signs (OR 0.84, CI 0.80–0.88), doxycycline (OR 0.85, CI 0.83–0.87), and prednisolone (OR 0.96, CI 0.94–0.99).

Weekly aggregate analysis: LRTI-ECLD in over 65 year olds

For the weekly aggregate analysis, 20 severity indicators had positive and significant CCs, 9 had negative and significant CCs, and 14 had non-significant positive or negative CCs. The five indicators with the largest positive CC were: BP Flag 1 (CC 0.84, CI 0.74–0.90), BP Flag 2 (CC 0.73, CI 0.58–0.83), O_2 sats Flag 1 (CC 0.69, CI 0.45–0.84), O_2 sats Flag 2 (CC 0.66, CI 0.41–0.82), and any sign Flag 2 (CC 0.66, CI 0.45–0.81).

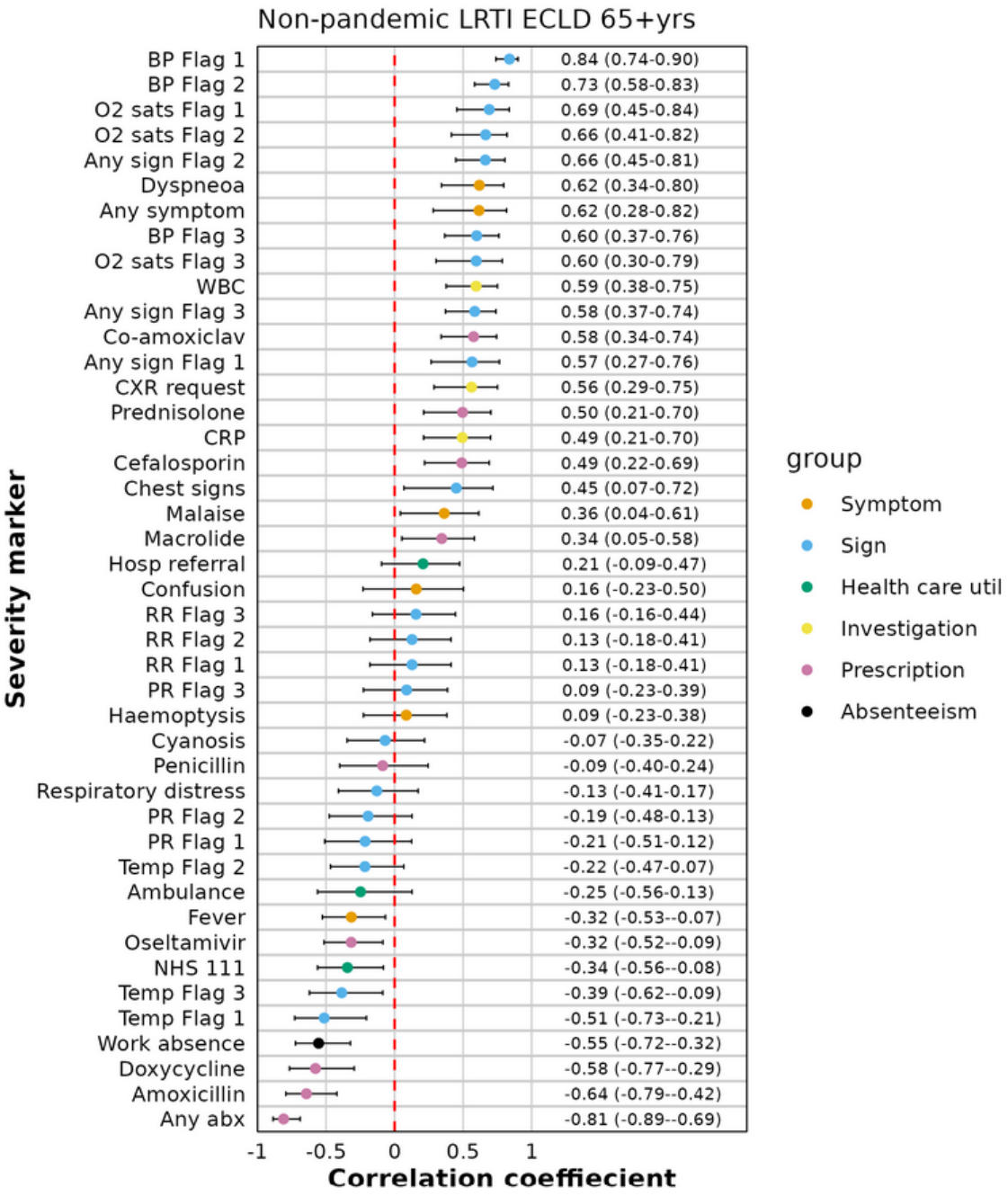


Figure 5.9: Forest plot showing unadjusted relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death

Several markers showed negative correlation with the outcome, including any antibiotic (CC –0.81, –0.89 to –0.69), amoxicillin (CC –0.64, CI –0.79 to –0.42), doxycycline (CC –0.58, CI –0.77 to –0.29), work absence (CC –0.55, CI –0.72 to –0.32), temperature Flag 1 (CC –0.51, CI –0.73 to –0.21), and temperature Flag 3 (CC –0.39, CI –0.62 to –0.09).

Severity marker ranking: all strata

After applying the filters (exclusion of non-associated and noisy signals (Section 5.3.8)), for non-pandemic URTI and LRTI-ECLD, a total of 27 non-unique severity markers remained (some appear in multiple strata). These comprised 14 clinical signs, 6 symptoms, 4 prescriptions and 3 investigations. Across both URTI and LRTI-ECLD analyses, there were 15 unique severity markers: 8 clinical signs, 3 symptoms, 2 prescriptions, and 2 investigations.

URTI: Five severity indicators remained for 15 to 64 year olds (Table 5.10). These were: malaise (GM 2.94, SNR 2.37), co-amoxiclav (GM 2.93, SNR 1.58), BP Flag 1 (GM 2.91, SNR 4.16), BP Flag 2 (GM 2.77, SNR 1.71) and any symptom (GM 2.65, SNR 1.88). For children <15 years 2 severity indicators remained: chest signs (GM 4.23, SNR 2.72) and prednisolone (GM 2.90, SNR 5.83).

Table 5.10: Top indicators for non-pandemic URTI

Predictor	OR (95% CI)	CCF (95% CI)	SNR	GM	Rank
15–64 yrs					
Malaise	1.45 (1.35–1.56)	0.71 (0.53–0.82)	2.37	2.94	1
Co-amoxiclav	1.78 (1.67–1.90)	0.64 (0.43–0.78)	1.58	2.93	2
BP Flag 1	1.48 (1.42–1.54)	0.69 (0.49–0.82)	4.16	2.91	3
BP Flag 2	1.58 (1.46–1.71)	0.63 (0.35–0.80)	1.71	2.77	4
Any symptom	1.51 (1.44–1.58)	0.61 (0.40–0.75)	1.88	2.65	5
<15 yrs					
Chest signs	3.48 (3.26–3.71)	0.73 (0.57–0.83)	2.72	4.23	1
Prednisolone	1.49 (1.41–1.57)	0.69 (0.50–0.81)	5.83	2.90	2

Notes: The geometric mean is of the scaled OR and CC. SNR: signal to noise ratio, GM: geometric mean, BP: blood pressure.

LRTI/ECLD: After applying filters (exclusion of non-associated and noisy signals (Section 5.3.8), 9 severity indicators remained for those aged 65+ years (Table 5.11). These

were BP Flag 2 (GM 4.48, SNR 1.21), BP Flag 1 (GM 4.23, SNR 2.98), any sign Flag 2 (GM 3.82, SNR 1.23), O_2 saturation Flag 2 (GM 3.76, SNR 1.11), O_2 saturation Flag 1 (GM 3.26, SNR 1.68), any symptom (GM 2.83, SNR 4.18), any sign Flag 1 (GM 2.80, SNR 2.08), dyspnoea (GM 2.67, SNR 4.47), and CXR request (GM 2.54, SNR 3.51). For adults aged 15–64 years, 10 severity markers remained: O_2 saturation Flag 2 (GM 4.45, SNR 1.18), BP Flag 2 (GM 3.62, SNR 1.54), BP Flag 1 (GM 3.40, SNR 3.99), O_2 saturation Flag 1 (GM 3.38, SNR 1.80), CRP (GM 3.37, SNR 1.29), co-amoxiclav (GM 3.17, SNR 1.66), CXR request (GM 3.16, SNR 3.83), prednisolone (GM 2.93, SNR 7.97), dyspnoea (GM 2.65, SNR 3.95), and any symptom (GM 2.60, SNR 3.17). For children <15 years, a single severity indicator remained: temperature Flag 1 (GM 1.83, SNR 3.95).

ILI: No indicators remained for ILI after filtering.

Table 5.11: Top indicators for non-pandemic LRTI-ECLD

Predictor	OR (95% CI)	CC (95% CI)	SNR	GM	Rank
65+ yrs					
BP Flag 2	3.92 (3.69–4.17)	0.73 (0.58–0.83)	1.21	4.48	1
BP Flag 1	2.74 (2.65–2.84)	0.84 (0.74–0.90)	2.98	4.23	2
Any sign Flag 2	3.19 (3.12–3.27)	0.66 (0.45–0.81)	1.23	3.82	3
O_2 sats Flag 2	3.07 (2.97–3.16)	0.66 (0.41–0.82)	1.11	3.76	4
O_2 sats Flag 1	2.00 (1.96–2.05)	0.69 (0.45–0.84)	1.68	3.26	5
Any symptom	1.73 (1.69–1.78)	0.62 (0.28–0.82)	4.18	2.83	6
Any sign Flag 1	2.02 (1.98–2.05)	0.57 (0.27–0.76)	2.08	2.80	7
Dyspnoea	1.47 (1.42–1.51)	0.62 (0.34–0.80)	4.47	2.67	8
CXR request	1.58 (1.54–1.63)	0.56 (0.29–0.75)	3.51	2.54	9
15–64 yrs					
O_2 sats Flag 2	3.88 (3.66–4.12)	0.73 (0.54–0.85)	1.18	4.45	1
BP Flag 2	2.20 (2.04–2.38)	0.76 (0.59–0.86)	1.54	3.62	2
BP Flag 1	1.71 (1.64–1.79)	0.80 (0.66–0.88)	3.99	3.40	3
O_2 sats Flag 1	2.05 (1.98–2.13)	0.72 (0.49–0.85)	1.80	3.38	4
CRP	3.75 (3.50–4.01)	0.52 (0.28–0.70)	1.29	3.37	5
Co-amoxiclav	1.97 (1.87–2.07)	0.67 (0.49–0.80)	1.66	3.17	6

Continued on next page

Table 5.11: Top indicators for non-pandemic LRTI-ECLD (continued)

Predictor	OR (95% CI)	CC (95% CI)	SNR	GM	Rank
CXR request	1.87 (1.81–1.94)	0.69 (0.44–0.83)	3.83	3.16	7
Prednisolone	1.29 (1.26–1.32)	0.74 (0.56–0.86)	7.97	2.93	8
Dyspnoea	1.81 (1.74–1.88)	0.56 (0.29–0.74)	3.95	2.65	9
Any symptom	1.81 (1.75–1.87)	0.54 (0.25–0.74)	3.17	2.60	10
<15 yrs					
Temp Flag 1	1.50 (1.43–1.58)	0.40 (0.07–0.66)	3.95	1.83	1

Notes: The geometric mean is of the scaled OR and CC. SNR: signal-to-noise ratio, GM: geometric mean, BP: blood pressure, O_2 sats: oxygen saturation, CXR: chest X-ray, CRP: C-reactive protein.

Figure 5.10 shows an example severity indicator and its comparison with the composite outcome. The top panel shows how the proportion of cases with blood pressure Flag 2 (top ranked indicator for LRTI-ECLD in over 65 year olds) changes in study Period 2 (the late pre-pandemic period). The lower panel shows how the proportion of cases with the outcome varies over time. Peaks in both broadly align, which is reflected in the correlation coefficient of 0.73 (CI 0.58-0.83). A notable finding is that severity peaks in the summer months, indicated by the arrows in the lower panel.

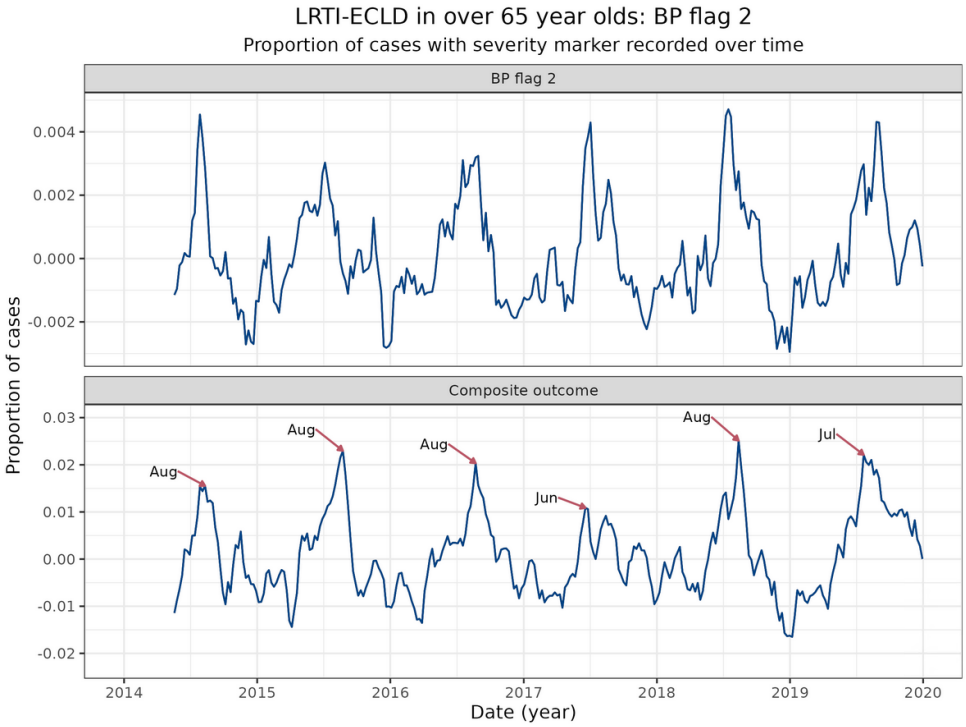


Figure 5.10: Example indicator: non-pandemic. TOP PANEL: Rolling 4-weekly average of proportion of cases of LRTI and ECLD combined in over 65 year olds for whom the blood pressure Flag 2 (systolic blood pressure (BP) less than 101) was present. BOTTOM PANEL: weekly proportion of the same cases with the composite outcome recorded. Arrows on bottom panel indicate the month in which the peak cases occurred during the given year. ***NOTE: The y-axis scale is centred around zero as a consequence of trend adjustment and therefore this doesn't reflect the true proportion***

5.4.3 Pandemic analysis

Study population

The pandemic analysis included 2,907,933 episodes of suspected COVID-19. The mean age of participants at the time of the episodes was 37.24 years and participants were more commonly women (59.1%). There were 540,500 episodes of suspected COVID-19 in over 65 year olds. Other demographic characteristics of individuals included in the study are presented in Table 5.12.

Table 5.12: Baseline characteristics of suspected COVID-19

Characteristic / level	Value
n	2,907,933
Rate per week[a]	20,273
Age band (%)	
Under 15yrs	777,840 (26.7%)
15 to 64yrs	1,589,593 (54.7%)
Over 65yrs	540,500 (18.6%)
Age years (mean (SD))	
Mean (SD)	37.24 (26.45)
Sex (%)	
Male	1,190,500 (40.9%)
Ethnicity (%)	
Asian	275,702 (9.5%)
Black	90,580 (3.1%)
Mixed	59,795 (2.1%)
White	2,222,238 (76.4%)
Other	51,890 (1.8%)
Missing	207,728 (7.1%)
Risk group (%)	
Risk group	1,134,916 (39.0%)
Resp risk group (%)	
Resp risk group	243,952 (8.4%)

Notes: Values are *n (percentage)* unless stated. Means (SD) shown for continuous variables.

[a] Average weekly rate calculated over 1,004 days (approximately 143 weeks).

Outcome and severity marker recording

Of the 540,500 included episodes of suspected COVID-19 in over 65 year olds, 88,471 (16.4%) had a recorded outcome (Table 5.13). Prescriptions were again the most commonly recorded severity markers, but were substantially less commonly recorded than during the non-pandemic period. For example, any antibiotic was recorded in 344,719 (63.8%) pandemic episodes compared with 1,390,330 (82.2%) non-pandemic episodes; amoxicillin in 181,758 (33.6%) vs. 863,964 (51.1%), and doxycycline in 126,667 (23.4%) vs. 424,434 (25.1%), respectively.

Table 5.13: Predictor recording: Pandemic, Suspected COVID, 65+yrs

Predictor	**n (%)**
Episodes	
Episodes n	540,500
Outcome n (%)	88,471 (16.4%)
- Hospital attendance	56,397 (10.4%)
- Hospital admission	27,569 (5.1%)
- ICU admission	166 (0.0%)
- Complication	4,910 (0.9%)
- Died	26,390 (4.9%)
Predictors (ordered by % desc.)	
Any antibiotic	344,719 (63.8%)
Amoxicillin	181,758 (33.6%)
Doxycycline	126,667 (23.4%)
Prednisolone	100,074 (18.5%)
Any-sign (Level 1)	95,486 (17.7%)
Any symptom	57,542 (10.6%)
Dyspnoea	44,628 (8.3%)
O_2 sats Flag 1	42,656 (7.9%)
PR Flag 1	33,938 (6.3%)
Any-sign (Level 2)	35,835 (6.6%)
CXR request	25,495 (4.7%)
NHS 111	21,958 (4.1%)
Ambulance	16,270 (3.0%)
Co-amoxiclav	16,154 (3.0%)

Continued on next page

Table 5.13: Predictor recording: Pandemic, Suspected COVID, 65+yrs (continued)

Predictor	n (%)
Penicillin	14,439 (2.7%)
Temp Flag 1	13,681 (2.5%)
BP Flag 1	13,692 (2.5%)
RR Flag 1	13,557 (2.5%)
RR Flag 2	13,461 (2.5%)
Any-sign (Level 3)	13,830 (2.6%)
Chest signs	12,603 (2.3%)
CRP	10,254 (1.9%)
O_2 sats Flag 2	18,078 (3.3%)
O_2 sats Flag 3	8,335 (1.5%)
WBC	8,772 (1.6%)
Malaise	7,457 (1.4%)
Fever	6,477 (1.2%)
Work absence	5,677 (1.0%)
RR Flag 3	5,183 (1.0%)
Macrolide	4,589 (0.8%)
PR Flag 2	4,602 (0.8%)
Cefalosporin	4,114 (0.8%)
BP Flag 2	4,260 (0.8%)
Confusion	3,079 (0.6%)
Haemoptysis	1,017 (0.2%)
BP Flag 3	1,117 (0.2%)
Hospital referral	1,305 (0.2%)
PR Flag 3	667 (0.1%)
Temp Flag 2	664 (0.1%)
Temp Flag 3	289 (0.1%)
Work of breathing	134 (0.0%)
Oseltamivir	174 (0.0%)
Cyanosis	58 (0.0%)

Individual-level analysis: Suspected COVID-19 in over 65 year olds

All predictors had a significant association with the composite outcome except for oseltamivir, chest signs and prednisolone (Figure 5.11). Thirty-five severity markers were positively and significantly associated with the outcome, with ORs ranging from 1.09 to

12.12. Five markers were negatively and significantly associated with the outcome, with ORs ranging from 0.41 to 0.84.

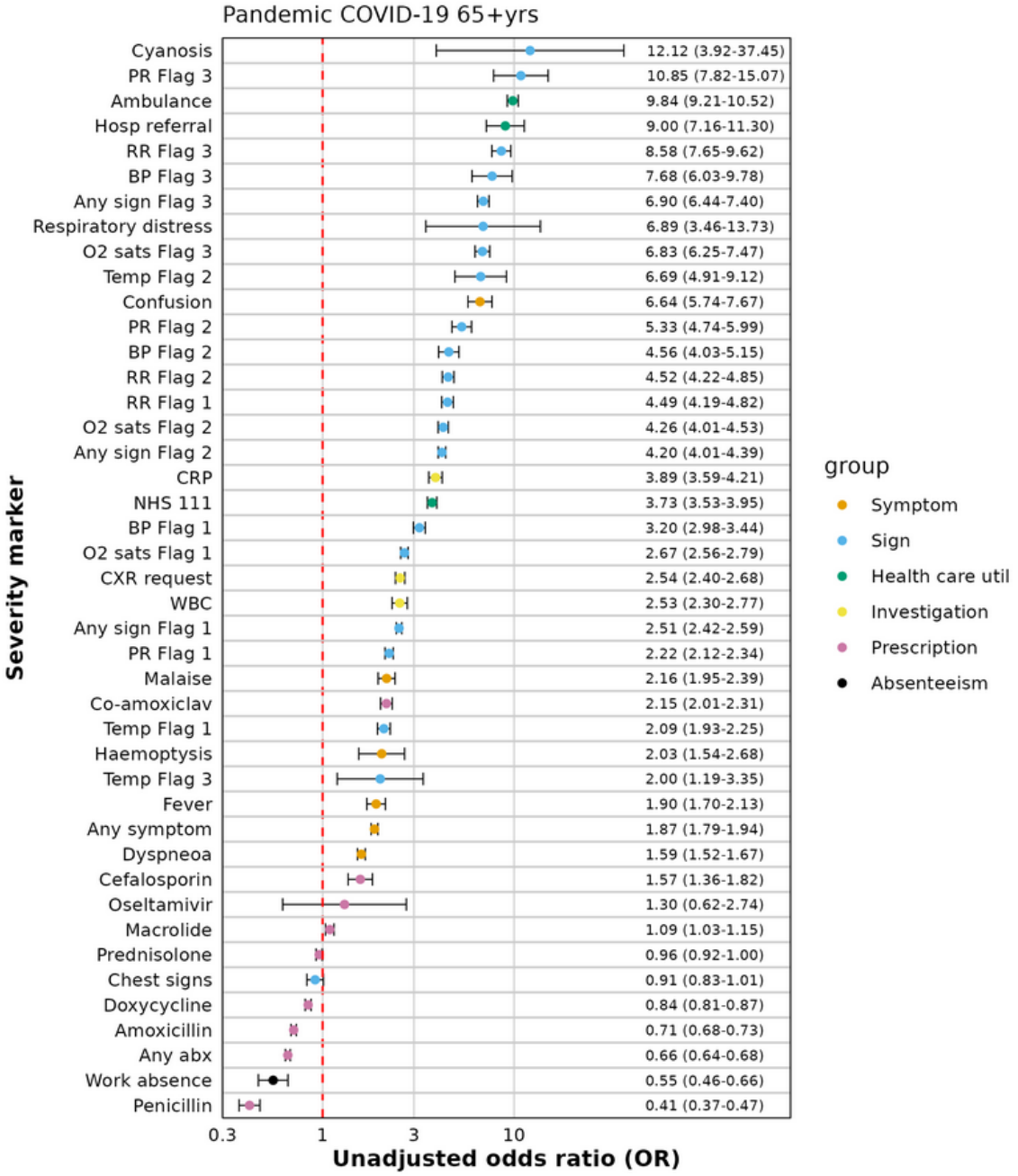


Figure 5.11: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death

Those with the strongest positive associations included: cyanosis (OR 12.12, CI 3.92 to 37.45), PR Flag 3 (OR 10.85, CI 7.82 to 15.07), ambulance encounter (OR 9.84, CI 9.21 to 10.52), hospital referral (OR 9.00, CI 7.16 to 11.30), and RR Flag 3 (OR 8.58, CI 7.65 to 9.62). Notably, cyanosis had wide confidence intervals, reflecting that it was the least

commonly recorded severity marker (Table 5.11).

Numeric clinical sign flags (except for temperature) again showed progressively stronger associations with the outcome from Flag 1 to Flag 3. For example, oxygen saturation Flag 1 was modestly associated with the outcome (OR 2.67, CI 2.56 to 2.79), while Flag 2 showed a stronger association (OR 4.26, CI 4.01 to 4.53), and Flag 3 the strongest (OR 6.83, CI 6.25 to 7.47).

Severity markers negatively associated with the outcome included: any antibiotic (OR 0.66, CI 0.64 to 0.68), amoxicillin (OR 0.71, CI 0.68 to 0.73), work absence (OR 0.55, CI 0.46 to 0.66), doxycycline (OR 0.84, CI 0.81 to 0.87) and penicillin (OR 0.41, CI 0.37 to 0.47).

Weekly aggregate analysis: Suspected COVID-19 in over 65 year olds

For the weekly aggregate analysis, 16 severity indicators had positive and significant CCs, 6 had negative and significant CCs, and 21 had non-significant positive or negative CCs. The five indicators with the largest positive CC were malaise (CC 0.88, CI 0.69 to 0.96), O_2 saturation Flag 3 (CC 0.83, CI 0.27 to 0.97), confusion (CC 0.81, CI 0.63 to 0.91), any sign Flag 3 (CC 0.80, CI 0.42 to 0.94), and co-amoxiclav (CC 0.79, CI 0.24 to 0.96). The strongest negative CCs were observed for amoxicillin (CC -0.74, CI -0.89 to -0.47), any antibiotic (CC -0.64, CI -0.88 to -0.16), PR Flag 1 (CC -0.59, CI -0.77 to -0.33), doxycycline (CC -0.58, CI -0.86 to -0.33), any sign Flag 1 (CC -0.45, CI -0.70 to -0.10), and prednisolone (CC -0.42, CI -0.61 to -0.16).

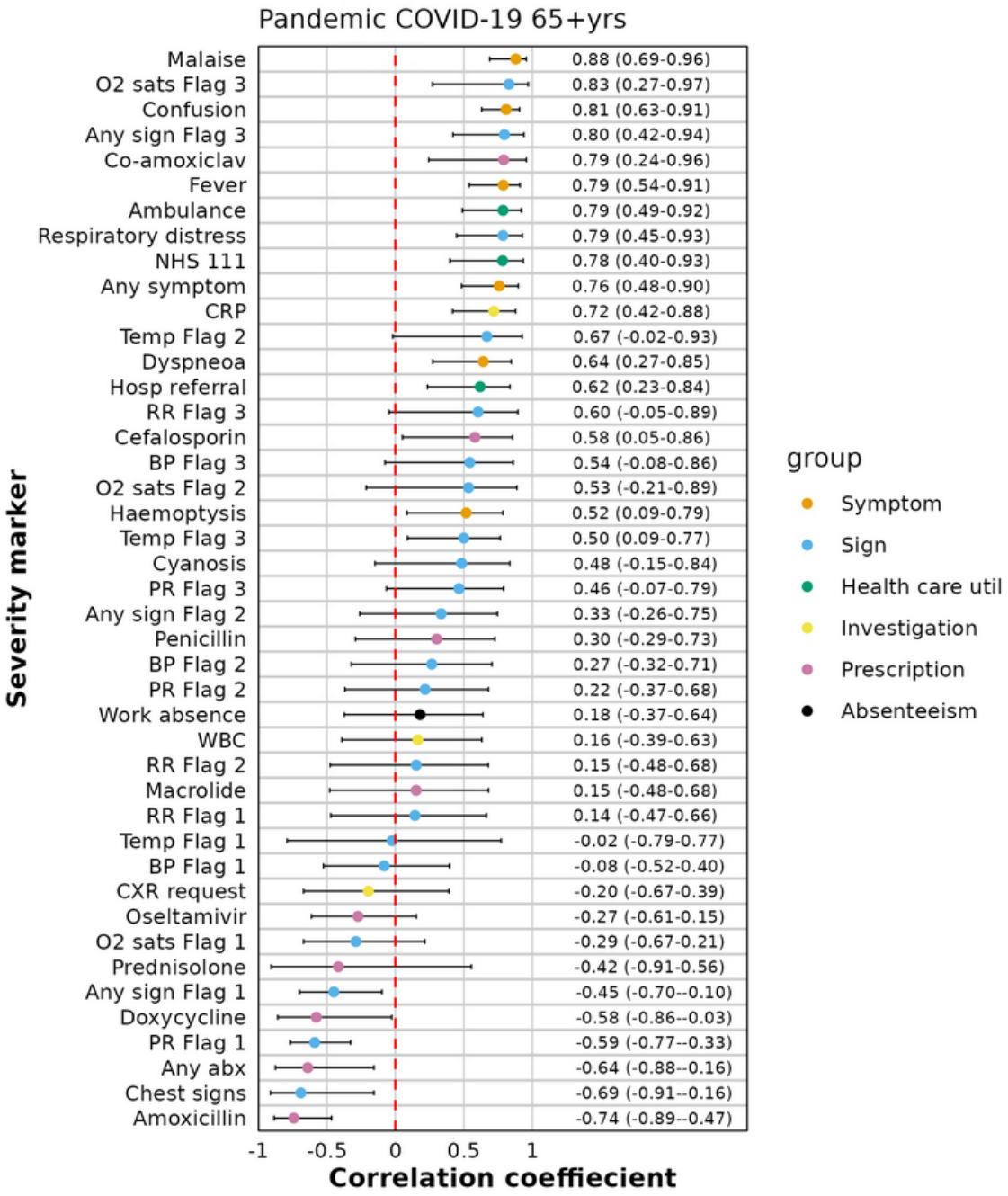


Figure 5.12: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death

Severity marker ranking: all strata

For suspected COVID-19 during the pandemic, a total of 21 non-unique severity markers were included after filtering (described in Section 5.3.8) (some appear in multiple strata). These comprised 10 symptoms, 5 health care utilisation indicators, 2 clinical sign flags, 2 investigations, and 2 prescriptions. Across both age strata, there were 14 unique

severity markers: 6 symptoms, 3 health care utilisation indicators, 2 clinical sign flags, 2 investigations, and 1 prescription.

Suspected COVID-19: After applying the relevant filters, 11 severity indicators remained for suspected COVID-19 in adults aged 65 years and older (Table 5.14). These comprised: 2 sign threshold markers (O_2 saturation Flag 3 and any sign Flag 3), 5 symptoms (confusion, malaise, fever, dyspnoea, and 'any symptom'), 2 health care utilisation indicators (ambulance and NHS 111), 1 investigation (CRP), and 1 prescription (co-amoxiclav). The top 5 ranked indicators by GM were ambulance (GM 7.17, SNR 4.85), O_2 saturation Flag 3 (GM 6.27, SNR 2.85), any sign Flag 3 (GM 6.13, SNR 3.36), confusion (GM 6.09, SNR 1.93), and NHS 111 (GM 4.60, SNR 7.51).

Table 5.14: Top indicators for pandemic suspected COVID

Predictor	OR (95% CI)	CC (95% CI)	SNR	GM	Rank
65+ yrs					
Ambulance	9.84 (9.21–10.52)	0.79 (0.49–0.92)	4.85	7.17	1
O_2 sats Flag 3	6.83 (6.25–7.47)	0.83 (0.27–0.97)	2.85	6.27	2
Any sign Flag 3	6.90 (6.44–7.40)	0.80 (0.42–0.94)	3.36	6.13	3
Confusion	6.64 (5.74–7.67)	0.81 (0.63–0.91)	1.93	6.09	4
NHS 111	3.73 (3.53–3.95)	0.78 (0.40–0.93)	7.51	4.60	5
CRP	3.89 (3.59–4.21)	0.72 (0.42–0.88)	1.79	4.41	6
Malaise	2.16 (1.95–2.39)	0.88 (0.69–0.96)	6.77	3.97	7
Co-amoxiclav	2.15 (2.01–2.31)	0.79 (0.24–0.96)	3.73	3.69	8
Fever	1.90 (1.70–2.13)	0.79 (0.54–0.91)	20.00	3.51	9
Any symptom	1.87 (1.79–1.94)	0.76 (0.48–0.90)	20.09	3.39	10
Dyspnoea	1.59 (1.52–1.67)	0.64 (0.27–0.85)	14.59	2.83	11
15–64 yrs					
Hosp referral	16.84 (14.19–19.97)	0.81 (0.64–0.91)	1.11	9.50	1
Ambulance	10.59 (9.88–11.36)	0.82 (0.65–0.92)	9.93	7.67	2
NHS 111	3.71 (3.57–3.85)	0.75 (0.50–0.88)	16.94	4.44	3
Haemoptysis	3.51 (2.89–4.28)	0.72 (0.43–0.87)	1.71	4.22	4
Dyspnoea	2.50 (2.38–2.62)	0.80 (0.59–0.91)	34.94	3.95	5

Continued on next page

Table 5.14: Top indicators for pandemic suspected COVID (continued)

Predictor	OR (95% CI)	CC (95% CI)	SNR	GM	Rank
Co-amoxiclav	2.08 (1.93–2.24)	0.84 (0.45–0.96)	6.74	3.80	6
Any symptom	2.16 (2.08–2.25)	0.79 (0.52–0.91)	45.73	3.68	7
Malaise	1.69 (1.55–1.85)	0.79 (0.42–0.94)	10.38	3.37	8
CXR request	3.60 (3.41–3.79)	0.49 (0.00–0.79)	6.80	3.15	9
Fever	1.70 (1.58–1.83)	0.71 (0.32–0.89)	35.31	3.12	10

Notes: The geometric mean is calculated from the scaled OR and CC. SNR: signal-to-noise ratio, GM: geometric mean, O_2 sats: oxygen saturation, CRP: C-reactive protein, CXR: chest X-ray.

After applying the relevant filters, 10 severity indicators remained for suspected COVID in adults aged 15–64 years during the pandemic period (Table 5.14). These comprised: 5 symptoms (haemoptysis, dyspnoea, malaise, fever, and the any symptom), 3 health care utilisation indicators (hospital referral, ambulance, and NHS 111), 1 investigation (CXR request), and 1 prescription (co-amoxiclav). The top 5 ranked indicators by GM were hospital referral (GM 9.50, SNR 1.11), ambulance (GM 7.67, SNR 9.93), NHS 111 (GM 4.44, SNR 16.94), haemoptysis (GM 4.22, SNR 1.71), and dyspnoea (GM 3.95, SNR 34.94).

Figure 5.13 shows an example severity indicator and its comparison with the composite outcome. The top panel shows how the proportion of cases with oxygen saturation Flag 3 changes in Period 3 (the pandemic period). The lower panel shows how the proportion of cases with the outcome varies over time. Peaks in both indicator and outcome are broadly aligned and are consistent with peaks in severity associated with differing waves of COVID-19 caused by the variants marked in the figure by the red arrows.

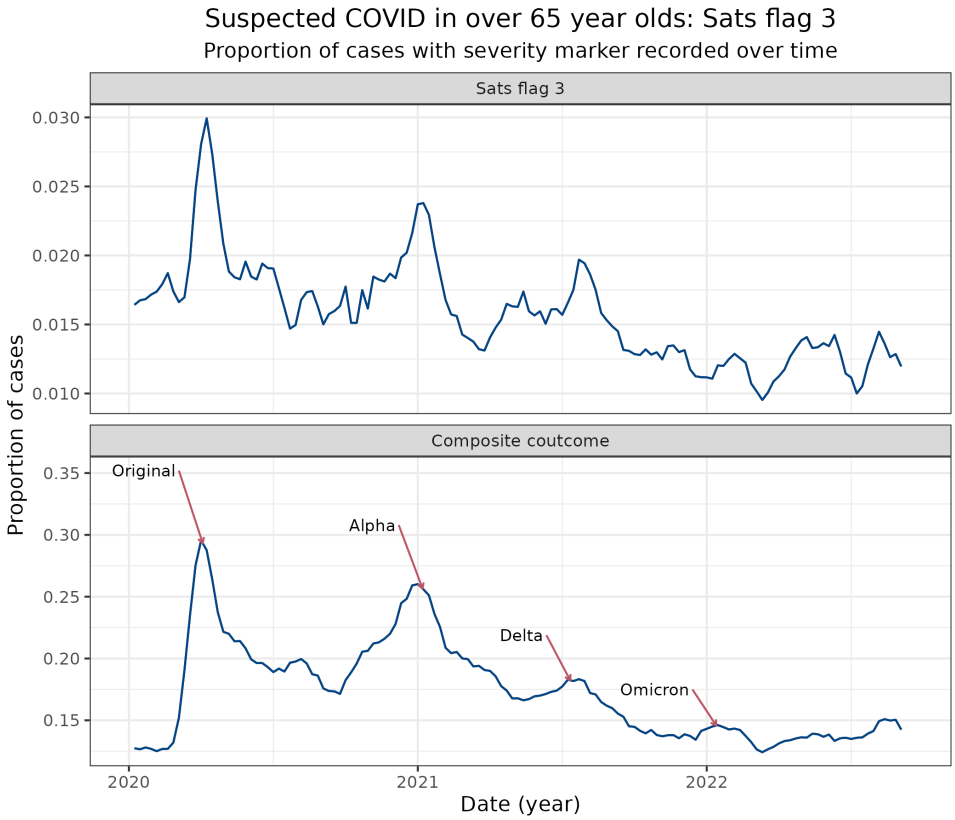


Figure 5.13: Example indicator: Pandemic. TOP PANEL: Weekly proportion of cases of suspected COVID in over 65 year olds for whom the oxygen saturation (sats) Flag 3 (less than 92%) was present. BOTTOM PANEL: weekly proportion of the same cases with the composite outcome recorded. Arrows on bottom panel indicate waves of COVID infection, labels indicate the dominant variant at that time. Original: Wuhan, 'wild' type

5.4.4 Comparison of non-pandemic and pandemic indicators

Geometric means of indicators that were retained for suspected COVID in the pandemic period were on average higher than those observed for non-pandemic respiratory infections. The median GM for suspected COVID in the pandemic was 3.97 (IQR 3.51–6.09), compared with 3.21 (IQR 2.77–3.65) for non-pandemic LRTI-ECLD and 2.91 (IQR 2.83–2.94) for non-pandemic URTI (Table 5.15). This suggests, in general, that indicators during the pandemic had a stronger association with outcome.

Table 5.15: Summary of GM and SNR by ARI subtype

ARI subtype	GM (median [LQ–UQ])	SNR (median [LQ–UQ])
LRTI-ECLD	3.21 (2.77–3.65)	2.53 (1.48–3.95)
URTI	2.91 (2.83–2.94)	2.37 (1.79–3.44)
suspected COVID	3.97 (3.51–6.09)	6.80 (3.36–16.94)

Notes: Values represent geometric means for severity markers that passed filtering. ARI: acute respiratory infection, IQR: inter-quartile range, Min: minimum, Max: maximum, URTI: upper respiratory tract infection, LRTI: lower respiratory tract infection, ECLD: exacerbation of chronic lung disease. GM: geometric mean, SNR: signal to noise ratio. LQ: lower quartile, UQ: upper quartile.

The SNRs of indicators that were retained for suspected COVID-19 were also on average higher than those observed for non-pandemic URTI and LRTI. The median SNR for suspected COVID-19 in the pandemic was 6.80 (IQR 3.36–16.94), compared with 2.53 (IQR 1.48–3.95) for non-pandemic LRTI-ECLD and 2.37 (IQR 1.79–3.44) for non-pandemic URTI (Table 5.15). This suggests, in general, that the signal strength of indicators during the pandemic was greater than for non-pandemic respiratory infections.

Table 5.16 summarises severity indicators performance across all strata. The ordering is based primarily on the number of strata in which each indicator met the predefined significance and correlation thresholds (as detailed in Section 5.3.8), and secondarily on the average geometric mean (GM) observed across those strata. Notably, 'Any symptom' (including malaise, dyspnoea, haemoptysis, fever and confusion) was retained in five of the eleven strata.

Table 5.16: Geometric mean values of predictors by all strata, ordered by frequency and average GM.

Predictor	Non-pandemic				Pandemic		Count	Avg GM
	URTI		LRTI-ECLD		Suspected COVID			
	<15	15–64	15–64	65+	15–64	65+		
Any symptom	—	2.65	2.60	2.83	3.68	3.39	5	3.03
Dyspnoea	—	—	2.65	2.67	3.95	2.83	4	3.03
Co-amoxiclav	—	2.93	3.17	—	3.80	3.69	4	3.40
BP Flag 2	—	2.77	3.62	4.48	—	—	3	3.62
BP Flag 1	—	2.91	3.40	4.23	—	—	3	3.51
Malaise	—	2.94	—	—	3.37	3.97	3	3.43
CXR request	—	—	3.16	2.54	3.15	—	3	2.95
O_2 sats Flag 2	—	—	4.45	3.76	—	—	2	4.11
CRP	—	—	3.37	—	—	4.41	2	3.89
O_2 sats Flag 1	—	—	3.38	3.26	—	—	2	3.32
Fever	—	—	—	—	3.12	3.51	2	3.32
Prednisolone	2.90	—	2.93	—	—	—	2	2.92
Ambulance	—	—	—	—	7.67	7.17	2	7.42
NHS 111	—	—	—	—	4.44	4.60	2	4.52
Hospital referral	—	—	—	—	9.50	—	1	9.50
O_2 sats Flag 3	—	—	—	—	—	6.27	1	6.27
Any sign Flag 3	—	—	—	—	—	6.13	1	6.13
Confusion	—	—	—	—	—	6.09	1	6.09
Chest signs	4.23	—	—	—	—	—	1	4.23
Haemoptysis	—	—	—	—	4.22	—	1	4.22
Any sign Flag 2	—	—	—	3.82	—	—	1	3.82
Any sign Flag 1	—	—	—	2.80	—	—	1	2.80

Only six severity indicators were retained in both pandemic and non-pandemic strata: Any symptom (GM=3.03), Dyspnoea (GM=3.03), Co-amoxiclav (GM=3.40), Malaise (GM=3.43), CXR request (GM=2.95), and C-reactive protein (GM=3.89). Oxygen saturation flags also appeared in both contexts but at different levels: Flags 1 and 2 were observed in non-pandemic LRTI-ECLD (15–64 years and ≥65 years), while Flag 3 was prominent in suspected COVID during the pandemic. The average GM for oxygen satura-

tion markers (Flag 1 = 3.32, Flag 2 = 4.11, Flag 3 = 6.27) was notably higher than for the other six severity markers occurring in both pandemic and non-pandemic strata.

The indicators with the highest GMs generally did not appear in both pandemic and non-pandemic periods. For example, Hospital referral (GM=9.50), Ambulance (GM=7.42), and Oxygen saturation Flag 3 (GM=6.27) occurred only in the pandemic strata, whereas Chest signs (GM=4.23) and Oxygen saturation Flag 2 (GM=4.11) were limited to the non-pandemic period.

5.5 DISCUSSION

5.5.1 Summary of main findings

In this chapter, I used primary care data from the ORCHID to evaluate the severity markers identified earlier in this DPhil. I examined whether markers recorded at or near the time of ARI presentation were associated with subsequent severe outcomes, and I ranked predictors by the combined strength of their individual-level and weekly aggregate associations with these outcomes.

I identified four main findings in this chapter. (1) Several predictors showed potential to serve as timely severity indicators for ARI surveillance. The composition of these indicators differed between the pandemic and non-pandemic periods: clinical signs such as blood pressure dominated the non-pandemic period, whereas symptoms and health-care utilisation markers were more prominent during the pandemic. (2) Some indicators, including symptoms, and some grouped sign thresholds and oxygen saturation were consistent across both periods. (3) After filtering (described in Section 5.3.8), no severity indicators remained for ILI in any age group, for URTI among those aged 65 years and over, and only one for LRTI in children under 15 years of age. (4) Outcome severity during the non-pandemic period peaked in summer rather than winter, an unexpected finding that may reflect environmental, patient, pathogen, or health-system factors.

5.5.2 Results in context

Previous studies

To the best of my knowledge, this study is the first to comprehensively assess candidate severity indicators as alternatives to the CHR and CFR for ARI surveillance. The identification of timely, primary care–based indicators could help overcome many of the challenges associated with these traditional measures. Use of such indicators would represent a substantial departure from previous surveillance practice, including that outlined in the WHO's PISA framework [23]. Relevant prior work falls into two main strands: clinical prediction models and other surveillance studies.

Primary care–based studies have shown that a number of clinical parameters can predict severe outcomes to some extent. These include RECAP [144], GRACE [145], and STARWAVe [146], which reported predictors of adverse outcomes including breathlessness, low oxygen saturation, diastolic blood pressure, crackles on auscultation, sputum severity, short illness duration ($\leq$ 3 days), vomiting, fever, clinician-observed subcostal recession, and wheeze [144–146]. Many of these parameters were also identified in the individual-level analysis of my study. These studies, however, are clinical models and therefore didn't assess aggregate-level association.

Some surveillance studies have undertaken aggregate-level analysis of primary care data, but these have tended to examine whether primary care ARI or COVID-19 incidence can act as early indicators of subsequent surges in hospitalisations or ICU attendance. The underlying rationale is that viral respiratory infections such as influenza can predispose individuals to secondary bacterial infections and complications, or exacerbate chronic lung disease, leading to more severe outcomes. Some studies indicate ILI consultations lead hospital admissions by approximately two weeks, supporting the potential of primary care data as an early warning signal for severe respiratory disease [147, 148]. Similarly, the CDC's FluSight programme publishes forecasts of influenza hospital admissions using multiple models. Because these approaches use population-level denominators rather than ARI case denominators, it is not possible to know whether surges in hospitalisations result

from an increase in severity or an increase in incidence. This is where severity indicators could add additional value.

Candidate indicators

The candidate indicators identified for the non-pandemic period for LRTI-ECLD were dominated by clinical signs flags. These included blood pressure, oxygen saturation and also 'any sign' flags. The pulse rate and respiratory rate flags did not appear, as none of these were associated with outcomes at the aggregate level. Although temperature flags were associated at the individual-level, they were negatively associated at the aggregate level. This could be because larger volumes of milder febrile infections occur in the winter.

Indicators based on clinical signs are appealing because they are objective and are commonly recorded at the time of the consultation, meaning they are likely to be particularly timely. 'Any sign' flags are a particularly attractive option as they consider multiple signs which increases robustness to reduced data recording and is supported by validation studies of the NEWS2 score. Because clinical signs are based on standardised measurement and widely used thresholds, they are less susceptible to variation in patient reporting, clinician interpretation, or cultural and language factors. This makes 'any sign'-based indicators less vulnerable to missingness or measurement error in any one sign, and more resilient to variation in recording practice. However, sign-based indicators performed less well during the pandemic, reflecting the reduced recording of the signs rather than their intrinsic association with severe disease. Innovation to allow primary care practitioners to assess clinical signs during remote consultations could help support use of clinical signs during future pandemics.

During the pandemic, the indicators were more heterogeneous, but included significantly more symptoms including: confusion, malaise, fever, any symptom and dyspnoea. Additionally, health care utilisation indicators were highly ranked including: ambulance, NHS 111 and emergency hospital referral indicators. Recording of health care utilisation indicators increased substantially compared to the non-pandemic assessment, increasing

from 1.5% to 4.1% for NHS 111 and 1.5% to 3% for ambulance encounters. The reduced number of clinical signs and increased symptoms probably reflects the shift to remote consulting, where clinical signs were not easily assessed and there was an increased reliance on the subjective patient history such as symptoms. Health care utilisation indicators also reflects the dramatic shift in how health services were accessed during that time.

Pandemic severity indicators generally had higher SNR and GMs than those observed in the non-pandemic period. Part of this difference may reflect the aggregation of all ARI subtypes into a single suspected COVID-19 category, which increased the total number of cases per unit time and thereby reduced random variability, leading to higher and more stable SNR values. The higher GMs may indicate stronger overall associations with severe outcomes. This likely reflects both the intrinsically greater severity of COVID-19 compared with most non-pandemic respiratory infections, and the clearer clinical and coding signals during the pandemic.

This contrasts with ILI, for which no indicators survived the filtering process. Part of the challenge with ILI is that it is a surveillance-specific construct rather than a widely used clinical term. Moreover, ILI episodes are generally less common than broader categories such as URTI or LRTI, reducing the statistical power to detect associations. The smaller number of cases also increases variability, leading to lower SNR values. As ILI has historically been an important surveillance category, this is a notable limitation; however, it does not preclude the use of severity indicators identified in other ARI subtypes in future pilots. Similarly, fewer severity indicators met the filtering criteria for URTI, likely reflecting the intrinsically milder nature of URTIs and their weaker association with severe outcomes.

Some indicators occurred across multiple strata and were present in both pandemic and non-pandemic periods. However, these were generally not the indicators with the highest GMs. For example, ‘Any symptom’ appeared in five strata with an average GM of 3.03, and Dyspnoea and Co-amoxiclav were observed in four strata each with average GMs of 3.03 and 3.40, respectively. Similarly, Malaise (GM=3.43) and CXR request (GM=2.95) were

identified in both periods, whereas indicators with higher average GMs, such as Hospital referral (GM=9.50), Ambulance (GM=7.42), and Oxygen saturation Flag 3 (GM=6.27), were only seen during the pandemic. These cross-period indicators therefore appear resilient to changes in healthcare practice, including increased remote consulting, but were in general less strongly associated with the outcome.

Timing of severity peaks

One of the more surprising findings of this study was that outcome severity typically peaked during the summer months in the non-pandemic period (Figure 5.10). This is in contrast to the incidence of ARI that is known to peak during the winter. There are a number of plausible explanations for this, including environmental, patient, pathogen and health-system factors.

Environmental: Some studies, including one from Spain in 2023, showed a similar summer peak in outcomes among hospitalised patients with respiratory disease and attributed this to high ambient temperatures during that time of year [149, 150]. Heatwaves in the UK are associated with excess mortality and increased hospitalisations but further work is required to establish this link more robustly [151, 152]. *Patient factors:* Firstly, although there are fewer infections in the summer, those that do occur could be disproportionately in high-risk groups such as those with chronic lung disease and older populations. Therefore fewer infections occur but with worse outcomes. Health seeking behaviours are also known to vary throughout the year, for example, people tend to have a higher threshold for seeking healthcare over Christmas, this may also be the case during the summer holidays. Therefore those that do attend primary care tend to be more unwell. *Pathogen:* The pathogen mix in summer may also have an impact, with low levels of circulating viruses compared to bacteria that may cause a greater proportion of infections and result in worse outcomes. *Health-system factors:* Factors that may contribute to this summer peak are seasonal variations in admission thresholds. In the winter, due to pressure on the health service the threshold for admission may be higher reducing the proportion of cases overall that go on to be hospitalised [153]. Finally workload pressures during winter in primary

care may also dictate the extent to which outcomes are recorded in primary care [154]. It is conceivable that in the winter months when primary care is at its busiest, less time is dedicated to clinical coding. This may underestimate severity at this time.

Severity indicators should therefore be interpreted alongside incidence, as high severity during periods of low incidence may not translate into substantial health system burden. Defining seasonal baselines or thresholds for severity indicators, similar to those used for incidence, may therefore support more meaningful interpretation across the year.

In contrast, during the pandemic, peaks in severe outcomes aligned with SARS-CoV-2 waves (Figure 5.13). This in part reflects the fact that other viruses were relatively suppressed as a consequence of reduced social mixing. This results in a cleaner severity signal during the pandemic. The highest peak in severity occurred during the initial 'wild type' wave in April 2020 where outcomes occurred in nearly 30% of all ARI cases in over 65-year-olds. This is closely followed by the peak in early January 2021 caused by the Alpha variant where outcomes occurred in around 25% of all cases. Small peaks in severity also occurred with the Delta and Omicron variants. The order of these peaks is generally supported by evidence where it is clear the first two waves were the most severe, and later waves in general diminished in severity as population immunity and vaccination programmes were instituted [155]. It is important to point out that comparing severity by wave is challenging due to the evolving testing, management and immunisation practices.

5.5.3 Limitations

Syndromic surveillance: The identification of summer peaks in severity during the non-pandemic period highlights a key limitation of syndromic rather than pathogen specific surveillance. Without virological confirmation, we do not know which pathogen causes a given ARI, and the circulating mix of pathogens will likely shape severity. For example, if respiratory viruses circulate less in summer, a larger share of presentations may be bacterial pneumonias in higher risk individuals, which could drive higher severity despite lower overall incidence. However, irrespective of the pathogen mix, these signals still tell

us how often people with ARIs progress to severe outcomes. As noted earlier, defining severity thresholds or seasonal baselines would support interpretation at different times of year. A rise above the expected level in winter may have a greater impact on health services because more infections are occurring, even if the absolute value of the severity indicator is lower than the summer peak.

Further analysis of the factors that influence the timing of severity peaks would help our understanding of severity indicator values. In particular, relating peak timing to case mix (for example, age and risk group) and to pathogen mix (dominant circulating pathogens) in each period can clarify drivers of severity in syndromic surveillance and improve interpretation of indicator values over time.

Pandemic based indicators: A number of potential severity indicators were identified during the pandemic period, but there is no guarantee they would be suitable in a future emergency. That period was dominated by a single pathogen, and the severity peaks we observed aligned with waves of SARS-CoV-2. A future event may involve a different pathogen with different symptoms and may have a different impact on care pathways. These shifts could change which indicators are most informative and how consistently they are captured.

This highlights the importance of innovation to identify objective clinical signs that continue to flag individuals at risk of severe outcomes, and to enable measurement of clinical signs in the home setting. In the short term, to improve resilience and preparedness, it would be sensible to define a compact panel that goes beyond clinical signs alone, combining a small set of reliable symptoms and measures of health care use, and to reassess performance as conditions change. However in the future emerging consumer technologies, such as home pulse oximeters and wearable devices capable of monitoring clinical signs, may help maintain the data quality of clinical sign data during remote consulting.

Methodological approach: Surveillance is, by definition, prospective and ongoing. In this analysis I used retrospective data to evaluate candidate severity indicators. Findings

may not fully bear out in a prospective setting, because some information is added to clinical records after the initial consultation. Retrospective and prospective data therefore are unlikely to be identical. A prospective pilot is needed to assess real-time recording, timeliness, and validity of the indicators under operational conditions.

Such a pilot should include the severe outcomes used here and, in parallel, a core set of clinical signs, a small number of symptoms, and measures of health care use. This would allow assessment of the timeliness and completeness of each predictor relative to the occurrence of severe outcomes, and help refine the structure and thresholds of the surveillance indicators before scaling up.

A number of potential sensitivity analyses could be considered as part of an additional validation step. These could include assessing associations with outcomes using different time windows to the event; for example, rather than using 56 days, periods of 28 or 14 days could be tested. Another approach could involve assessing the inclusion of predictors that occurred only on the same day as the ARI event. Furthermore, alternative cut offs for clinical signs could be evaluated to test the robustness of associations. Because of the underlying data structure, undertaking these analyses is likely to be challenging.

5.5.4 Summary

For this study I undertook a comprehensive evaluation of the candidate severity indicators identified in earlier chapters. I used a novel ranking framework to prioritise those with the strongest association with severe outcomes. The analysis indicates that a compact set of indicators centred on objective clinical signs, complemented by a small number of symptoms and health care utilisation measures, could form the basis of a practical, timely severity indicator set. This approach is likely to be more robust to increases in remote consulting as occurred during the pandemic. However, innovations in how clinical signs are recorded during telephone and video consultations are likely to be required to improve data quality and optimise preparedness for future pandemics. Finally, defining baselines or thresholds for severity indicators could be a practical way to interpret indicator values

given their highly seasonal nature. In the final chapter of my DPhil, I describe a clear and actionable path from retrospective analysis to prospective, timely severity surveillance for ARIs.

Chapter 6

Final discussion

6.1 INTRODUCTION

In the final section of this thesis, I summarise the work that has been undertaken and the main findings in relation to the original objectives. I also discuss the implications for surveillance and for the wider public, focusing on three principal themes: epidemic and pandemic preparedness, data quality and timeliness, and innovation and technology. Specifically, how epidemic and pandemic preparedness is a central aim of CMR-based surveillance, but it is limited by issues of data quality. Innovation is required to overcome these challenges and unlock its full potential for public health benefit. I go on to critically reflect on the strengths and weaknesses of the methodological approaches used in this DPhil. I then set out the critical next steps in the development of primary care CMR-based severity indicators for ARIs. Finally, I round off the chapter with a closing summary.

The work undertaken in this DPhil aimed to support timely public health surveillance of ARIs through the development and evaluation of severity indicators derived from primary care CMRs. Specifically, the work sought to complement incidence figures with reliable, timely markers of severity and to strengthen readiness for both seasonal and pandemic threats.

Primary care data were used to address challenges that limit severity reporting from other sources, namely a lack of timeliness, small sample sizes, and case ascertainment bias [23, 28–31]. The work was structured into four main chapters, each addressing a distinct objective:

1. **Develop** and **validate** a digital phenotyping algorithm to identify cases of ARI from primary care CMRs.

2. **Identify** candidate markers of severity present in primary care CMRs.

3. **Assess** the data quality of these markers, including recording rates and temporal dynamics.

4. **Evaluate** the identified markers at the individual level and at the weekly aggregate level to determine the severity indicators with the greatest potential for prospective surveillance.

6.2 SUMMARY MAIN FINDINGS

This section concisely presents the core outputs of the DPhil across its four main chapters. It highlights what was done, the main findings and brief implications for surveillance practice including the surveillance challenges that have been addressed.

6.2.1 Chapter 2: Case identification

What was done: Development and validation of an ARI digital phenotyping algorithm for primary care CMR-based surveillance. I used the SNOMED polyhierarchy with Expression Constraint Language to create clear, rule based intensional code lists. ARI case detection forms the foundation of severity indicator development.

Main findings: Increased ARI case detection and accuracy by including symptom codes, ECLDs, and ARI-NOS, and excluding chronic disease codes. Creation of a shareable and transparent algorithm.

Brief implications: Increased case detection sensitivity helps to address the issue of small sample sizes. Measuring severity in a broader range of cases may also help reduce case ascertainment bias. Greater case identification reduces incidence and severity indicator volatility providing clearer signals and enhanced outbreak detection. Taken together this

new PhA contributes to enhanced epidemic and pandemic preparedness. The algorithm is peer reviewed, openly available, and integrated into the RSC workflow informing UKHSA, ECDC, and WHO reporting.

6.2.2 Chapter 3: Candidate severity markers

What was done: A systematic review to identify candidate severity markers suitable for primary care surveillance. Severity markers were grouped into severe outcomes (for example hospital attendance, admission, ICU admission, complications, death) and predictors of severe outcomes (for example symptoms, clinical signs, treatments, health care utilisation).

Main findings: Thirty severity markers were identified: 7 severe outcomes and 23 predictors of severe outcomes. Severe outcomes are less timely and are often recorded with a delay due to reliance on hospital discharge summaries. These served as the outcome in the later evaluation. Predictors are more likely to be recorded at or near the index ARI consultation and therefore can provide timelier estimates of severity. These formed the candidate set tested against the outcome.

Brief implications: A set of candidate severity markers including possible outcomes and timely predictors for testing in subsequent sections of this DPhil was defined. This work has been peer reviewed and published and can also be used by others considering defining severity indicators in different surveillance settings. This helps to address the challenge of timeliness of severity indicators for ARI.

6.2.3 Chapter 4: Data quality of severity markers

What was done: A data quality assessment using a large CMR repository based at the RSC focusing on the completeness of recording and temporal dynamics of severity marker recording.

Main findings: Clinical sign recording was initially low but rose to about 55 percent of

ARI presentations in the post-pandemic period. Recording was more likely when perceived severity or risk was higher, which means sign-based indicators may overestimate severity. During the COVID-19 pandemic, sign recording fell from 47 percent to 29 percent and symptom recording rose from 4 percent to 7 percent.

Brief implications: Completeness of recording varies by ARI subtype, age, and risk status, so indicators should account for this. Indicators can become vulnerable to failure when recording practices change, particularly during pandemics, and sign-based indicators may overestimate severity in the primary care population. The work helps quantify known issues of data quality which can act as a catalyst for innovation to improve data quality and thereby support provision of more complete and accurate intelligence to UKHSA.

6.2.4 Chapter 5: Evaluation of markers

What was done: A stratified individual and aggregate-level analyses assessing the association between predictors and severe outcomes were performed in order to rank severity indicators according to association with outcomes.

Main findings: I identified four main findings in this chapter. (1) Several predictors showed potential as timely severity indicators, with clinical signs being most important in the non-pandemic period and symptoms and health care utilisation markers being more prominent during the pandemic. (2) A small set of indicators, particularly symptoms, grouped sign thresholds and oxygen saturation, were consistent across both periods. (3) No indicators were retained for ILI, none for URTI in older adults and only one for LRTI-ECLD in children. (4) Severity in the non-pandemic period peaked in summer rather than winter, likely reflecting environmental, patient, pathogen or health system factors.

Brief implications: Severity indicators derived from the primary care CMR have potential to support timely surveillance of ARIs. The optimal indicators differ between pandemic and non-pandemic periods, in part, because of changes in data quality that occurred during the pandemic. Some were consistent across both periods. A set of symptom, clinical sign

and health care utilisation-based severity indicators could form the basis of a set for further prospective evaluation.

6.2.5 Overall contribution

Primary care CMR based severity indicators have potential to enhance surveillance of ARIs through data driven resource planning and public health decision making, thus contributing to epidemic and pandemic preparedness. Further prospective piloting work and innovation is required to improve data quality and operationalise these indicators.

6.3 IMPLICATIONS OF FINDINGS

In this section, I interpret the findings of the DPhil with particular consideration of three recurrent themes: epidemic and pandemic preparedness, data quality, and innovation and technology. I also relate the findings to the challenges that motivated this work, specifically timeliness, small sample sizes, and case ascertainment bias.

6.3.1 Integrated surveillance: defining ARI for a post-pandemic context

The new ARI digital phenotyping algorithm provides a machine-processable, flexible, and transparent approach to case definition that supports integrated ARI surveillance [44]. It enables automated identification of ARI episodes from routinely collected primary care CMRs using a single rule-based algorithm built on SNOMED's ECL. This replaces the legacy system of multiple syndrome-specific algorithms with a coherent phenotype that captures the full range of clinical presentations associated with respiratory viruses and bacteria. In line with open science, it is published and shared, and is now integrated into the RSC surveillance workflow reports to UKHSA, ECDC, and WHO. It also defined the ARI cohort used throughout this DPhil.

Although unified, the algorithm retains the ability to consider syndromes separately and maintains historically important indicators. Continuity of the ILI indicator remains valu-

able, as it has traditionally served as an influenza-specific measure because of its higher specificity for laboratory confirmed influenza infection [156]. However, this distinction is evolving, as COVID-19 often presents with a clinical picture similar to influenza [157].

Incidence and severity estimates in populations with small sample sizes can be noisy, which makes early detection of meaningful signals challenging [158]. The new algorithm increases case detection through inclusion of relevant symptoms and ECLDs, which can reduce noise in incidence signals and increases confidence when interpreting changes [44]. It also indirectly reduces noise in severity indicators because a larger denominator yields more severe events. Together, these changes make the system more sensitive to small temporal shifts and support more timely identification of abnormal signals. Taken together, reduced signal noise and increased syndrome breadth strengthen epidemic and pandemic preparedness by enabling earlier and more reliable detection of threats from respiratory pathogens.

This framework is well suited to the post-pandemic landscape, where epidemic and pandemic prone respiratory viruses such as SARS-CoV-2 variants can vary substantially in their clinical characteristics [159]. Within the context of this DPhil, the integrated approach allows severity to be assessed across ARI subtypes. Some syndromes map more closely to certain pathogens, for example ILI has historically been more specific for influenza, and bronchiolitis for RSV in young children [156, 160]. However, presentations remain highly variable. Individuals infected with different pathogens may exhibit similar symptom patterns, and the same pathogen result in varied presentations in differing individuals. This underscores the importance of viewing primary care CMR-based syndromic surveillance as one component within a broader system that also includes laboratory and hospital-based surveillance.

6.3.2 Data quality

Data quality is central to the reliability of any surveillance system and determines the extent to which changes in indicator signals reflect true variation in disease severity rather

than artefacts of recording practice [161]. The analyses in this work highlight that the completeness of recording varies by indicator type, age, and ARI subtype. Clinical sign recording, for example, has increased substantially over time but remains incomplete, with approximately half of all ARI presentations now including at least one recorded sign. Recording is also non-random, being more likely when clinicians perceive greater severity or risk. Whilst assessment of severity in primary care can reduce case ascertainment bias compared with hospital-based measures, bias in the recording of clinical signs may still lead to overestimation of severity in the primary care population.

During the COVID-19 pandemic, the quality of clinical data changed markedly. Recording of signs fell from 47 percent before the pandemic to 29 percent during it, while symptom recording increased from 4 to 7 percent. This almost certainly reflects the shift toward remote consulting, which reduced opportunities to measure vital signs directly. Consequently, fewer clinical sign-based indicators were ranked highly during the pandemic. This is not because they were intrinsically unassociated with the outcomes, but because they were under-recorded. Understanding and accounting for such patterns is essential when interpreting apparent changes in indicator performance across differing contexts. It should be noted that CMR data quality issues during the pandemic were not limited to primary care [162] having also been noted in hospital settings as well.

Maintaining data quality during times of operational disruption requires proactive communication and coordination between stakeholders, including clinicians, data engineers, informaticians, and public health authorities [163]. Early dialogue at the onset of a public health emergency can ensure that updates to clinical systems, terminology sets such as SNOMED, and the phenotyping algorithms themselves are made rapidly and consistently across platforms. This multidisciplinary communication is essential for maintaining the resilience of the surveillance infrastructure and ensuring that changes in clinical practice do not compromise the performance of surveillance indicators. As we work towards greater epidemic and pandemic preparedness, this coordination becomes even more important, as reliable indicators are most needed when the risk of signal failure is greatest. Sustained

innovation will also be critical to strengthening data quality and preventing such failures in the future.

6.3.3 Indicator selection

Severe outcomes and their predictors

The severity indicators used in this work were classified into two groups: concrete severe outcomes that are reported with delay, such as hospitalisation and death, and timely predictors that are typically available at the time of diagnosis and are associated with severe outcomes. Whilst this DPhil has aimed to identify timely indicators, both severe outcomes and predictors have value.

Severe outcomes can serve as retrospective measures of severity, for example, in end of season reports and vaccine effectiveness studies where timeliness is less critical [12]. Predictors, on the other hand, are likely to be more valuable as early warnings of increased severity in prospective, near real-time surveillance systems. These therefore have potential to influence the public health response to emerging threats. Together, severe outcomes and their predictors add value to surveillance systems and both should be considered important in any future work in this field.

Innovation should also play a key role in the future development of severity indicators. For example, making hospitalisation and death data more readily available would allow concrete outcomes to become more timely predictors. If primary and secondary care systems were better linked, an automated hospital admission record could be embedded in the primary care CMR at the point of admission, helping to reduce reporting delays. However, even if this were achieved, it would not overcome the inevitable delay between a primary care encounter and a subsequent hospital admission. Predictors of severe outcomes would therefore continue to be the most timely severity indicators.

Choice of specific indicators

Twenty-two indicators were retained after filtering out those not associated with the outcome and those that had high volatility. Identifying which of these indicators would be the most robust to carry forward was challenging, because they behaved differently in different strata and in the pandemic and non-pandemic periods. Indicators present in multiple strata have potential to be more flexible. For example, 'Any symptom' was retained in five strata in both the pandemic and non-pandemic periods. This could therefore serve as a flexible severity indicator across contexts.

In the non-pandemic strata numeric clinical signs were, in general, the highest ranked indicators. This included blood pressure and oxygen saturation thresholds, as well as a composite sign threshold that considered blood pressure, pulse rate, respiratory rate, and temperature using cut offs based on the NEWS2 score [84]. Oxygen saturation Flags was the only clinical sign that appeared in both pandemic and non-pandemic periods, albeit at different NEWS2 levels.

Clinical signs are attractive indicators because they are standard measures used to assess and communicate the severity of acute illness. They are objective, and there is substantial evidence that they predict adverse outcomes at the patient level [84]. Their completeness can also be measured directly, which facilitates routine monitoring of data quality. Most importantly, because clinical signs are often recorded at the time of an ARI consultation, they offer a practical way to address the timeliness challenge that motivated this DPhil.

Symptoms and healthcare utilisation markers ranked more highly in the pandemic analyses. Prioritising these non-sign indicators could make severity surveillance more resilient to recording changes driven by remote consulting. Although health care utilisation markers were not retained in the non-pandemic analysis this could reflect the relatively small amount of post-pandemic data available in the study and the recording quality of these could have now increased in the aftermath of the pandemic. Future piloting work should retain these indicators to allow continued monitoring of their suitability in non-pandemic contexts.

With respect to other predictor groups, co-amoxiclav and prednisolone were the only treatments retained, with co-amoxiclav appearing in four strata (both pandemic and non-pandemic) and prednisolone in two strata. Chest X-ray request and CRP request were also retained in the investigation group.

Co-amoxiclav is not a first line antibiotic for the treatment of ARIs in primary care, therefore, its prescribing is likely to occur in higher risk cases or in those that have not responded to first line antibiotics. Interestingly, other antibiotics (particularly amoxicillin and doxycycline) were often negatively associated with the outcome at both the individual and aggregate levels. The reasons for this are not entirely clear but may reflect the effectiveness of these antibiotics in preventing adverse outcomes or, alternatively, relate to over prescribing in milder cases.

The retention of prednisolone in younger age groups may reflect that these individuals often have existing asthma, which could increase their likelihood of hospital admission during an ARI episode. Although investigation requests are somewhat more timely than severe outcomes, they are slightly less appealing as indicators because they are unlikely to occur at the exact point of contact with a healthcare professional due to short delays in receiving results.

Notably, no severity indicators were retained in five of the eleven strata, including all ILI age bands, URTI in adults aged over 65 years, and LRTI-ECLD in children under 15 years. For ILI, this may reflect the limited case load (only approximately 2% of all ARIs in the non-pandemic period). This makes identifying associations more challenging. Whilst many associations were evident at the individual level, few remained at the aggregate level, and those that did were highly volatile. As ILI is an important indicator, this was a disappointing finding. ILI is not a term or diagnosis commonly used by clinicians, as it is primarily a specialist surveillance concept. Clinicians with greater familiarity with surveillance are likely to record it more consistently. The RSC can help support practices in the appropriate use of ILI codes through education and communication.

My recommendation would be further prospective evaluation of all symptoms, numeric clinical signs and healthcare utilisation indicators, and also inclusion of all severe outcomes.

Outcome indicators during the pandemic period followed a predictable pattern, with peaks in outcomes coinciding with known waves of SARS-CoV-2 infection in England. Strong correlations between outcomes and the top ranked predictors during this period suggest that these markers provided useful information for public health authorities about changes in rates of severe outcomes in the community. In the non-pandemic period, however, peaks in severity were seasonal, with the greatest severity occurring in the summer months. This finding was somewhat unexpected but may reflect a combination of high risk individuals being most susceptible to infection at this time, ambient temperatures, and higher rates of bacterial pathogens, although further work is required to establish the underlying drivers.

6.4 STRENGTHS AND LIMITATIONS

Here I focus on the strengths and limitations of the work undertaken in this DPhil. I critically appraise the data, methods and inferences made from the studies included and how this impacts on their real-world utility.

6.4.1 Data and study setting

Real-world surveillance

These studies were undertaken within a real-world surveillance system, which increases potential to practically apply the outputs from this work. As a case in point, this setting enabled direct implementation of the ARI PhA into routine reporting, so it now contributes to RSC surveillance outputs to UKHSA and beyond. Data flows from EMIS practices to the RSC are timely, with a lag of about two to four days. This supports compilation of prompt incidence figures and offers a suitable test bed for developing timely severity indicators was a core objective of this DPhil.

The present analyses were retrospective which doesn't completely reflect the real-world system, which operates prospectively. In order to ensure the data I used better reflected the real-world setting, I limited the analysis to records (diagnoses of ARI and severity markers) that were recorded and entered on the same day. This theoretically reduces the chance of retrospectively entered data being included in the analysis. However, this approach still relies on clinicians appropriately back-dating certain details when a retrospective record is entered, which, from personal experience, does not always happen. As a result, some records that appear prospective may in fact describe events that occurred earlier. Further prospective piloting is therefore required, and the RSC environment is a suitable place for such work.

Primary care CMR data are rich, which allowed exploration of a wide range of candidate severity markers. Some additional metadata would further strengthen interpretation. In particular, flags indicating consultation mode (face to face, remote) and whether entries derived from secondary sources such as hospital discharge summaries would help distinguish contemporaneous from delayed recording and reduce inclusion on non-relevant records. These data are theoretically available, but have not yet been extracted or engineered by colleagues in the technical team, and are therefore currently unavailable for analysis.

Representativeness

The dataset is largely representative of the UK, comprising 25.7 million ARI episodes in nearly 9 million individuals [40]. Because most of the population is registered with a general practice, this surveillance has potential reach across the vast majority of people and includes a range of socioeconomic and ethnic groups. Nonetheless, not everyone attends general practice, and people who are not registered are disproportionately among the least advantaged, for example people experiencing homelessness or those in prison. In addition, current RSC ARI surveillance aggregates by subtype, age band, and NHS region and does not yet integrate ethnicity or socioeconomic status. This reflects a deliberate balance between simplicity and granularity in operational surveillance. As computational

capacity and data engineering mature, integration of key demographic features may be feasible. It is also relevant that these analyses draw on EMIS practices only, due to earlier data governance constraints. EMIS coverage is strongest in the West of England, London, and much of the South, which slightly limits regional generalisability. With governance issues now resolved, future work can extend to non-EMIS practices.

Free text versus coded data

This DPhil used only coded data from the primary care CMR because free text is not routinely available for secondary use. Access to free text is restricted because it may contain sensitive information that risks patient privacy and security. This has implications for measuring completeness, both for ARI diagnostic codes and for severity marker recording. Although diagnostic (problem) codes are generally well recorded, primary care staff often use free text to capture nuanced details of a consultation. Consequently, the completeness estimates reported here cannot fully reflect the extent to which these concepts are documented in the CMR. Reassuringly, recording of several key severity markers, particularly clinical signs, has increased markedly over the last decade. Advances in NLP and large language model (LLM) methods for de-identification and information extraction could enable safe use of free text and improve the sensitivity and timeliness of severity indicators. Robust governance and regulatory mechanisms are needed to facilitate this.

6.4.2 Reference standards

ARI case detection

Absence of a definitive reference standard in primary care CMR data emerged repeatedly during this DPhil. To validate the new ARI PhA, I benchmarked it against the existing RSC algorithm, recognising that the CMR alone cannot determine with certainty which encounters truly meet the inclusion criteria for an ARI episode. To minimise subjectivity, the PhA and its associated code lists were specified a priori, independently reviewed, and peer reviewed, providing transparent scrutiny and supporting their validity. Using ARI cases with virological confirmation would allow severity to be characterised for pathogen

specific cohorts, but this is not a solution for syndromic surveillance, which is intentionally agnostic to laboratory confirmation.

Severity markers

Lack of reference standards during the data quality assessment also limited the extent to which completeness could be measured. In the absence of a reference standard, completeness cannot be measured directly for binary indicators such as symptoms. To aid interpretation, I introduced three different data entry contexts for severity markers: (1) Markers where completeness can be validly inferred from recording rates, for example numeric clinical signs. Then, (2) structurally complete markers that should always be recorded when they occur because the system enforces or automates entry (for example antibiotic prescribing). (3) Finally, structurally incomplete markers where recording is optional and missingness cannot be quantified (for example symptoms). Definitive measurement of completeness would require prospective observation of documentation practice in primary care, which was outside the scope of this DPhil.

Outcomes

Outcomes in this study were defined from primary care records. The definitive sources for hospitalisation and death data are HES and the ONS. At the start of this DPhil access to linked HES and ONS data was available, which would have enabled gold standard outcome definitions, but access was subsequently revoked due to governance changes. Consequently, I used all cause hospitalisation and all cause death from primary care CMR, where reason for admission and cause of death are not routinely available. This introduces classification bias because some admissions and deaths will be unrelated to the index ARI. Despite this, in many setting using 'all-cause' outcomes is often preferred. Future linkage to HES and ONS via the secure hashed NHS number would be highly desirable to validate and refine outcome definitions in primary care.

6.4.3 Methodological choices

Systematic review

On reflection, there were drawbacks to using a systematic review to define candidate severity markers. The process was extremely time consuming, with over two thousand titles and abstracts independently screened by two reviewers. Although clear definitions were set out and articles were screened independently, there was inevitable subjectivity in deciding what constituted a severity marker. The primary care informatics review helped mitigate this by providing an additional layer of expert scrutiny.

An alternative approach could have used a different systematic review design. For example, I could have sought only predictors that had demonstrated an association with adverse outcomes. The downside is that this would have narrowed the range of markers considered and would probably have missed healthcare utilisation markers, particularly those specific to primary care. Moreover, identifying markers of little value is itself informative.

A Delphi study could also have been used. A panel of experts could iteratively reach consensus on candidate markers, offering a more efficient and targeted way to identify relevant indicators while reducing subjectivity. It would have required only a single step rather than a systematic review followed by an expert review. It is likely that many of the same markers would have been identified, albeit more efficiently. However, the set produced here was coherent, with plausible associations between many predictors and severe outcomes.

Finally, a data driven approach could have been used whereby all SNOMED codes were recorded at and around the time of the ARI diagnosis. This could then have been synthesised into groups, much like in the systematic review. This approach would still require considerable expert review at the end to ensure that these are correctly interpreted. However a data quality assessment could be integrated into the step, which would potentially be more efficient. Ultimately, each of these process would likely have resulted in a similar set of severity markers.

Evaluation framework

To the best of my knowledge no previous study has systematically explored the potential for defining severity indicators in the manner of this DPhil. The literature clearly identifies the value of such indicators, but focuses principally on measurements of severity using traditional methods: case hospitalisation and case fatality ratio, which inherently have a substantial lag [23, 24]. These guidelines also tend to focus on the measurements of clinical severity for laboratory confirmed cases such as influenza and SARS-CoV-2. This created a challenge in that no prior methodological precedent was set for how this should be approached. The final analysis work flow used in Chapter 5 is novel, however, each element of this work flow utilises well recognised statistical techniques such as ORs and CCs.

One deliberate methodological decision in Chapter 5 was to retain individual predictors rather than build multivariable models. The primary motivation was operational simplicity. While constructing multivariable models is relatively straightforward, maintaining and interpreting severity indicators based on models would require resources beyond those currently available to the RSC surveillance programme. In addition, multivariable models reduce interpretability and make face validation more difficult. They would also necessitate development of an aggregate severity score, introducing additional technical and computational demands for prospective, near real-time surveillance. This supports stability and simplicity, two key attributes in any public health surveillance system [161]. As the computational capacity of the surveillance system grows, multivariable models could be considered in the future which could also help manage confounding.

Changes in recording quality during the pandemic exposed vulnerabilities in CMR-based surveillance. However, pandemics are not the only potential source of disruption. The technical and regulatory landscape is evolving rapidly. The emergence of TREs as the principal route to access highly sensitive patient datasets has occurred over a short period, creating operational challenges for the RSC and for other programmes that use CMR data for secondary purposes. Similar shifts are likely in the future and could affect not only

severity indicators but entire surveillance workflows.

Technological changes will also arise. Some will be small, such as SNOMED version updates that require periodic review of the PhA and the code lists used for severity markers. Others could be more disruptive, for example the introduction of new primary care IT systems that displace EMIS and competitors. Recently the first new primary care system in 25 years was released [164]. This system is cloud based and accessed via a browser and is integrated with other NHS services. It has potential to disrupt the primary care IT system landscape. If current systems such as EMIS do not also keep pace they could become obsolete which would have major implications for the RSC surveillance system.

6.5 FUTURE WORK

In this section, I set out a plan of further work in this field. I focus on the key next steps in testing and operationalising severity indicators.

6.5.1 Piloting

Moving forward, the first and most critical step is to develop a pilot study using prospectively collected surveillance data. Based on the findings of this DPhil, I would focus on a compact set of predictors centred on key numeric clinical signs, top-ranked symptoms, and health care utilisation indicators from the pandemic, as well as the composite outcome. This approach ensures that the data used closely reflect the real-world setting, making the findings more robust and relevant. It also allows more accurate assessment of timeliness by measuring the exact date a record becomes available to the surveillance system, rather than relying on event dates or recorded dates.

One of the limitations of collecting data prospectively is the time required to accumulate a sufficiently large dataset to power a rigorous analysis. A valuable parallel piece of work would be to refine retrospective datasets to improve our ability to identify appropriate

records. For example, mechanisms could be developed to label consultation types as face to face, remote, or encoded retrospectively (for instance, from a discharge summary). These labels could help select surveillance records more accurately, both in terms of the cases identified and the associated severity markers. In addition, a curated retrospective dataset could be developed to more closely reflect real-world surveillance conditions.

6.5.2 Outcome validation

A parallel consideration would be clarification and validation of outcomes. Here I have used primary care outcomes, which was all that was available to me. Validation of these outcomes through comparison to the gold standard hospital and death data (HES and ONS) can provide more confidence that primary care outcomes can be used. This is a relatively straightforward study requiring a simple set analysis. Although access to HES and ONS datasets are more heavily regulated and the bar to accessing this data is higher this study should be prioritised in the near future.

6.5.3 Engagement of stakeholders

Once the outcome is clearly defined, the aims and objectives and action to be taken based on the pilot should be pre-specified. While this DPhil was intentionally exploratory given its novelty and the absence of prior evidence, the pilot will be designed to make a definitive decision about embedding severity indicators into routine surveillance reports. Accordingly, clear, pre-agreed objectives and decision criteria are required so that stakeholders are prepared to integrate the indicators if those criteria are met. Key partners in this process should include: UKHSA colleagues, NHS partners, RSC surveillance leads, RSC, software engineers, statisticians, member practices and the public. Table 6.1 indicates the role of each stakeholder involved.

Table 6.1: Key stakeholders and their roles in ARI surveillance

Stakeholder	**Description of role**
The public	*Enable surveillance by consenting to share their data (i.e., not opting out). Public involvement and PPI work are critical to ensure trust, guide priorities, and shape how results are communicated. Their perspectives help refine severity indicators and make outputs more meaningful.*
Member practices	*Provide the essential data that make surveillance possible. Their engagement supports data quality improvement, and highly involved practices can pilot new initiatives within the RSC network.*
RSC surveillance leads	*Oversee the project as a whole, coordinating with member practices and developing initiatives to improve data quality. They provide academic and operational leadership to ensure that indicators are scientifically robust and practically useful.*
UKHSA	*UKHSA are the principal funders of the surveillance at the RSC. Provide the national public health perspective and help determine which information is most valuable to support decision making. They place severity indicators within the broader context of respiratory surveillance and advise on policy relevance.*
NHS partners	*Supply linked health service data (e.g., Hospital Episode Statistics, ONS mortality data) that can validate outcomes. NHS partners also help ensure that indicators align with service planning and operational needs.*

Table 6.1: Key stakeholders and their roles in ARI surveillance (continued)

Stakeholder	Description of role
IT providers	*Supply the electronic health record systems that enable primary care data extraction. Early engagement with both established and new providers is essential to shape how their platforms can best support surveillance, data quality, and real-time indicator development.*
Software/database engineers	*Implement any required changes to the surveillance system and maintain the infrastructure underpinning severity indicators.*
Statisticians	*Define the most appropriate methodologies for baselines, thresholds, and models. They ensure that the chosen methods are statistically sound, reproducible, and suitable for integration into a pilot.*

6.5.4 Equity and representativeness

Work in this DPhil has focused on supplementing the current surveillance report with severity indicators. At present, incidence reports are aggregated by region and age, but not by other demographic parameters such as ethnicity. This level of aggregation provides a useful overview, but it does not highlight differences in how severe disease affects specific groups. A more granular understanding of who is most impacted by severe disease could help address surveillance disparities. This could include the consideration of variables such as ethnicity, deprivation, gender, and vaccination status.

Keeping surveillance simple is important, because excessive complexity can lead to instability. However, planned assessments of the individuals most commonly affected by severe disease could provide valuable insights into long-term trends in disparities. These insights could also ensure that resources reach those who most need them.

6.5.5 Analytical approach

Concurrently, the specific analytical approach should be defined. This approach should take account of the objectives of respiratory surveillance and be aimed at the provision of actionable public health intelligence. Whereas, this DPhil has assessed individual and aggregate-level associations between predictors, a future analysis could look more closely at the ability of indicators to predict severe outcomes and therefore health system pressures. A number of frameworks exist for such aggregate level (time series data) analysis with a range of complexity levels. Table 6.2 sets out some of the potential approaches.

Table 6.2: Baseline, threshold, and forecasting options for primary care ARI severity indicators

Approach	Description (what it is, pros/cons, examples)
1. Five-year seasonal baseline	Compute weekly expected values from the average of the past ≥ 5 comparable seasons (optionally excluding outliers) and compare current values to that baseline. **Pros:** Simple, transparent, quick to implement; intuitive seasonal context. **Cons:** Ignores underlying trend and reporting delays; sensitive to regime changes and outliers; limited uncertainty quantification. **Examples:** RSC ARI surveillance reports [165].
2. Intensity thresholds	Derive pre-epidemic and epidemic intensity thresholds from historical seasons (e.g., Moving Epidemic Method (MEM) or percentile bands [166]) to classify current activity. **Pros:** Standardised categories across seasons; widely used for influenza; easy to communicate. **Cons:** Needs multiple stable pre-pandemic seasons; assumes consistent seasonality and coding; classifies intensity rather than forecasting; not effective in pandemic settings. **Examples:** ECDC/European influenza intensity (MEM); UK national influenza intensity reporting.

Table 6.2: Baseline, threshold, and forecasting options (continued)

Approach	Description (what it is, pros/cons, examples)
3. Regression	Over-dispersed Poisson or negative binomial regression models with trend, seasonality, and calendar effects to estimate expected counts and quantify excess activity. Includes established aberration-detection approaches such as the Farrington and Farrington Flexible algorithms, which fit quasi-Poisson regression models to historical windows to generate expected counts and alert thresholds. RAMMIE regression is conceptually similar, extending this framework for UK operational syndromic surveillance [167]. **Pros:** Handles trend/seasonality and over-dispersion; produces excess estimates and supports alerting; established operational workflows exist (e.g., RAMMIE, `surveillance` R package). **Cons:** Requires upkeep and re-baselining after structural changes; assumes reasonably consistent recording practices; may be unstable with small counts; Farrington-type models depend on stable historical windows. **Examples:** UKHSA syndromic surveillance (RAMMIE regression); Farrington/Farrington Flexible algorithm; EuroMOMO network [168].
4. Short-horizon forecasts	Statistical time-series forecasts 1–4 weeks ahead with prediction intervals; alerts when observations exceed forecast bands. **Pros:** Operationally useful for planning; can include exogenous predictors; quantifies uncertainty. **Cons:** More complex; sensitive to structural breaks; needs ongoing monitoring and retraining. **Examples:** CDC FluSight forecasting [169].

6.5.6 Data quality improvements

Initiatives to improve data quality would be highly desirable. Here, stakeholder engagement is key, as data quality can be influenced by multiple groups: IT providers through user interface design; clinicians through member practice training and resource development; software and database engineers through optimisation of data processing; and statisticians through adjusting for missing data or biases.

Innovation can also support improved data quality, for example, integrating wearable technology with the patient record could help increase completeness of clinical sign recording. An initial approach would likely focus on supporting surveillance practices in coding fully characterised ARI cases, through education on the most valuable severity markers to record. Furthermore, the use of generative artificial intelligence (AI) in consultations can help support increased data quality through automated coding. Such systems are already in place in parts of England.

6.5.7 Summary of future work

Collectively, this approach lays out a clear path for both the essential and desirable pieces of work required to operationalise severity indicators within the surveillance system. This will involve the active engagement of all stakeholders to ensure that the work remains relevant and equitable. It will also require appropriate resources to sustain the workforce and technical infrastructure needed to maintain these indicators in routine practice.

6.6 CLOSING REMARKS

Respiratory viruses represent one of the greatest infectious threats to public health, as starkly demonstrated by the COVID-19 pandemic. Seasonal respiratory viruses such as influenza and RSV also impose a substantial burden, driving winter pressures across the health system, from community pharmacies through to the ICU. Epidemic and pandemic preparedness remains a central aim of respiratory virus surveillance, and it has been a

principal theme running through this DPhil.

Preparedness is built on a state of readiness to manage pathogenic threats, and surveillance is a cornerstone of this readiness. This DPhil has contributed to preparedness by developing a new PhA, openly sharing the algorithm for transparency and scrutiny, and by introducing severity indicators designed to move primary care-based surveillance beyond just incidence measures. These indicators aim to capture changes in clinical severity that more accurately reflect pressures on health services, thereby generating more actionable intelligence for decision makers.

The work has also highlighted the importance of data quality as both an enabler and a limitation of surveillance. Fluctuations in coding, susceptibility to disruption, and incomplete recording reduce the reliability of primary care CMR-based intelligence. These issues cannot be ignored if surveillance is to support epidemic and pandemic preparedness effectively. Improving data quality through stakeholder engagement, education, and innovation should therefore remain a priority for future system development.

Finally, this work is one part of a large, interconnected global effort to reduce the impact of ARIs. At the heart of this effort are the people whose health surveillance ultimately seeks to protect. By developing severity indicators within primary care CMR-based surveillance systems, the aim is to contribute a focused but meaningful piece of work that complements wider initiatives. While this thesis represents a substantial body of research in its own right, it is best understood as one contribution to the collective endeavour to strengthen surveillance, support epidemic and pandemic preparedness, and safeguard population health, continuing a longstanding public health priority, exemplified by the seventeenth-century Bills of Mortality, to provide timely and actionable intelligence on disease severity [170].

List of Abbreviations

AI artificial intelligence.

ARI acute respiratory infection.

AUROC area under the receiver operating characteristic curve.

BP blood pressure.

CABI Centre for Agriculture and Bioscience International.

CC correlation coefficient.

CDC United States Centers for Disease Control and Prevention.

CFR case fatality ratio.

CHR case hospitalisation ratio.

CI confidence interval.

CMR computerised medical record.

COPD chronic obstructive pulmonary disease.

COVID-19 coronavirus disease 2019.

CRP C-reactive protein.

DPhil Doctor of Philosophy.

ECDC European Centre for Disease Prevention and Control.

ECL Expression Constraint Language.

ECLD exacerbation of chronic lung disease.

ED emergency department.

EMIS Egton Medical Information Systems.

ERVISS European Respiratory Virus Surveillance Summary.

EU European Union.

FDA Food and Drug Administration.

GCS Glasgow Coma Scale.

GISRS Global Influenza Surveillance and Response System.

GM geometric mean.

GRIPP2-SF Guidance for Reporting Involvement of Patients and the Public, Short Form.

HES Hospital Episode Statistics.

ICD International Classification of Diseases.

ICU intensive care unit.

ILI influenza-like illness.

IQR interquartile range.

ISO International Organization for Standardization.

IV intravenous.

LLM large language model.

LOESS locally estimated scatterplot smoothing.

LRTI lower respiratory tract infection.

MAARI medically attended acute respiratory infection.

MERS-CoV Middle East respiratory syndrome coronavirus.

MeSH Medical Subject Headings.

NDOO National Data Opt-out.

NEWS National Early Warning Score.

NEWS2 National Early Warning Score 2.

NHS National Health Service.

NICE National Institute for Health and Care Excellence.

NLP natural language processing.

NOS not otherwise specified.

ONS Office for National Statistics.

OR odds ratio.

ORCHID Oxford Clinical Informatics Digital Hub.

PDS Personal Demographics Service.

PEWS Paediatric Early Warning System.

PhA digital phenotyping algorithm.

PISA Pandemic Influenza Severity Assessment.

PPI patient and public involvement.

PPV positive predictive value.

PRIMIS University of Nottingham, Primary Care Information Services.

PSAF Pandemic Severity Assessment Framework.

QoF Quality and Outcomes Framework.

RESST Real-time Syndromic Surveillance Team.

RSC Oxford–Royal College of General Practitioners Research and Surveillance Centre.

RSV respiratory syncytial virus.

RWD real-world data.

SARI severe acute respiratory infection.

SARI Watch Severe Acute Respiratory Infection Watch surveillance system.

SARS-CoV-1 severe acute respiratory syndrome coronavirus 1.

SARS-CoV-2 severe acute respiratory syndrome coronavirus 2.

SGSS Second Generation Surveillance System.

SIRS Systemic Inflammatory Response Syndrome.

SNOMED Systematized Nomenclature of Medicine.

SNR signal-to-noise ratio.

SOFA Sequential Organ Failure Assessment.

TESSy The European Surveillance System.

TPP The Phoenix Partnership (SystmOne).

TRE Trusted Research Environment.

UK United Kingdom.

UKHSA United Kingdom Health Security Agency.

URTI upper respiratory tract infection.

US United States.

WHO World Health Organization.

Glossary of Terms

acute respiratory infection Acute respiratory infection (ARI) is a surveillance concept representing a broad range of acute infective respiratory illnesses. In this DPhil, it encompasses influenza-like illness (ILI), upper respiratory tract infection (URTI), lower respiratory tract infection (LRTI), exacerbation of chronic lung disease (ECLD), suspected COVID-19, and ARI not otherwise specified (NOS).

age band The categorical grouping of individuals by age used for stratified analysis and reporting of incidence and severity indicators. In this DPhil, two age banding structures were applied. For the data quality assessment, five RSC age bands were used following the UKHSA surveillance reporting standard: under 1 year, 1–4 years, 5–14 years, 15–64 years, and 65 years and over. For the evaluation stage, these were merged into three broader groups to simplify stratification: <15 years (combining under 1, 1–4, and 5–14 years), 15–64 years, and 65 years.

aggregate-level analysis An analytical approach that examines relationships between variables summarised at a group or time-unit level rather than at the level of individual patients. In this DPhil, aggregate-level analysis was used to assess temporal correlations between weekly trends in candidate severity indicators and observed severe outcomes, providing a population-level perspective on the relationship.

case ascertainment bias A limitation in surveillance where only a subset of true cases are identified or recorded, resulting in a distorted view of disease severity or incidence. In traditional hospital-based surveillance, severe cases are more likely to be captured, while milder cases remain unseen. Primary care computerised medical records reduce this bias by capturing a broader spectrum of illness, improving representativeness and strengthening epidemic and pandemic preparedness.

case fatality ratio A population-level measure of disease severity defined as the proportion of identified cases that result in death within a specified time period.

case hospitalisation ratio A population-level measure of disease severity defined as the proportion of identified cases that result in hospital admission within a specified time period.

Centers for Disease Control and Prevention The national public health agency of the United States responsible for protecting health and safety through disease control, prevention, and health promotion activities.

code list A defined set of clinical codes used to identify diagnoses, symptoms, treatments, or outcomes in the computerised medical record. In this DPhil, code lists were used to define acute respiratory infection cases and candidate severity markers within primary care data. Code lists may be constructed either extensionally, by listing all included codes, or intensionally, using logical rules expressed in the Expression Constraint Language.

composite outcome The combined endpoint used in this DPhil was the recording of a hospital attendance, admission, complication of ARI (sepsis or acute respiratory failure) or death from any cause within 56 days of the index ARI case.

computerised medical record A digital version of a patient's health record, containing coded entries of diagnoses, symptoms, test results, prescriptions, and other clinical data recorded during consultations.

cross-correlation analysis A statistical method used to assess the temporal association between two time series. In this DPhil, cross-correlation analysis was used to evaluate the relationship between candidate severity indicators and severe outcomes at the aggregate level.

digital phenotyping algorithm A method for identifying clinical concepts such as acute respiratory infection episodes from a computerised medical record. The algorithm developed in this DPhil combines diagnostic, symptom, and treatment codes.

ecological fallacy An error arising when inferences about individuals are made from aggregate-level data. In this DPhil, both individual and aggregate-level analyses were conducted to mitigate this risk.

EMIS A primary care electronic health record system widely used in England. EMIS practices contribute data to the RSC network for respiratory surveillance.

European Centre for Disease Prevention and Control A European Union agency responsible for strengthening infectious disease surveillance and response across member states. The RSC contributes to ECDC influenza and respiratory surveillance through the UKHSA partnership.

exacerbation of chronic lung disease A sudden deterioration in respiratory symptoms such as cough, sputum production, or breathlessness in an individual with underlying chronic lung disease, including chronic obstructive pulmonary disease, asthma, or other chronic respiratory disorders.

expression constraint A formal rule used within SNOMED CT to define a subset of clinical concepts based on logical relationships in the terminology's polyhierarchy.

Expression Constraint Language A standard query language used to define subsets of SNOMED CT concepts based on logical expressions. It enables rule-based construction of intensional code lists, supporting transparent and reproducible code definitions.

extensional code list A code list created by explicitly enumerating all included codes. Extensional code lists are static and may require frequent manual updates as clinical terminologies evolve. Earlier versions of RSC surveillance algorithms used extensional lists, later replaced by intensional lists built with the Expression Constraint Language.

flag A categorical indicator used to classify clinical sign values according to severity thresholds. In this DPhil, flags were derived from NEWS2 scores for adults and

equivalent PEWS thresholds for children under 15 years. Flag 1 represents the lowest severity level and Flag 3 the highest.

fold change A relative measure of change in a variable, typically expressed as a ratio. In this DPhil, fold change was used to compare recording rates of severity markers across different study periods.

geometric mean A multiplicative average used to combine values that are expressed as ratios or on different scales. In this DPhil, the geometric mean was used to combine scaled odds ratios and cross-correlation coefficients into a single balanced metric, providing a summary measure that gives equal weight to relative differences across indicators.

Hospital Episode Statistics A national database containing details of all admissions, attendances, and outpatient appointments at NHS hospitals in England. Not used in this DPhil due to data governance constraints.

individual-level analysis An analytical approach that assesses associations between variables measured at the level of individual patient or episode. In this DPhil, individual-level analysis was used to quantify the strength of association between recorded predictors of severe outcome—such as symptoms, signs, or treatments—and the occurrence of the outcome.

influenza-like illness Defined by the Research and Surveillance Centre (RSC) as an acute respiratory infection with a measured (or clinically plausible) temperature of 38°C (except in older adults who may be afebrile), accompanied by cough, systemic upset such as headache or myalgia, and a sudden onset, in the absence of a more plausible diagnosis.

integrated respiratory surveillance A respiratory surveillance framework that monitors multiple respiratory pathogens. This supports surveillance of novel, emerging, and re-emerging pathogens such as SARS-CoV-2, as well as influenza.

intensional code list A code list defined by a logical rule or expression rather than by manual enumeration. Intensional code lists in this DPhil were constructed using the Expression Constraint Language within SNOMED, allowing dynamic, reproducible selection of all relevant descendant or related concepts.

lower respiratory tract infection An acute infection of the respiratory tract below the larynx, including the bronchi, bronchioles, and alveoli. LRTI encompass clinical entities such as bronchitis, bronchiolitis, and pneumonia.

NEWS2 score The National Early Warning Score 2, a standardised tool for assessing the severity of acute illness through measurement of clinical signs. Used in this DPhil to define cut-offs for numeric clinical signs and for composite clinical sign severity thresholds.

non-pandemic period Combined late pre-pandemic (Period 2) and post-pandemic (Period 4) phases used in the evaluation stage of this DPhil.

numeric severity marker A type of severity marker represented by a numeric measurement such as blood pressure, temperature, or oxygen saturation. These markers can be assessed for completeness and temporal plausibility and are often used to construct objective severity indicators.

Office for National Statistics deaths data Mortality records compiled by the Office for National Statistics, including date and cause of death. In this DPhil, these data are used as the reference standard for validating death outcomes recorded in primary care computerised medical records.

Oxford-Royal College of General Practitioners Research and Surveillance Centre A syndromic and sentinel surveillance network of general practices coordinated by the Royal College of General Practitioners, Oxford University, and the UK Health Security Agency. The RSC collects and analyses pseudonymised primary care data to monitor the incidence of infectious diseases, including acute respiratory infections.

It also undertakes virological sampling and supports vaccine effectiveness studies. Data from the RSC are used for all empirical studies in this DPhil.

Pandemic Influenza Severity Assessment A framework developed by the World Health Organization to assess the population-level severity of influenza epidemics and pandemics.

pandemic period The timeframe encompassing the COVID-19 pandemic (Period 3). Used in the evaluation stage of this DPhil for comparison with the non-pandemic period (Periods 2 and 4).

Pandemic Severity Assessment Framework A framework developed by the United States Centers for Disease Control and Prevention to evaluate the potential severity of emerging influenza pandemics.

population-level severity A measure of the seriousness of disease at the population level. It reflects the extent to which circulating respiratory pathogens are associated with adverse outcomes such as hospitalisation, intensive care admission, or death. In the Pandemic Severity Assessment Framework (PISA) framework, this corresponds to the *seriousness of disease* dimension and complements measures of incidence to estimate overall impact.

practice liaison team A group within the RSC responsible for engaging and supporting member practices, promoting data quality, and facilitating new surveillance initiatives.

predictor of severe outcome A type of severity marker representing a characteristic recorded around the time of acute respiratory infection onset, such as symptoms, signs, or treatments, that may predict subsequent severe outcomes. Predictors are more timely than severe outcomes and can support earlier assessment of severity.

rolling mean A smoothing technique that calculates the average value over a moving time window to reduce short-term fluctuations in surveillance data.

sentinel surveillance A surveillance system that relies on a network of selected reporting sites (such as general practices, hospitals, or laboratories) to provide high-quality data on specific diseases, often including virological sampling to confirm respiratory infections.

severe outcome A type of severity marker representing a definitive endpoint of disease, such as hospitalisation or death. Severe outcomes are reliable but are usually less timely because of the lag between onset and their availability in primary care computerised medical records.

severity indicator A type of surveillance indicator that specifically quantifies the proportion of acute respiratory infection (ARI) episodes resulting in severe outcomes within a defined period. It is expressed as a ratio, with the denominator representing the total number of ARI episodes and the numerator representing those episodes meeting predefined severity criteria (i.e., severity markers).

severity marker The criteria used to decide whether an acute respiratory infection episode is classified as severe. They include two main types: severe outcomes (e.g., hospitalisation or death) and predictors of severe outcomes (e.g., symptoms, signs, or treatments). Severity markers define the numerator of severity indicators.

signal-to-noise ratio A measure of the strength of a true signal relative to background variability. Used in this DPhil to identify stable and reliable severity indicators for surveillance.

SNOMED CT The Systematized Nomenclature of Medicine – Clinical Terms, an internationally standardised, polyhierarchically structured clinical terminology used in UK primary care and globally to record clinical concepts in computerised medical records.

stratification In this DPhil, stratification was applied differently in the pandemic and non-pandemic analyses. During the pandemic period (Period 3), two strata were analysed, corresponding to adults (15–64 years, and 65 years). During the non-

pandemic period (Periods 2 and 4), nine strata were analysed, defined by three age bands (<15 years, 15–64 years, and 65 years) crossed with three acute respiratory infection subtypes: upper respiratory tract infection (URTI), influenza-like illness (ILI), and lower respiratory tract infection combined with exacerbation of chronic lung disease (LRTI–ECLD).

structural completeness Structurally complete severity markers are those for which systems are in place to enforce recording of a given field. For example, prescription recording is almost universally structurally complete.

structurally incomplete Structurally incomplete severity markers are those where recording is at the clinician's discretion, meaning completeness cannot be reliably assessed directly from the computerised medical record.

study period A defined timeframe used to structure temporal comparisons in data analysis. In this DPhil, four study periods (P1–P4) were specified to examine the temporal dynamics and data quality of severity marker recording within primary care computerised medical records. The study period spanned ISO week 40 of 2008 (29th September 2008) to ISO week 39 of 2024 (29th September 2024) and was divided into four phases centred on key milestones in the COVID-19 pandemic: P1 – early pre-pandemic; P2 – late pre-pandemic; P3 – pandemic; and P4 – post-pandemic. All four periods were included in data quality assessment; however, P1 was excluded from the evaluation stage due to its earlier timeframe and differences in recording completeness.

surveillance The ongoing, systematic collection, analysis, and interpretation of health data for public health action.

surveillance indicator A quantitative measure derived from surveillance data that describes monitoring of a specific surveillance concept over time. Typically, for acute respiratory infection (ARI) surveillance these are incidence, case positivity or population-level severity. Changes in the indicators over time, for example, an increase in incidence, can indicate impending public health pressures.

syndromic surveillance A surveillance approach based on the real-time or near real-time collection, analysis, and interpretation of health-related data that precede diagnosis, focusing on clinical syndromes such as influenza-like illness or acute respiratory infection, rather than laboratory-confirmed diagnoses.

TPP A UK supplier of primary care electronic health record systems. TPP practices are increasingly being incorporated into the RSC surveillance network.

UK Health Security Agency (UKHSA) The United Kingdom's national public health agency responsible for protecting health from infectious diseases and other hazards. The UKHSA funds and collaborates with the RSC to deliver national primary care surveillance for respiratory infections and supports integration of severity indicators into routine monitoring.

valid completeness measure A measure of data completeness where recorded values accurately reflect whether information was collected. For example, recording rates of numeric clinical signs such as blood pressure or oxygen saturation can serve as valid completeness measures because these are expected to be present for all assessed patients.

World Health Organization An agency of the United Nations responsible for global public health. The WHO coordinates international efforts to prevent, detect, and respond to infectious and other global health threats, including epidemics and pandemics.

Bibliography

[1] Paul Slack. *The impact of plague in Tudor and Stuart England : Slack, Paul : Free Download, Borrow, and Streaming : Internet Archive*. Clarendon Press, Oxford, 1990. URL https://archive.org/details/impactofplaguein0000slac_b5v7/page/n7/mode/2up.

[2] J. C. Robertson. Reckoning with London: interpreting the Bills of Mortality before John Graunt. *Urban History*, 23(3):325–350, 1996. ISSN 1469-8706. doi: 10.1017/S0963926800016898. URL https://www.cambridge.org/core/journals/urban-history/article/reckoning-with-london-interpreting-the-bills-of-mortality-before-john-graunt/D6A416132817560572374BE2EB740642.

[3] John Bell. *London's remembrancer: or, a true accompt of every particular weeks christnings and mortality in all the years of pestilence ... being XVIII years / Taken out of the Register of the Company of Parish Clerks of London. | Wellcome Collection*. Wellcome collection, London, 1665. URL https://wellcomecollection.org/works/rhgppmqj/items?canvas=15.

[4] S Declich1 and A O Carter2. Reviews /Analyses Public health surveillance: historical origins, methods and evaluation. *Bulletin of the World Health Organization*, 72(2):285–304, 1994.

[5] World Health Organization. International Health Regulations Third Edition (2005). In *International Health Regulations*, volume Third Edition, pages 10–10. World Health Organization, 2005. ISBN 9789241580496. URL https://apps.who.int/iris/rest/bitstreams/1031116/retrieve.

[6] Richard J. Hatchett, Carter E. Mecher, and Marc Lipsitch. Public health interven-

tions and epidemic intensity during the 1918 influenza pandemic. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18):7582–7587, 5 2007. ISSN 00278424. doi: 10.1073/PNAS.0610941104/SUPPL{_}FILE/10941TABLE11.XLS. URL https://www.pnas.org/doi/abs/10.1073/pnas.0610941104.

[7] Alan J. Hay and John W. McCauley. The WHO global influenza surveillance and response system (GISRS)-A future perspective. *Influenza and other respiratory viruses*, 12(5):551–557, 9 2018. ISSN 1750-2659. doi: 10.1111/IRV.12565. URL https://pubmed.ncbi.nlm.nih.gov/29722140/.

[8] Jobie Budd, Benjamin S. Miller, Erin M. Manning, Vasileios Lampos, Mengdie Zhuang, Michael Edelstein, Geraint Rees, Vincent C. Emery, Molly M. Stevens, Neil Keegan, Michael J. Short, Deenan Pillay, Ed Manley, Ingemar J. Cox, David Heymann, Anne M. Johnson, and Rachel A. McKendry. Digital technologies in the public-health response to COVID-19. *Nature Medicine 2020 26:8*, 26(8):1183–1192, 8 2020. ISSN 1546-170X. doi: 10.1038/s41591-020-1011-4. URL https://www.nature.com/articles/s41591-020-1011-4.

[9] Tom Chen, Wenjun Li, Bob Zambarano, and Michael Klompas. Small-area estimation for public health surveillance using electronic health record data: reducing the impact of underrepresentation. *BMC Public Health*, 22(1):1–10, 12 2022. ISSN 14712458. doi: 10.1186/S12889-022-13809-2/TABLES/3. URL https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-022-13809-2.

[10] Nidhi Ghildayal, Kshema Nagavedu, Jennifer L. Wiltz, Soowoo Back, Tegan K. Boehmer, Christine Draper, Adi V. Gundlapalli, Casie Horgan, Keith A. Marsolo, Nik R. Mazumder, Juliane Reynolds, Matthew Ritchey, Sharon Saydah, Yacob G. Tedla, Thomas W. Carton, and Jason P. Block. Public Health Surveillance in Electronic Health Records: Lessons From PCORnet. *Preventing Chronic Disease*, 21, 2024. ISSN 1545-1151. doi: 10.5888/PCD21.230417.

[11] World Health Organization: Global Influenza Programme (GIP). Implementing the integrated sentinel surveillance of influenza and other respiratory viruses of epidemic and pandemic potential by the Global Influenza Surveillance and Response System. Technical report, World Health Organization, 12 2024. URL https://www.who.int/publications/i/item/9789240101432.

[12] R. Pebody, F. Warburton, N. Andrews, J. Ellis, B. Von Wissmann, C. Robertson, I. Yonova, S. Cottrell, N. Gallagher, H. Green, C. Thompson, M. Galiano, D. Marques, R. Gunson, A. Reynolds, C. Moore, D. Mullett, S. Pathirannehelage, M. Donati, J. Johnston, S. De Lusignan, J. McMenamin, and M. Zambon. Effectiveness of seasonal influenza vaccine in preventing laboratory-confirmed influenza in primary care in the United Kingdom: 2014/15 end of season results. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 20(36), 9 2015. ISSN 1560-7917. doi: 10.2807/1560-7917.ES.2015.20.36.30013. URL https://pubmed.ncbi.nlm.nih.gov/26535911/.

[13] Nicola Newson. Respiratory syncytial virus and its impact on the NHS. Technical report, Jouse of Lords, 2022. URL https://lordslibrary.parliament.uk/respiratory-syncytial-virus-and-its-impact-on-the-nhs/.

[14] UKHSA. Guidance on use of antiviral agents for the treatment and prophylaxis of seasonal influenza. Technical report, UKHSA, 2021.

[15] Richard Fry, Joe Hollinghurst, Helen R. Stagg, Daniel A. Thompson, Claudio Fronterre, Chris Orton, Ronan A. Lyons, David V. Ford, Aziz Sheikh, and Peter J. Diggle. Real-time spatial health surveillance: Mapping the UK COVID-19 epidemic. *International Journal of Medical Informatics*, 149:104400, 5 2021. ISSN 18728243. doi: 10.1016/J.IJMEDINF.2021.104400. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7843148/.

[16] Her Majesty's Government. COVID-19 Response: Living with COVID-19 - GOV.UK. Technical report, Cabinet Office, 2022. URL https://www.gov.

uk/government/publications/covid-19-response-living-with-covid-19/covid-19-response-living-with-covid-19.

[17] Simon de Lusignan, Praveen SebastianPillai, Omid Parvizi, Cecilia Okusi, Mark Joy, Shuma Banik, Fatima Batool, Katja Hoschler, Beatrix Kele, Angie Lackenby, Joanna Ellis, Richard Pebody, Conall Watson, Jamie Lopez Bernal, and Maria Zambon. Enabling genomic surveillance from 30 years of linked English sentinel network data: The Wellcome Quinquagenarian (QQG) Biomedical Resource. *Wellcome Open Research*, 10:411, 8 2025. doi: 10.12688/WELLCOMEOPENRES.23653.1.

[18] Simon de Lusignan, FD Richard Hobbs, and Aziz Sheikh. Lessons from the English primary care sentinel network's response to the COVID-19 pandemic. *The Lancet Infectious Diseases*, 24(1):14–16, 1 2024. ISSN 14744457. doi: 10.1016/S1473-3099(23)00736-3. URL https://www.thelancet.com/action/showFullText?pii=S1473309923007363https://www.thelancet.com/action/showAbstract?pii=S1473309923007363https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(23)00736-3/abstract.

[19] Thedi Ziegler, Ann Moen, Wenqing Zhang, and Nancy J. Cox. Global Influenza Surveillance and Response System: 70 years of responding to the expected and preparing for the unexpected. *The Lancet*, 400(10357):981–982, 9 2022. ISSN 1474547X. doi: 10.1016/S0140-6736(22)01741-X. URL http://www.thelancet.com/article/S014067362201741X/fulltexthttp://www.thelancet.com/article/S014067362201741X/abstracthttps://www.thelancet.com/journals/lancet/article/PIIS0140-6736(22)01741-X/abstract.

[20] William Elson, Maria Zambon, and Simon de Lusignan. Integrated respiratory surveillance after the COVID-19 pandemic. *The Lancet*, 400(10367):1924–1925, 12 2022. ISSN 1474547X. doi: 10.1016/S0140-6736(22)02325-X. URL http://www.thelancet.com/article/S014067362202325X/fulltexthttp://www.thelancet.com/article/S014067362202325X/abstracthttps:

//www.thelancet.com/journals/lancet/article/PIIS0140-6736(22)02325-X/abstract.

[21] Royal College of General Practitioners- Research and Surveillance Centre. Weekly surveillance report archives. URL https://www.rcgp.org.uk/representing-you/research-at-rcgp/research-surveillance-centre/public-health-data.

[22] Richard S. Hopkins, Michael Landen, and Megan Toe. Development of Indicators for Public Health Surveillance of Substance Use and Mental Health. *Public Health Reports*, 133(5):523, 9 2018. ISSN 14682877. doi: 10.1177/0033354918784913. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6134563/.

[23] Global Influenza Programme (GIP). Pandemic influenza severity assessment (PISA). Technical report, World Health Organization, 2024. URL https://www.who.int/publications/i/item/9789240093881.

[24] Carrie Reed, Matthew Biggerstaff, Lyn Finelli, Lisa M. Koonin, Denise Beauvais, Amra Uzicanin, Andrew Plummer, Joe Bresee, Stephen C. Redd, and Daniel B. Jernigan. Novel Framework for Assessing Epidemiologic Effects of Influenza Epidemics and Pandemics - Volume 19, Number 1—January 2013 - Emerging Infectious Diseases journal - CDC. *Emerging Infectious Diseases*, 19(1):85–91, 1 2013. ISSN 10806040. doi: 10.3201/EID1901.120124. URL https://wwwnc.cdc.gov/eid/article/19/1/12-0124_article.

[25] Homa Attar Cohen, Samuel Mesfin, Juniorcaius Ikejezie, Zyleen Kassamali, Finlay Campbell, Sandra Adele, Noe Guinko, Friday Idoko, Bernadette Basuta Mirembe, Maria Elizabeth Mitri, Ingrid Nezu, Kazuki Shimizu, Ajong Brian Ngongheh, Nikola Sklenovska, Nicksy Gumede, Fausta Shakiwa Mosha, Basant Mohamed, Aura Corpuz, Richard Pebody, Marco Marklewitz, Lionel Gresh, Jairo A.Mendez Rico, Kareena Hundal, Masaya Kato, Amarnath Babu, Brett N. Archer, Olivier Le Polain de Waroux, Maria D. Van Kerkhove, Abdirahman Mahamud, Lorenzo

Subissi, and Boris I. Pavlin. Surveillance for variants of SARS-CoV-2 to inform risk assessments. *Bulletin of the World Health Organization*, 101(11):707–716, 11 2023. ISSN 1564-0604. doi: 10.2471/BLT.23.290093. URL https://pubmed.ncbi.nlm.nih.gov/37961054/.

[26] Lone Simonsen, Elizabeth Higgs, and Robert J. Taylor. Clinical research networks are key to accurate and timely assessment of pandemic clinical severity. *The Lancet Global Health*, 6(9):e956–e957, 9 2018. ISSN 2214109X. doi: 10.1016/S2214-109X(18)30304-8. URL http://www.thelancet.com/article/S2214109X18303048/fulltexthttp://www.thelancet.com/article/S2214109X18303048/abstracthttps://www.thelancet.com/journals/langlo/article/PIIS2214-109X(18)30304-8/abstract.

[27] Jamie Lopez Bernal, Mary A. Sinnathamby, Suzanne Elgohari, Hongxin Zhao, Chinelo Obi, Laura Coughlan, Vasileios Lampos, Ruth Simmons, Elise Tessier, Helen Campbell, Suzanna McDonald, Joanna Ellis, Helen Hughes, Gillian Smith, Mark Joy, Manasa Tripathy, Rachel Byford, Filipa Ferreira, Simon de Lusignan, Maria Zambon, Gavin Dabrera, Kevin Brown, Vanessa Saliba, Nick Andrews, Gayatri Amirthalingam, Sema Mandal, Michael Edelstein, Alex J. Elliot, and Mary Ramsay. The impact of social and physical distancing measures on COVID-19 activity in England: Findings from a multi-tiered surveillance system. *Eurosurveillance*, 26(11):2001062, 3 2021. ISSN 15607917. doi: 10.2807/1560-7917.ES.2021.26.11.2001062/CITE/PLAINTEXT. URL https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2021.26.11.2001062.

[28] Deaths registered weekly in England and Wales, provisional - Office for National Statistics. URL https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending4july2025.

[29] Lone Simonsen, Elizabeth Higgs, Robert J. Taylor, Deborah Wentworth, Al Cozzi-

Lepri, Sarah Pett, Dominic E. Dwyer, Richard Davey, Ruth Lynfield, Marcelo Losso, Kathleen Morales, Marshall J. Glesby, Jozef Weckx, Dianne Carey, Cliff Lane, and Jens Lundgren. Using Clinical Research Networks to Assess Severity of an Emerging Influenza Pandemic. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 67(3):341, 7 2018. ISSN 15376591. doi: 10.1093/CID/CIY088. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6248856/.

[30] Fatimah S. Dawood, A. Danielle Iuliano, Carrie Reed, Martin I. Meltzer, David K. Shay, Po Yung Cheng, Don Bandaranayake, Robert F. Breiman, W. Abdullah Brooks, Philippe Buchy, Daniel R. Feikin, Karen B. Fowler, Aubree Gordon, Nguyen Tran Hien, Peter Horby, Q. Sue Huang, Mark A. Katz, Anand Krishnan, Renu Lal, Joel M. Montgomery, Kåre Mølbak, Richard Pebody, Anne M. Presanis, Hugo Razuri, Anneke Steens, Yeny O. Tinoco, Jacco Wallinga, Hongjie Yu, Sirenda Vong, Joseph Bresee, and Marc Alain Widdowson. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet. Infectious diseases*, 12(9):687–695, 9 2012. ISSN 1474-4457. doi: 10.1016/S1473-3099(12)70121-4. URL https://pubmed.ncbi.nlm.nih.gov/22738893/.

[31] Estimating mortality from COVID-19: Scientific brief, 4 August 2020. URL https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci-Brief-Mortality-2020.1.

[32] Shaun R. Seaman, Tommy Nyberg, Christopher E. Overton, David J. Pascall, Anne M. Presanis, and Daniela De Angelis. Adjusting for time of infection or positive test when estimating the risk of a post-infection outcome in an epidemic. *Statistical methods in medical research*, 31(10):1942–1958, 10 2022. ISSN 1477-0334. doi: 10.1177/09622802221107105. URL https://pubmed.ncbi.nlm.nih.gov/35695245/.

[33] Anne M. Presanis, Daniela De Angelis, Angela Hagy, Carrie Reed, Steven Ri-

ley, Ben S. Cooper, Lyn Finelli, Paul Biedrzycki, Marc Lipsitch, Joel Ackelsberg, Alys Adamski, Gail Adman, Elisabeth Agbor-Tabi, Christopher Aston, Josephine Atamian, Peter Backman, Sharon Balter, Oxiris Barbot, Sara T. Beatrice, Gary Beaudry, Elizabeth Begier, Geraldine Bell, Debra Berg, Magdalena Berger, James Betz, Susan Blank, Katherine Bornschlegel, Brooke Bregman, Meghan Burke, Barbara Butts, Liqun Cai, Alejandro Cajigal, Marilyn Campbell, Lorraine Camurati, Shadi Chamany, Dan Cimini, James Cone, Heather Cook, Debra Cook, Catherine Corey, Roseann Costarella, Christiana Coyle, Bindy Crouch, Cherry Ann Da Costa, Alexandria Daniels, Berta Darkins, Arlene DeGrasse, Susanne DeGrechie, Otto Del Cid, Bisram Deocharan, Luis Diaz, Kathleen DiCaprio, Laura DiGrande, Damon Duquaine, James Durrah, Joanna Eavey, Zadkijah Edghill, Barbara Edwin, Joseph Egger, Donna Eisenhower, Martin Evans, Shannon Farley, Richard Feliciano, Marcial Fernandez, Christine Fils-Aime, Anne Fine, Ana Maria Fireteanu, Kelly Fitzgerald, Anne Marie France, Thomas Frieden, Stephen Friedman, Jie Fu, Lawrence Fung, Latchmidat Girdharrie, Michelle Glaser, Christopher Goranson, Francine Griffing, Leena Gupta, Carol Hamilton, Heather Hanson, Scott Harper, Ian Hartman-O'Connell, Qazi Hasnain, Sonia Hedge, Michael Heller, Debra Hendrickson, Arnold Herskovitz, Kinjia Hinterland, Roosevelt Holmes, Jeanne Hom, Jeffrey Hon, Tana Hopke, Jennifer Hsieh, Scott Hughes, Stephen Immerwahr, Anne Marie Incalicchio, John Jasek, Julia Jimenez, Michael Johns, Lucretia Jones, Hannah Jordan, Chrispin Kambili, Jisuk Kang, Deborah Kapell, Adam Karpati, Bonnie Kerker, Kevin Konty, John Kornblum, Gary Krigsman, Fabienne Laraque, Marcelle Layton, Ellen Lee, Lillian Lee, Stephen Lee, Sungwoo Lim, Melissa Marx, Emily McGibbon, Kevin Mahoney, Gilbert Marin, Thomas Matte, Rene McAnanama, Ryan McKay, Carolyn McKay, Katherine McVeigh, Eric Medina, Wanda Medina, Danielle Michelangelo, Juliet Milhofer, Irina Milyavskaya, Mark Misener, Joseph Mizrahi, Linda Moskin, Matt Motherwell, Christa Myers, Hemanth P. Nair, Trang Nguyen, Diana Nilsen, Janet Nival, Jennifer Norton, William Oleszko, Carolyn Olson, Marc Paladini, Lucille Palumbo, Peter Papadopoulos,

Hilary Parton, Jacob Paternostro, Lynn Paynter, Krystal Perkins, Sharon Perlman, Haresh Persaud, Charles Peters, Melissa Pfeiffer, Roger Platt, Lindsay Pool, Amado Punsalang, Zahedur Rasul, Valerie Rawlins, Vasudha Reddy, Anne Rinchiuso, Teresa Rodriguez, Ramon Rosal, Maureen Ryan, Michael Sanderson, Allison Scaccia, Amber Levanon Seligson, Jantee Seupersad, Joanne Severe-Dildy, Asma Siddiqi, Ulirike Siemetzki, Tejinder Singh, Sally Slavinski, Meredith Slopen, Timothy Snuggs, David Starr, Catherine Stayton, Alaina Stoute, Jacqueline Terlonge, Alexandra Ternier, Lorna Thorpe, Catherine Travers, Benjamin Tsoi, Kimberly Turner, Joan Tzou, Shameeka Vines, Elizabeth Needham Waddell, Donald Walker, Connie Warner, Isaac Weisfuse, Don Weiss, Antoinette Williams-Akita, Elisha Wilson, Eliza Wilson, Marie Wong, Charles Wu, David Yang, Mohammad Younis, Sulaimon Yusuff, Christopher Zimmerman, and Jane Zucker. The Severity of Pandemic H1N1 Influenza in the United States, from April to July 2009: A Bayesian Analysis. *PLOS Medicine*, 6(12):e1000207, 12 2009. ISSN 1549-1676. doi: 10.1371/JOURNAL.PMED.1000207. URL https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000207.

[34] Marc Lipsitch, Steven Riley, Simon Cauchemez, Azra C. Ghani, and Neil M. Ferguson. Managing and Reducing Uncertainty in an Emerging Influenza Pandemic. *The New England journal of medicine*, 361(2):112, 7 2009. ISSN 0028-4793. doi: 10.1056/NEJMP0904380. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3066026/.

[35] Gavin Jamie, William Elson, Debasish Kar, Rashmi Wimalaratna, Uy Hoang, Bernardo Meza-Torres, Anna Forbes, William Hinton, Sneha Anand, Filipa Ferreira, Rachel Byford, Jose Ordonez-Mena, Utkarsh Agrawal, and Simon de Lusignan. Phenotype execution and modeling architecture to support disease surveillance and real-world evidence studies: English sentinel network evaluation. *JAMIA open*, 7(2), 4 2024. ISSN 2574-2531. doi: 10.1093/JAMIAOPEN/OOAE034. URL https://pubmed.ncbi.nlm.nih.gov/38737141/.

[36] Hadeel Alzoubi, Raid Alzubi, Naeem Ramzan, Daune West, Tawfik Al-Hadhrami, and Mamoun Alazab. A Review of Automatic Phenotyping Approaches using Electronic Health Records. *Electronics 2019, Vol. 8, Page 1235*, 8(11):1235, 10 2019. ISSN 2079-9292. doi: 10.3390/ELECTRONICS8111235. URL https://www.mdpi.com/2079-9292/8/11/1235/htmhttps://www.mdpi.com/2079-9292/8/11/1235.

[37] Katherine P. Liao, Tianxi Cai, Guergana K. Savova, Shawn N. Murphy, Elizabeth W. Karlson, Ashwin N. Ananthakrishnan, Vivian S. Gainer, Stanley Y. Shaw, Zongqi Xia, Peter Szolovits, Susanne Churchill, and Isaac Kohane. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *The BMJ*, 350:h1885, 4 2015. ISSN 17561833. doi: 10.1136/BMJ.H1885. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4707569/.

[38] NHS England Digital. Quality and Outcomes Framework, 2022-23: Main findings. Technical report, NHS England Digital, 2023. URL https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-prevalence-and-exceptions-data/2022-23/main-findings.

[39] SNOMED International. SNOMED CT, 2025. URL https://www.snomed.org/what-is-snomed-ct.

[40] Meredith Leston, William H Elson, Conall Watson, Anissa Lakhani, Carole Aspden, Clare R Bankhead, Ray Borrow, Elizabeth Button, Rachel Byford, Alex J Elliot, Xuejuan Fan, Uy Hoang, Ezra Linley, Jack Macartney, Brian D Nicholson, Cecilia Okusi, Mary Ramsay, Gillian Smith, Sue Smith, Mark Thomas, Dan Todkill, Ruby RS Tsang, William Victor, Alice Williams, John Williams, Maria Zambon, Gary Howsam, Gayatri Amirthalingam, Jamie Lopez-Bernal, FD Richard Hobbs, and Simon de Lusignan. Representativeness, Vaccination Uptake, and COVID-19 Clinical Outcomes 2020-2021 in the UK Oxford-Royal College of General

Practitioners Research and Surveillance Network: Cohort Profile Summary. *JMIR public health and surveillance*, 8(12), 12 2022. ISSN 2369-2960. doi: 10.2196/39141. URL https://pubmed.ncbi.nlm.nih.gov/36534462/.

[41] Ling Chu, Vaishnavi Kannan, Mujeeb A. Basit, Diane J. Schaeflein, Adolfo R. Ortuzar, Jimmie F. Glorioso, Joel R. Buchanan, and Duwayne L. Willett. SNOMED CT Concept Hierarchies for Computable Clinical Phenotypes From Electronic Health Record Data: Comparison of Intensional Versus Extensional Value Sets. *JMIR Medical Informatics*, 7(1), 1 2019. ISSN 22919694. doi: 10.2196/11487. URL /pmc/articles/PMC6351992//pmc/articles/PMC6351992/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6351992/.

[42] World Health Organization. International Classification of Diseases (ICD), 2024. URL https://www.who.int/standards/classifications/classification-of-diseases.

[43] SNOMED International. Expression Constraint Language - Specification and Guide - Expression Constraint Language - SNOMED Confluence, 2024. URL https://confluence.ihtsdotools.org/display/DOCECL.

[44] William H. Elson, Gavin Jamie, Rashmi Wimalaratna, Anna Forbes, Meredith Leston, Cecilia Okusi, Rachel Byford, Utkarsh Agrawal, Dan Todkill, Alex J. Elliot, Conall Watson, Maria Zambon, Roger Morbey, Jamie L. Lopez Bernal, F. D.Richard Hobbs, and Simon de Lusignan. Validation of an acute respiratory infection phenotyping algorithm to support robust computerised medical record-based respiratory sentinel surveillance, England, 2023. *Eurosurveillance*, 29(35), 8 2024. ISSN 15607917. doi: 10.2807/1560-7917.ES.2024.29.35.2300682.

[45] European Union. Commission Implementing Decision (EU) 2018/ 945 - of 22 June 2018 - on the communicable diseases and related special health issues to be covered by epidemiological surveillance as well as relevant case definitions. *Official Journal of the European Union*, 61:24, 2018. URL https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2018:170:FULL.

[46] Martin Chapman, Shahzad Mumtaz, Luke V. Rasmussen, Andreas Karwath, Georgios V. Gkoutos, Chuang Gao, Dan Thayer, Jennifer A. Pacheco, Helen Parkinson, Rachel L. Richesson, Emily Jefferson, Spiros Denaxas, and Vasa Curcin. Desiderata for the development of next-generation electronic health record phenotype libraries. *GigaScience*, 10(9):1–13, 9 2021. ISSN 2047217X. doi: 10.1093/GIGASCIENCE/GIAB059. URL https://dx.doi.org/10.1093/gigascience/giab059.

[47] Emily R. Pfaff, Andrew T. Girvin, Miles Crosskey, Srushti Gangireddy, Hiral Master, Wei Qi Wei, V. Eric Kerchberger, Mark Weiner, Paul A. Harris, Melissa Basford, Chris Lunt, Christopher G. Chute, Richard A. Moffitt, and Melissa Haendel. De-black-boxing health AI: demonstrating reproducible machine learning computable phenotypes using the N3C-RECOVER Long COVID model in the All of Us data repository. *Journal of the American Medical Informatics Association : JAMIA*, 30(7):1305, 7 2023. ISSN 1527974X. doi: 10.1093/JAMIA/OCAD077. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10280348/.

[48] Richard Williams, Evangelos Kontopantelis, Iain Buchan, and Niels Peek. Clinical code set engineering for reusing EHR data for research: A review. *Journal of Biomedical Informatics*, 70:1–13, 6 2017. ISSN 1532-0464. doi: 10.1016/J.JBI.2017.04.010.

[49] UK Health Security Agency. Immunisation against infectious disease - GOV.UK, 2025. URL https://www.gov.uk/government/collections/immunisation-against-infectious-disease-the-green-book.

[50] UK Health Security Agency. Immunisation of individuals with underlying medical conditions: the green book, chapter 7 - GOV.UK, 2020. URL https://www.gov.uk/government/publications/immunisation-of-individuals-with-underlying-medical-conditions-the-green-book-chapter-7.

[51] University of Nottingham. University of Nottingham, Primary Care Information Services (PRIMIS), 2025. URL https://www.nottingham.ac.uk/primis/.

[52] Food and Agriculture Organization of the United Nations; World Health Organization; World Organisation for Animal Health. Global technical meeting on MERS-CoV and other emerging zoonotic coronaviruses - meeting summary. Technical report, Food and Agriculture Organization of the United Nations; World Health Organization; World Organisation for Animal Health, 2022. URL https://www.who.int/publications/m/item/global-technical-meeting-on-mers-cov-and-other-emerging-zoonotic-coronaviruses-meeting-summary.

[53] Regional Office for Africa World Health Organization. Severe Acute Respiratory Syndrome ({SARS}) Preparedness and Response in the {WHO} African Region. Technical report, World Health Organization, Regional Office for Africa, 2003. URL https://www.afro.who.int/sites/default/files/sessions/working_documents/AFR-RC53-INF-DOC.2.pdf.

[54] Bennett J. Waxse, Fausto Andres Bustos Carrillo, Tam C. Tran, Huan Mo, Emily E. Ricotta, and Joshua C. Denny. Computable phenotypes to identify respiratory viral infections in the All of Us research program. *Scientific Reports 2025 15:1*, 15 (1):1–14, 5 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-02183-9. URL https://www.nature.com/articles/s41598-025-02183-9.

[55] Jane Y. Tong, Amanda Wong, Daniel Zhu, Judd H. Fastenberg, and Tristan Tham. The Prevalence of Olfactory and Gustatory Dysfunction in COVID-19 Patients: A Systematic Review and Meta-analysis. *Otolaryngology–head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery*, 163(1):3–11, 7 2020. ISSN 1097-6817. doi: 10.1177/0194599820926473. URL https://pubmed.ncbi.nlm.nih.gov/32369429/.

[56] Elisabeth Mahase. Bird flu: US reports first severe human case as California declares state of emergency. *BMJ*, 387:q2859, 12 2024. ISSN 1756-1833. doi: 10.1136/BMJ.Q2859. URL https://www.bmj.com/content/387/bmj.q2859https://www.bmj.com/content/387/bmj.q2859.abstract.

[57] National flu and COVID-19 surveillance report: 6 November 2025 (week 45) -

GOV.UK. Technical report. URL https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports-2025-to-2026-season/national-flu-and-covid-19-surveillance-report-6-november-2025-week-45?utm_source=chatgpt.com.

[58] Joint Committee on Vaccination and Immunisation. RSV immunisation programme: JCVI advice, 7 June 2023 (updated 11 September 2023) - GOV.UK. Technical report, Department of Health and Social Care, 6 2023. URL https://www.gov.uk/government/publications/rsv-immunisation-programme-jcvi-advice-7-june-2023.

[59] Noelle M. Cocoros, Karen Eberhardt, Vu Thuy Nguyen, Catherine M. Brown, Alfred DeMaria, Lawrence C. Madoff, Liisa M. Randall, and Michael Klompas. Electronic Health Record–Based Algorithm for Monitoring Respiratory Virus–Like Illness - Volume 30, Number 6—June 2024 - Emerging Infectious Diseases journal - CDC. *Emerging Infectious Diseases*, 30(6), 6 2024. ISSN 10806059. doi: 10.3201/EID3006.230473. URL https://wwwnc.cdc.gov/eid/article/30/6/23-0473_article.

[60] Patrick Saunders-Hastings, Sze Wing Heong, Pradeep Rajan, Timothy Burrell, Jeff Beers, Judith Cope, Deborah Thompson, and Cindy Ke Zhou. CBER Surveillance Program Biologics Effectiveness and Safety Initiative A Structured Review of Electronic Coding Algorithms for Pneumonia Using Administrative Claims and Electronic Health Records Final Report. Technical report, U.S. Food and drug Administration, 2021.

[61] Musaab Elkheder, Arturo Gonzalez-Izquierdo, Muhammad Qummer Ul Arfeen, Valerie Kuan, R Thomas Lumbers, Spiros Denaxas, and Anoop D Shah. Translating and evaluating historic phenotyping algorithms using SNOMED CT. *Journal of the American Medical Informatics Association : JAMIA*, 30(2):222, 1 2022. ISSN 1067-5027. doi: 10.1093/JAMIA/OCAC158. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9846670/.

[62] Duwayne L. Willett, Vaishnavi Kannan, Ling Chu, Joel R. Buchanan, Ferdinand T. Velasco, John D. Clark, Jason S. Fish, Adolfo R. Ortuzar, Josh E. Youngblood, Deepa G. Bhat, and Mujeeb A. Basit. SNOMED CT Concept Hierarchies for Sharing Definitions of Clinical Conditions Using Electronic Health Record Data. *Applied clinical informatics*, 9(3):667–682, 7 2018. ISSN 1869-0327. doi: 10.1055/S-0038-1668090. URL https://pubmed.ncbi.nlm.nih.gov/30157499/.

[63] Oxford Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC). Oxford Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC): Surveillance reports. Technical report, Oxford Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC), 2024. URL https://www.rcgp.org.uk/representing-you/research-at-rcgp/research-surveillance-centre/public-health-data.

[64] UK Health Security Agency. Seasonal influenza: guidance, data and analysis - GOV.UK. Technical report, UK Health Security Agency, 2025. URL https://www.gov.uk/government/collections/seasonal-influenza-guidance-data-and-analysis.

[65] NHS England Digital. The NHS England Terminology Server - NHS England Digital, 2025. URL https://digital.nhs.uk/services/terminology-server.

[66] John C. O'Horo, Mikhail Dziadzko, Amra Sakusic, Rashid Ali, M. Rizwan Sohail, Daryl J. Kor, and Ognjen Gajic. Seeking out SARI: an automated search of electronic health records. *Epidemiology and Infection*, 146(8):1065, 6 2018. ISSN 14694409. doi: 10.1017/S0950268818000699. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5997502/.

[67] Goldacre Ben. Better, broader, safer: using health data for research and analysis - GOV.UK. Chapter 6 Data Curation. Technical report, Department of Health and Social Care, London, 2022. URL https://www.gov.uk/government/public

ations/better-broader-safer-using-health-data-for-research-and-analysis.

[68] Joan Puig-Barberà, Anita Tormos, Anna Sominina, Elena Burtseva, Odile Launay, Meral A. Ciblak, Angels Natividad-Sancho, Amparo Buigues-Vila, Sergio Martínez-Úbeda, and Cedric Mahé. S061: First-year results of the Global Influenza Hospital Surveillance Network: 2012-2013 Northern hemisphere influenza season. *BMC public health*, 14(1), 6 2014. ISSN 1471-2458. doi: 10.1186/1471-2458-14-564. URL https://pubmed.ncbi.nlm.nih.gov/24903737/.

[69] Minh Nhat Le, Lay Myint Yoshida, Motoi Suzuki, Hien Anh Nguyen, Huu Tho Le, Hiroyuki Moriuchi, Duc Anh Dang, and Koya Ariyoshi. S065: Impact of 2009 pandemic influenza among Vietnamese children based on a population-based prospective surveillance from 2007 to 2011. *Influenza and other respiratory viruses*, 8(4):389–396, 2014. ISSN 1750-2659. doi: 10.1111/IRV.12244. URL https://pubmed.ncbi.nlm.nih.gov/24602158/.

[70] Ruth Lynfield, Richard Davey, Dominic E. Dwyer, Marcelo H. Losso, Deborah Wentworth, Alessandro Cozzi-Lepri, Kathy Herman-Lamin, Grazyna Cholewinska, Daniel David, Stefan Kuetter, Zelalem Ternesgen, Timothy M. Uyeki, H. Clifford Lane, Jens Lundgren, and James D. Neaton. S082: Outcomes of influenza A(H1N1)pdm09 virus infection: results from two international cohort studies. *PloS one*, 9(7), 7 2014. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0101785. URL https://pubmed.ncbi.nlm.nih.gov/25004134/.

[71] William Elson, Anna Forbes, Jamie Lopez Bernal, Roger Morbey, FD Richard Hobbs, and Simon de Lusignan. Systematic review of outcomes of respiratory virus infections for use in clinical severity indicators in public health surveillance systems. *PROSPERO International prospective register of systematic reviews*, 2023. URL https://www.crd.york.ac.uk/PROSPERO/view/CRD42023460281.

[72] Gavin Jamie. *Starting Snomed: A Beginner's Guide to the Snomed CT Healthcare Terminology*. Independently Published, 2019.

[73] Andrew Meci, Florence Du Breuil, Ana Vilcu, Thibaud Pitel, Caroline Guerrisi, Quentin Robard, Clément Turbelin, Thomas Hanslik, Louise Rossignol, Cécile Souty, and Thierry Blanchon. The Sentiworld project: global mapping of sentinel surveillance networks in general practice. *BMC primary care*, 23(1), 12 2022. ISSN 2731-4553. doi: 10.1186/S12875-022-01776-X. URL https://pubmed.ncbi.nlm.nih.gov/35836123/.

[74] Clarivate. EndNote, 2024.

[75] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid. Rayyan—a web and mobile app for systematic reviews, 2016.

[76] Jotform Inc. Jotform , 2024.

[77] World Bank Country and Lending Groups – World Bank Data Help Desk. URL https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

[78] Charlie McLeod, Richard Norman, Edward Litton, Benjamin R. Saville, Steve Webb, and Thomas L. Snelling. Choosing primary endpoints for clinical trials of health care interventions. *Contemporary Clinical Trials Communications*, 16: 100486, 12 2019. ISSN 24518654. doi: 10.1016/J.CONCTC.2019.100486. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6881606/.

[79] Mohammad Reza Yousefi, Mehrdad Karajizadeh, Mehdi Ghasemian, and Shahram Paydar. Comparing NEWS2, TRISS, and RTS in predicting mortality rate in trauma patients based on prehospital data set: a diagnostic study. *BMC Emergency Medicine*, 24(1):1–10, 12 2024. ISSN 1471227X. doi: 10.1186/S12873-024-01084-W/TABLES/8. URL https://bmcemergmed.biomedcentral.com/articles/10.1186/s12873-024-01084-w.

[80] Shengfeng Wei, Dan Xiong, Jia Wang, Xinmeng Liang, Jingxian Wang, and Yuee Chen. The accuracy of the National Early Warning Score 2 in predicting early death

in prehospital and emergency department settings: a systematic review and meta-analysis. *Annals of translational medicine*, 11(2):95–95, 1 2023. ISSN 2305-5839. doi: 10.21037/ATM-22-6587. URL https://pubmed.ncbi.nlm.nih.gov/36819553/.

[81] Global Influenza Programme (GIP). Global epidemiological surveillance standards for influenza. Technical report, The World Health Organization, 2013. URL https://www.who.int/publications/i/item/9789241506601.

[82] Bronke Boudewijns, Saverio Caini, Marco Del Riccio, Marta C Nunes, Sandra S Chaves, Melissa K Andrew, Justin R Ortiz, Oana Săndulescu, Joseph S Bresee, Elena Burtseva, Daouda Coulibaly, Daria M Danilenko, Kirill Stolyarov, Anca C Drăgănescu, Mine Durusu Tanriover, Heloisa I G Giamberardino, Parvaiz A Koul, F. Xavier Lopez-Labrador, Shelly A McNeil, Ainara Mira-Iglesias, Alejandro Orrico-Sanchez, Nancy A Otieno, Jorim Ayugi, Sonia M Raboni, and Peter Spreeuwenberg. Severity Scale of Influenza and Acute Respiratory Illness Hospitalizations to Support Viral Genomic Surveillance: A Global Influenza Hospital Surveillance Network Pilot Study. *Influenza and Other Respiratory Viruses*, 19 (3):e70085, 3 2025. ISSN 1750-2659. doi: 10.1111/IRV.70085. URL https://onlinelibrary.wiley.com/doi/full/10.1111/irv.70085https://onlinelibrary.wiley.com/doi/abs/10.1111/irv.70085https://onlinelibrary.wiley.com/doi/10.1111/irv.70085.

[83] F. Martinón-Torres, M. Carmo, L. Platero, G. Drago, Jl López-Belmonte, M. Bangert, and J. Díez-Domingo. Clinical and economic hospital burden of acute respiratory infection (BARI) due to respiratory syncytial virus in Spanish children, 2015–2018. *BMC Infectious Diseases*, 23(1):385, 12 2023. ISSN 14712334. doi: 10.1186/S12879-023-08358-X. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10249572/.

[84] Bryan Williams. The National Early Warning Score: from concept to NHS im-

plementation. *Clinical Medicine*, 22(6):499–505, 11 2022. ISSN 1470-2118. doi: 10.7861/CLINMED.2022-NEWS-CONCEPT.

[85] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22(7):707–710, 1996. ISSN 03424642. doi: 10.1007/BF01709751/METRICS. URL https://link.springer.com/article/10.1007/BF01709751.

[86] Chendi Cui, Tristan T. Timbrook, Cate Polacek, Zoe Heins, and Ning A. Rosenthal. Disease burden and high-risk populations for complications in patients with acute respiratory infections: a scoping review. *Frontiers in Medicine*, 11:1325236, 5 2024. ISSN 2296858X. doi: 10.3389/FMED.2024.1325236/BIBTEX.

[87] Ahmed M. Al Rajeh, Abdallah Y. Naser, Rayan Siraj, Abdulrhman Alghamdi, Jaber Alqahtani, Yousef Aldabayan, Abdulelah Aldhahir, Ahmed Al Haykan, and Yousif Mohammed Elmosaad. Acute upper respiratory infections admissions in England and Wales. *Medicine*, 102(21):e33616, 5 2023. ISSN 15365964. doi: 10.1097/MD.0000000000033616. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10219745/.

[88] UK Health Security Agency. FluSurvey, 2025. URL https://flusurvey.net/en.

[89] Saskia Jünger, Sheila A. Payne, Jenny Brine, Lukas Radbruch, and Sarah G. Brearley. Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative Medicine*, 31(8):684–706, 9 2017. ISSN 1477030X. doi: 10.1177/0269216317690685. URL https://journals.sagepub.com/doi/10.1177/0269216317690685.

[90] NHS England Digital. Hospital Episode Statistics (HES), 2025. URL https://digital.nhs.uk/services/hospital-episode-statistics.

[91] General Medical Council. Good medical practice - professional standards - GMC. Domain 3: Colleagues, culture and safety. Technical report, General Medical Council, London, 2024. URL https://www.gmc-uk.org/professional-standards/the-professional-standards/good-medical-practice.

[92] NHS England. High quality patient records, 2022. URL https://www.england.nhs.uk/long-read/high-quality-patient-records/.

[93] Nicole G. Weiskopf, George Hripcsak, Sushmita Swaminathan, and Chunhua Weng. Defining and measuring completeness of electronic health records for secondary use. *Journal of Biomedical Informatics*, 46(5):830–836, 10 2013. ISSN 1532-0464. doi: 10.1016/J.JBI.2013.06.010.

[94] Robert A. Verheij, Vasa Curcin, Brendan C. Delaney, and Mark M. McGilchrist. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. *J Med Internet Res 2018;20(5):e185 https://www.jmir.org/2018/5/e185*, 20 (5):e9134, 5 2018. ISSN 14388871. doi: 10.2196/JMIR.9134. URL https://www.jmir.org/2018/5/e185.

[95] Nicole Gray Weiskopf and Chunhua Weng. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association : JAMIA*, 20(1):144, 2013. ISSN 10675027. doi: 10.1136/AMIAJNL-2011-000681. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3555312/.

[96] Michael G Kahn, Tiffany J Callahan, Juliana Barnard, Alan E Bauck, Michael G ; Kahn, Tiffany J ; Callahan, Juliana ; Barnard, Alan E ; Bauck, Jeff ; Brown, Bruce N ; Davidson, Hossein ; Estiri, Carsten ; Goerg, Erin ; Holve, Steven G ; Johnson, Siaw-Teng ; Liaw, Marianne ; Hamilton-Lopez, Daniella ; Meeker, Toan C ; Ong, Patrick ; Ryan, Ning ; Shang, Nicole G ; Weiskopf, Chunhua ; Weng, Meredith N ; Zozus, and Lisa Schilling. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic

Health Record Data. *eGEMs*, 4(1):1244, 9 2016. doi: 10.13063/2327-9214.1244. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5051581/.

[97] Meng Chen Hsu, Chi Chuan Wang, Ling Ya Huang, Chih Ying Lin, Fang Ju Lin, and Sengwee Toh. Effect of ICD-9-CM to ICD-10-CM coding system transition on identification of common conditions: An interrupted time series analysis. *Pharmacoepidemiology and drug safety*, 30(12):1653–1674, 12 2021. ISSN 1099-1557. doi: 10.1002/PDS.5330. URL https://pubmed.ncbi.nlm.nih.gov/34258812/.

[98] Thomas Beaney, Jonathan Clarke, David Salman, Thomas Woodcock, Azeem Majeed, Mauricio Barahona, and Paul Aylin. Identifying potential biases in code sequences in primary care electronic healthcare records: a retrospective cohort study of the determinants of code frequency. *BMJ Open*, 13(9):e072884, 9 2023. ISSN 2044-6055. doi: 10.1136/BMJOPEN-2023-072884. URL https://bmjopen.bmj.com/content/13/9/e072884https://bmjopen.bmj.com/content/13/9/e072884.abstract.

[99] Patrick Rockenschaub, Vincent Nguyen, Robert W. Aldridge, Dionisio Acosta, Juan Miguel García-Gómez, and Carlos Sáez. Data-driven discovery of changes in clinical code usage over time: a case-study on changes in cardiovascular disease recording in two English electronic health records databases (2001–2015). *BMJ Open*, 10(2):e034396, 2 2020. ISSN 2044-6055. doi: 10.1136/BMJOPEN-2019-034396. URL https://bmjopen.bmj.com/content/10/2/e034396https://bmjopen.bmj.com/content/10/2/e034396.abstract.

[100] I. M. Carey, C. M. Nightingale, S. DeWilde, T. Harris, P. H. Whincup, and D. G. Cook. Blood pressure recording bias during a period when the Quality and Outcomes Framework was introduced. *Journal of Human Hypertension 2009 23:11*, 23(11):764–770, 3 2009. ISSN 1476-5527. doi: 10.1038/jhh.2009.18. URL https://www.nature.com/articles/jhh200918.

[101] Helen J. Curtis, Brian MacKenna, Milan Wiedemann, Louis Fisher, Richard Croker, Caroline E. Morton, Peter Inglesby, Alex J. Walker, Jessica Morley, Amir

Mehrkar, Sebastian C.J. Bacon, George Hickman, David Evans, Tom Ward, Simon Davy, William J. Hulme, Orla Macdonald, Robin Conibere, Tom Lewis, Martin Myers, Shamila Wanninayake, Kiren Collison, Charles Drury, Miriam Samuel, Harpreet Sood, Andrea Cipriani, Seena Fazel, Manuj Sharma, Wasim Baqir, Chris Bates, John Parry, and Ben Goldacre. OpenSAFELY NHS Service Restoration Observatory 2: changes in primary care clinical activity in England during the COVID-19 pandemic. *The British Journal of General Practice*, 73 (730):e318, 5 2023. ISSN 14785242. doi: 10.3399/BJGP.2022.0301. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10131234/.

[102] NHS England Digital. National Data Opt-Out, . URL https://digital.nhs.uk/services/national-data-opt-out.

[103] Digital Health. Fit notes finally go electronic | Digital Health, 2012. URL https://www.digitalhealth.net/2012/08/fit-notes-finally-go-electronic.

[104] NHS England. Personal demographic service (PDS), 2023. URL https://www.england.nhs.uk/long-read/personal-demographic-service-pds/.

[105] NHS England Digital. Mandating mortality data updates onto the Personal Demographics Service , 2021. URL https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/directions-and-data-provision-notices/data-provision-notices-dpns/mortality-data-flows.

[106] Martin C Gulliford, Judith Charlton, Olga Boiko, Joanne R Winter, Emma Rezel-Potts, Xiaohui Sun, Caroline Burgess, Lisa McDermott, Catey Bunce, James Shearer, Vasa Curcin, Robin Fox, Alastair D Hay, Paul Little, Michael V Moore, and Mark Ashworth. Sepsis recording in primary care electronic health records, linked hospital episodes and mortality records. *Health Services and Delivery Research*, 2021. URL https://www.ncbi.nlm.nih.gov/books/NBK570463/.

[107] UKHSA. Public reminded to stay safe as COVID-19 England restrictions lift -

GOV.UK, 2022. URL https://www.gov.uk/government/news/public-reminded-to-stay-safe-as-covid-19-england-restrictions-lift.

[108] B. Holden, A. Quinney, S. Padfield, W. Morton, S. Coles, P. Manley, A. Wensley, C. Hutchinson, P. J. Lillie, C. J.A. Duncan, M. L. Schmid, A. Li, K. Foster, S. Anaraki, G. Dabrera, M. Zambon, G. J. Hughes, and M. Gent. COVID-19: public health management of the first two confirmed cases identified in the UK. *Epidemiology & Infection*, 148:e194, 2020. ISSN 0950-2688. doi: 10.1017/S0950268820001922. URL https://www.cambridge.org/core/journals/epidemiology-and-infection/article/covid19-public-health-management-of-the-first-two-confirmed-cases-identified-in-the-uk/86BFB6BE58CA5260915C3E4B45D31D59.

[109] Claire Blacklock, Tanya Ali Haj-Hassan, and Matthew J. Thompson. When and how do GPs record vital signs in children with acute infections? A cross-sectional study. *The British Journal of General Practice*, 62(603):e679, 10 2012. ISSN 09601643. doi: 10.3399/BJGP12X656810. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3459775/.

[110] Nick A. Francis, Jochen W. Cals, Christopher C. Butler, Kerenza Hood, Theo Verheij, Paul Little, Herman Goossens, and Samuel Coenen. Severity assessment for lower respiratory tract infections: potential use and validity of the CRB-65 in primary care. *Primary care respiratory journal : journal of the General Practice Airways Group*, 21(1):65–70, 3 2012. ISSN 1475-1534. doi: 10.4104/PCRJ.2011.00083. URL https://pubmed.ncbi.nlm.nih.gov/21938349/.

[111] Merijn H. Rijk, Tamara N. Platteel, Marissa M.M. Mulder, Geert Jan Geersing, Frans H. Rutten, Maarten van Smeden, Roderick P. Venekamp, and Tuur M. Leeuwenberg. Incomplete and possibly selective recording of signs, symptoms, and measurements in free text fields of primary care electronic health records of adults with lower respiratory tract infections. *Journal of Clinical Epidemiology*, 166:111240, 2 2024. ISSN 0895-4356. doi: 10.1016/J.JCLINEPI.2023.111240.

[112] Mairead Murphy, Lauren J. Scott, Chris Salisbury, Andrew Turner, Anne Scott, Rachel Denholm, Rhys Lewis, Geeta Iyer, John Macleod, and Jeremy Horwood. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. *The British Journal of General Practice*, 71(704):e166, 3 2021. ISSN 14785242. doi: 10.3399/BJGP.2020.0948. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7909923/.

[113] Office for National Statistics. Coronavirus (COVID-19) Infection Survey technical article, 2021. URL https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticle/wavesandlagsofcovid19inenglandjune2021.

[114] Salwa S. Zghebi, David Reeves, Christos Grigoroglou, Brian McMillan, Darren M. Ashcroft, Rosa Parisi, and Evangelos Kontopantelis. Clinical code usage in UK general practice: a cohort study exploring 18 conditions over 14 years. *BMJ Open*, 12(7):e051456, 7 2022. ISSN 20446055. doi: 10.1136/BMJOPEN-2021-051456. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9328099/.

[115] Cliodna McNulty, Meredith Hawking, Donna Lecky, Leah Jones, Rebecca Owens, André Charlett, Chris Butler, Philippa Moore, and Nick Francis. Effects of primary care antimicrobial stewardship outreach on antibiotic use by general practice staff: pragmatic randomized controlled trial of the TARGET antibiotics workshop. *Journal of Antimicrobial Chemotherapy*, 73(5):1423, 5 2018. ISSN 14602091. doi: 10.1093/JAC/DKY004. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5909634/.

[116] Virginia Hernandez-Santiago, Charis A. Marwick, Andrea Patton, Peter G. Davey, Peter T. Donnan, and Bruce Guthrie. Time series analysis of the impact of an intervention in Tayside, Scotland to reduce primary care broad-spectrum antimicrobial use. *Journal of Antimicrobial Chemotherapy*, 70(8):2397–2404,

8 2015. ISSN 0305-7453. doi: 10.1093/JAC/DKV095. URL https://dx.doi.org/10.1093/jac/dkv095.

[117] GOV.UK. Helping parents spot the signs of sepsis, 2016. URL https://www.gov.uk/government/news/helping-parents-spot-the-signs-of-sepsis.

[118] NICE. Overview | Suspected sepsis: recognition, diagnosis and early management | Guidance | , 2016. URL https://www.nice.org.uk/guidance/ng51.

[119] Zheyuan Yang, Sabine Bou-Antoun, Sarah Gerver, Thomas E. Cowling, and Rachel Freeman. Sustained increases in antibiotic prescriptions per primary care consultation for upper respiratory tract infections in England during the COVID-19 pandemic. *JAC-Antimicrobial Resistance*, 5(1), 12 2022. ISSN 26321823. doi: 10.1093/JACAMR/DLAD012. URL https://dx.doi.org/10.1093/jacamr/dlad012.

[120] Alice P. McCloskey, Lucy Malabar, Philippa G. McCabe, Andrew Gitsham, and Ian Jarman. Antibiotic prescribing trends in primary care 2014–2022. *Research in Social and Administrative Pharmacy*, 19(8):1193–1201, 8 2023. ISSN 1551-7411. doi: 10.1016/J.SAPHARM.2023.05.001.

[121] Nina J. Zhu, Monsey McLeod, Cliodna A.M. McNulty, Donna M. Lecky, Alison H. Holmes, and Raheelah Ahmad. Trends in Antibiotic Prescribing in Out-of-Hours Primary Care in England from January 2016 to June 2020 to Understand Behaviours during the First Wave of COVID-19. *Antibiotics 2021, Vol. 10, Page 32*, 10(1): 32, 1 2021. ISSN 2079-6382. doi: 10.3390/ANTIBIOTICS10010032. URL https://www.mdpi.com/2079-6382/10/1/32/htmhttps://www.mdpi.com/2079-6382/10/1/32.

[122] B. Adam Williams, Charles H. Jones, Verna Welch, and Jane M. True. Outlook of pandemic preparedness in a post-COVID-19 world. *npj Vaccines 2023 8:1*, 8 (1):1–12, 11 2023. ISSN 2059-0105. doi: 10.1038/s41541-023-00773-0. URL https://www.nature.com/articles/s41541-023-00773-0.

[123] Oliver Eales, Michael J. Plank, Benjamin J. Cowling, Benjamin P. Howden, Adam J. Kucharski, Sheena G. Sullivan, Katelijn Vandemaele, Cecile Viboud, Steven Riley, James M. McCaw, and Freya M. Shearer. Key Challenges for Respiratory Virus Surveillance while Transitioning out of Acute Phase of COVID-19 Pandemic. *Emerging infectious diseases*, 30(2):e1–e9, 2 2024. ISSN 1080-6059. doi: 10.3201/EID3002.230768. URL https://pubmed.ncbi.nlm.nih.gov/38190760/.

[124] HDR UK. The Sudlow Review. URL https://www.hdruk.ac.uk/helping-with-health-data/the-sudlow-review/.

[125] NHS England Digital. Civil Registrations of Death, . URL https://digital.nhs.uk/services/data-access-request-service-dars/dars-products-and-services/data-set-catalogue/civil-registrations-of-death.

[126] S. Schwartz. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *https://doi.org/10.2105/AJPH.84.5.819*, 84(5):819–824, 10 2011. ISSN 00900036. doi: 10.2105/AJPH.84.5.819. URL https://ajph.aphapublications.org/doi/10.2105/AJPH.84.5.819.

[127] John C. Marshall, Srinivas Murthy, Janet Diaz, Neil Adhikari, Derek C. Angus, Yaseen M. Arabi, Kenneth Baillie, Michael Bauer, Scott Berry, Bronagh Blackwood, Marc Bonten, Fernando Bozza, Frank Brunkhorst, Allen Cheng, Mike Clarke, Vu Quoc Dat, Menno de Jong, Justin Denholm, Lennie Derde, Jake Dunning, Xiaobin Feng, Tom Fletcher, Nadine Foster, Rob Fowler, Nina Gobat, Charles Gomersall, Anthony Gordon, Thomas Glueck, Michael Harhay, Carol Hodgson, Peter Horby, Yae Jean Kim, Richard Kojan, Bharath Kumar, John Laffey, Denis Malvey, Ignacio Martin-Loeches, Colin McArthur, Danny McAuley, Stephen McBride, Shay McGuinness, Laura Merson, Susan Morpeth, Dale Needham, Mihai Netea, Myoung Don Oh, Sabai Phyu, Simone Piva, Ruijin Qiu, Halima Salisu-Kabara, Lei Shi, Naoki Shimizu, Jorge Sinclair, Steven Tong, Alexis Turgeon, Tim Uyeki, Frank van de Veerdonk, Steve Webb, Paula Williamson,

Timo Wolf, and Junhua Zhang. A minimal common outcome measure set for COVID-19 clinical research. *The Lancet Infectious Diseases*, 20(8):e192–e197, 8 2020. ISSN 14744457. doi: 10.1016/S1473-3099(20)30483-7. URL https://www.thelancet.com/action/showFullText?pii=S1473309920304837https://www.thelancet.com/action/showAbstract?pii=S1473309920304837https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30483-7/abstract.

[128] Faiez Zannad, Kenneth Stein, Angeles Alonso Garcia, Stefan D. Anker, Paul W. Armstrong, Gonzalo Calvo, John G.F. Cleland, Jay N. Cohn, Kenneth Dickstein, Michael J. Domanski, Inger Ekman, Gerasimos S. Filippatos, Mihai Gheorghiade, Adrian F. Hernandez, Tiny Jaarsma, Joerg Koglin, Marvin Konstam, Stuart Kupfer, Aldo P. Maggioni, Alexandre Mebazaa, Marco Metra, Christina Nowack, Burkert Pieske, Ileana L. Piña, Stuart J. Pocock, Piotr Ponikowski, Giuseppe Rosano, Luis M. Ruilope, Frank Ruschitzka, Thomas Severin, Scott Solomon, Norman L. Stockbridge, Wendy Gattis Stough, Karl Swedberg, Luigi Tavazzi, Adriaan A. Voors, Scott M. Wasserman, Holger Woehrle, Andrew Zalewski, and John J.V. McMurray. Clinical outcome endpoints in heart failure trials: a European Society of Cardiology Heart Failure Association consensus document. *European Journal of Heart Failure*, 15(10):1082–1094, 10 2013. ISSN 1879-0844. doi: 10.1093/EURJHF/HFT095. URL https://onlinelibrary.wiley.com/doi/full/10.1093/eurjhf/hft095https://onlinelibrary.wiley.com/doi/abs/10.1093/eurjhf/hft095https://onlinelibrary.wiley.com/doi/10.1093/eurjhf/hft095.

[129] Jeffrey C. Kwong, Kevin L. Schwartz, Michael A. Campitelli, Hannah Chung, Natasha S. Crowcroft, Timothy Karnauchow, Kevin Katz, Dennis T. Ko, Allison J. McGeer, Dayre McNally, David C. Richardson, Laura C. Rosella, Andrew Simor, Marek Smieja, George Zahariadis, and Jonathan B. Gubbay. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. *New England Journal of Medicine*, 378(4):345–353, 1 2018. ISSN 0028-4793. doi: 10.1056/NEJMOA

1702090/SUPPL{_}FILE/NEJMOA1702090{_}DISCLOSURES.PDF. URL https://www.nejm.org/doi/full/10.1056/NEJMoa1702090.

[130] Jennifer A. Davidson, Amitava Banerjee, Liam Smeeth, Helen I. McDonald, Daniel Grint, Emily Herrett, Harriet Forbes, Richard Pebody, and Charlotte Warren-Gash. Risk of acute respiratory infection and acute cardiovascular events following acute respiratory infection among adults with increased cardiovascular risk in England between 2008 and 2018: a retrospective, population-based cohort study. *The Lancet Digital Health*, 3(12):e773–e783, 12 2021. ISSN 25897500. doi: 10.1016/S2589-7500(21)00203-X. URL https://www.thelancet.com/action/showFullText?pii=S258975002100203Xhttps://www.thelancet.com/action/showAbstract?pii=S258975002100203Xhttps://www.thelancet.com/journals/landig/article/PIIS2589-7500(21)00203-X/abstract.

[131] Madeleine W. Sumner, Alicia Kanngiesser, Kosar Lotfali-Khani, Nidhi Lodha, Diane Lorenzetti, Anna L. Funk, and Stephen B. Freedman. Severe Outcomes Associated With SARS-CoV-2 Infection in Children: A Systematic Review and Meta-Analysis. *Frontiers in Pediatrics*, 10:916655, 6 2022. ISSN 22962360. doi: 10.3389/FPED.2022.916655/FULL. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9218576/.

[132] NHS England. NHS England » National paediatric early warning system (PEWS), 2018. URL https://www.england.nhs.uk/get-involved/cyp/pews/.

[133] R. C. Bone, R. A. Balk, F. B. Cerra, R. P. Dellinger, A. M. Fein, W. A. Knaus, R. M.H. Schein, and W. J. Sibbald. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest*, 101(6):1644–1655, 1992. ISSN 0012-3692. doi: 10.1378/CHEST.101.6.1644. URL https://pubmed.ncbi.nlm.nih.gov/1303622/.

[134] National Institute of Clinical Excellence. Recommendations | Suspected acute respiratory infection in over 16s: assessment at first presentation and initial management

| Guidance | NICE, 2023. URL https://www.nice.org.uk/guidance/ng237/chapter/Recommendations.

[135] BARNET WOOLF. ON ESTIMATING THE RELATION BETWEEN BLOOD GROUP AND DISEASE. *Annals of Human Genetics*, 19(4):251–253, 5 1955. ISSN 1469-1809. doi: 10.1111/J.1469-1809.1955.TB01348.X. URL https://onlinelibrary.wiley.com/doi/full/10.1111/j.1469-1809.1955.tb01348.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1955.tb01348.xhttps://onlinelibrary.wiley.com/doi/10.1111/j.1469-1809.1955.tb01348.x.

[136] Morten W Fagerland. Exact and mid-p confidence intervals for the odds ratio. *The Stata Journal*, 12(3):505–514, 2012.

[137] Titus von der Malsburg and Bernhard Angele. False Positives and Other Statistical Errors in Standard Analyses of Eye Movements in Reading. *Journal of memory and language*, 94:119, 6 2016. ISSN 0749596X. doi: 10.1016/J.JML.2016.10.003. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5461930/.

[138] Craig B. Borkowf, Paul S. Albert, and Christian C. Abnet. Using lowess to remove systematic trends over time in predictor variables prior to logistic regression with quantile categories. *Statistics in medicine*, 22(9):1477–1493, 5 2003. ISSN 0277-6715. doi: 10.1002/SIM.1507. URL https://pubmed.ncbi.nlm.nih.gov/12704611/https://pubmed.ncbi.nlm.nih.gov/12704611/?utm_source=chatgpt.com.

[139] Robert H.. Shumway and David S.. Stoffer. *Time series analysis and its applications : with R examples*. Springer, 2025. ISBN 9783031705847.

[140] Eric H.Y. Lau, Calvin K.Y. Cheng, Dennis K.M. Ip, and Benjamin J. Cowling. Potential use of multiple surveillance data in the forecast of hospital admissions. *Online Journal of Public Health Informatics*, 5(1):e168, 5 2013. ISSN 19326203.

doi: 10.1371/JOURNAL.PONE.0038346. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3692818/.

[141] Jens Peter Kreiss and Soumendra Nath Lahiri. Bootstrap Methods for Time Series. *Handbook of Statistics*, 30:3–26, 1 2012. ISSN 0169-7161. doi: 10.1016/B978-0-444-53858-1.00001-6.

[142] Robert H Shumway, David S Stoffer, and With R Examples. *Time Series Analysis and It's Applications EZ Green Edition*. Springer, 2013. URL http://en.wikibooks.org/wiki/Subject:K-12_mathematics.

[143] Christine P.. Dancey and John. Reidy. *Statistics without maths for psychology : using SPSS for Windows*. Pearson/Prentice Hall, 2007. ISBN 0132051605. URL https://books.google.com/books/about/Statistics_Without_Maths_for_Psychology.html?id=QjfQ0_DqyNQC.

[144] Ana Espinosa-Gonzalez, Denys Prociuk, Francesca Fiorentino, Christian Ramtale, Ella Mi, Emma Mi, Ben Glampson, Ana Luisa Neves, Cecilia Okusi, Laiba Husain, Jack Macartney, Martina Brown, Ben Browne, Caroline Warren, Rachna Chowla, Jonty Heaversedge, Trisha Greenhalgh, Simon de Lusignan, Erik Mayer, and Brendan C. Delaney. Remote COVID-19 Assessment in Primary Care (RECAP) risk prediction tool: derivation and real-world validation studies. *The Lancet Digital Health*, 4(9):e646–e656, 9 2022. ISSN 25897500. doi: 10.1016/S2589-7500(22)00123-6. URL https://www.thelancet.com/action/showFullText?pii=S2589750022001236https://www.thelancet.com/action/showAbstract?pii=S2589750022001236https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00123-6/abstract.

[145] Robin Bruyndonckx, Niel Hens, Theo J.M. Verheij, Marc Aerts, Margareta Ieven, Christopher C. Butler, Paul Little, Herman Goossens, and Samuel Coenen. Development of a prediction tool for patients presenting with acute cough in primary care: a prognostic study spanning six European countries. *The British Journal of General*

Practice, 68(670):e342, 5 2018. ISSN 14785242. doi: 10.3399/BJGP18X695789. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5916081/.

[146] Alastair D. Hay, Niamh M. Redmond, Sophie Turnbull, Hannah Christensen, Hannah Thornton, Paul Little, Matthew Thompson, Brendan Delaney, Andrew M. Lovering, Peter Muir, John P. Leeming, Barry Vipond, Beth Stuart, Tim J. Peters, and Peter S. Blair. Development and internal validation of a clinical rule to improve antibiotic use in children presenting to primary care with acute respiratory tract infection and cough: a prognostic cohort study. *The Lancet. Respiratory Medicine*, 4(11):902, 11 2016. ISSN 22132619. doi: 10.1016/S2213-2600(16)30223-5. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5080970/.

[147] Wilke Hendriks, Hendriek Boshuizen, Arnold Dekkers, Mirjam Knol, Ge A. Donker, Arie van der Ende, and Hester Korthals Altes. Temporal cross-correlation between influenza-like illnesses and invasive pneumococcal disease in The Netherlands. *Influenza and other respiratory viruses*, 11(2):130–137, 3 2017. ISSN 1750-2659. doi: 10.1111/IRV.12442. URL https://pubmed.ncbi.nlm.nih.gov/27943624/https://pubmed.ncbi.nlm.nih.gov/27943624/?utm_source=chatgpt.com.

[148] Lin Yang, Chit Ming Wong, Eric H.Y. Lau, King Pan Chan, Chun Quan Ou, and Joseph S.M. Peiris. Synchrony of Clinical and Laboratory Surveillance for Influenza in Hong Kong. *PLoS ONE*, 3(1):e1399, 1 2008. ISSN 19326203. doi: 10.1371/JOURNAL.PONE.0001399. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC2151138/.

[149] Hicham Achebak, Judith Garcia-Aymerich, Grégoire Rey, Zhaoyue Chen, Raúl Fernando Méndez-Turrubiates, and Joan Ballester. Ambient temperature and seasonal variation in inpatient mortality from respiratory diseases: a retrospective observational study. *The Lancet Regional Health - Europe*, 35:100757, 12 2023. ISSN 26667762. doi: 10.1016/j.lanepe.2023.100757. URL https://www.thelancet.com/action/showFullText?pii=S2666776

22300176Xhttps://www.thelancet.com/action/showAbstract?pii=S266677622300176Xhttps://www.thelancet.com/journals/lanepe/article/PIIS2666-7762(23)00176-X/abstract.

[150] Renjie Chen, Peng Yin, Lijun Wang, Cong Liu, Yue Niu, Weidong Wang, Yixuan Jiang, Yunning Liu, Jiangmei Liu, Jinlei Qi, Jinling You, Haidong Kan, and Maigeng Zhou. Association between ambient temperature and mortality risk and burden: time series study in 272 main Chinese cities. *BMJ*, 363, 2018. ISSN 17561833. doi: 10.1136/bmj.k4306.

[151] S. Hajat, R. S. Kovats, and K. Lachowycz. Heat-related and cold-related deaths in England and Wales: who is at risk? *Occupational and Environmental Medicine*, 64(2):93, 2 2006. ISSN 13510711. doi: 10.1136/OEM.2006.029017. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC2078436/.

[152] UKHSA. Health Effects of Climate Change (HECC) in the UK: 2023 report Chapter 2. Temperature effects on mortality in a changing climate. Technical report, UKHSA, 2023.

[153] NHS England » Winter pressures, . URL https://www.england.nhs.uk/long-read/winter-pressures/?utm_source=chatgpt.com.

[154] Sinead Millwood, Peter Tomlinson, and Jon Hopwood. Evaluation of winter pressures on general practice in Manchester: a cross-sectional analysis of nine GP practices. *BJGP Open*, 5(1):bjgpopen20X101138, 2021. ISSN 23983795. doi: 10.3399/BJGPOPEN20X101138. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7960527/.

[155] Excess deaths in England and Wales - Office for National Statistics. URL https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/excessdeathsinenglandandwalesmarch2020todecember2022/2023-03-09#excess-deaths-by-month.

[156] Meredith G. Wesley, Yeny Tinoco, Archana Patel, Piyarat Suntarratiwong, Danielle Hunt, Chalinthorn Sinthuwattanawibool, Giselle Soto, Wanitchaya Kittikraisak, Prabir Kumar Das, Carmen Sofia Arriola, Danielle Hombroek, Joshua Mott, Kunal Kurhe, Savita Bhargav, Amber Prakash, Richard Florian, Oswaldo Gonzales, Santiago Cabrera, Edwin Llajaruna, Tana Brummer, Parker Malek, Siddhartha Saha, Shikha Garg, Eduardo Azziz-Baumgartner, Mark G. Thompson, and Fatimah S. Dawood. Performance of Symptom-Based Case Definitions to Identify Influenza Virus Infection Among Pregnant Women in Middle-Income Countries: Findings From the Pregnancy and Influenza Multinational Epidemiologic (PRIME) Study. *Clinical Infectious Diseases*, 73(11):e4321–e4328, 12 2021. ISSN 1058-4838. doi: 10.1093/CID/CIAA1697. URL https://dx.doi.org/10.1093/cid/ciaa1697.

[157] Similarities and Differences between Flu and COVID-19 | Influenza (Flu) | CDC. URL https://www.cdc.gov/flu/about/flu-vs-covid19.html?utm_source=chatgpt.com.

[158] C. P. Farrington, N. J. Andrews, A. D. Beale, and M. A. Catchpole. A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 159(3):547–563, 5 1996. ISSN 0964-1998. doi: 10.2307/2983331. URL https://dx.doi.org/10.2307/2983331.

[159] Tommy Nyberg, Neil M. Ferguson, Sophie G. Nash, Harriet H. Webster, Seth Flaxman, Nick Andrews, Wes Hinsley, Jamie Lopez Bernal, Meaghan Kall, Samir Bhatt, Paula Blomquist, Asad Zaidi, Erik Volz, Nurin Abdul Aziz, Katie Harman, Sebastian Funk, Sam Abbott, Jamie Lopez Bernal, Nurin Abdul Aziz, Russell Hope, Andre Charlett, Meera Chand, Azra C. Ghani, Shaun R. Seaman, Gavin Dabrera, Daniela De Angelis, Anne M. Presanis, and Simon Thelwall. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. *The Lancet*,

399(10332):1303–1312, 4 2022. ISSN 1474547X. doi: 10.1016/S0140-6736(22)00462-7. URL https://www.thelancet.com/action/showFullText?pii=S0140673622004627https://www.thelancet.com/action/showAbstract?pii=S0140673622004627https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(22)00462-7/abstract.

[160] Knut Øymar, Håvard O. Skjerven, and Ingvild B. Mikalsen. Acute bronchiolitis in infants, a review. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 22(1):23–, 4 2014. ISSN 17577241. doi: 10.1186/1757-7241-22-23/FIGURES/1. URL https://sjtrem.biomedcentral.com/articles/10.1186/1757-7241-22-23.

[161] German RR. Updated Guidelines for Evaluating Public Health Surveillance Systems. *MMWR*, 50((RR13)):1–35, 2001. URL https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm.

[162] Samar Binkheder, Mohammed Ahmed Asiri, Khaled Waleed Altowayan, Turki Mohammed Alshehri, Mashhour Faleh Alzarie, Raniah N. Aldekhyyel, Ibrahim A. Almaghlouth, and Jwaher A. Almulhem. Real-World Evidence of COVID-19 Patients' Data Quality in the Electronic Health Records. *Healthcare*, 9(12):1648, 12 2021. ISSN 22279032. doi: 10.3390/HEALTHCARE9121648. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8701465/.

[163] Macarena Garcia, Nikolay Lipskiy, James Tyson, Roniqua Watkins, E. Stein Esser, and Teresa Kinley. Centers for Disease Control and Prevention 2019 novel coronavirus disease (COVID-19) information management: addressing national healthcare and public health needs for standardized data definitions and codified vocabulary for data exchange. *Journal of the American Medical Informatics Association : JAMIA*, 27(9):1476–1487, 9 2020. ISSN 1527-974X. doi: 10.1093/JAMIA/OCAA141. URL https://pubmed.ncbi.nlm.nih.gov/32940705/.

[164] NHS England » NHS approves new IT system for GPs to help transform care, .

URL https://www.england.nhs.uk/2025/06/nhs-approves-new-it-system-for-gps-to-help-transform-care/.

[165] RCGP-RSC. RCGP-RSC Annual report 2019/20. Technical report. URL https://www.rcgp.org.uk/representing-you/research-at-rcgp/research-surveillance-centre/public-health-data.

[166] Tomás Vega, Jose Eugenio Lozano, Tamara Meerhoff, René Snacken, Joshua Mott, Raul Ortiz de Lejarazu, and Baltazar Nunes. Influenza surveillance in Europe: Establishing epidemic thresholds by the Moving Epidemic Method. *Influenza and other Respiratory Viruses*, 7(4):546–558, 7 2013. ISSN 17502640. doi: 10.1111/J.1750-2659.2012.00422.X.

[167] Bushra Zareie, Jalal Poorolajal, Amin Roshani, and Manoochehr Karami. Outbreak detection algorithms based on generalized linear model: a review with new practical examples. *BMC Medical Research Methodology*, 23(1):235–, 12 2023. ISSN 14712288. doi: 10.1186/S12874-023-02050-Z/TABLES/3. URL https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-02050-zhttp://creativecom-mons.org/publicdomain/zero/1.0/.

[168] EUROMOMO. URL https://www.euromomo.eu/.

[169] FluSight | FluSight | CDC. URL https://www.cdc.gov/flu-forecasting/index.html.

[170] Niall Boyce. Bills of Mortality: tracking disease in early modern London. *Lancet (London, England)*, 395(10231):1186, 4 2020. ISSN 1474547X. doi: 10.1016/S0140-6736(20)30725-X. URL /pmc/articles/PMC7154511/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7154511/.

[171] Kim A. Lindblade, Wences Arvelo, Jennifer Gray, Alejandra Estevez, Gal Frenkel, Lissette Reyes, Fabiola Moscoso, Juan Carlos Moir, Alicia M. Fry, and Sonja J. Olsen. S001: A Comparison of the Epidemiology and Clinical Presentation of Seasonal Influenza A and 2009 Pandemic Influenza A (H1N1) in Guatemala. *PLoS*

ONE, 5(12):e15826, 2010. ISSN 19326203. doi: 10.1371/JOURNAL.PONE.0015826. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3012722/.

[172] Oliver Viera-Segura, Natali Vega-Magaña, Mariel García-Chagollán, Marcela Peña-Rodríguez, Germán Muñoz-Sánchez, Ahtziri Socorro Carranza-Aranda, Iris Monserrat Llamas-Covarrubias, Moisés Ramos-Solano, Jesús Mora-Mora, Carlos Daniel Díaz-Palomera, Gabriela Espinoza De León, José Sergio Zepeda-Nuño, Enrique Santillán-López, Samuel García-Arellano, Christian David Hernández-Silva, Darbi Alfredo Zerpa-Hernandez, Guillermina Muñoz-Rios, J. Samael Rodríguez-Sanabria, and José Francisco Muñoz-Valle. S002: A Comprehensive Descriptive Epidemiological and Clinical Analysis of SARS-CoV-2 in West-Mexico during COVID-19 Pandemic 2020. *International journal of environmental research and public health*, 18(20), 10 2021. ISSN 1660-4601. doi: 10.3390/IJERPH182010644. URL https://pubmed.ncbi.nlm.nih.gov/34682388/.

[173] Darpanarayan Hazra, Gina Maryann Chandy, Abirahmi Thanjavurkar, Karthik Gunasekaran, Ankita Chowdary Nekkanti, Rathijit Pal, Mahesh Moorthy, and Kundavaram Paul Prabhakar Abhilash. S003: A clinico-epidemiological profile, coinfections and outcome of patients with Influenza Like Illnesses (ILI) presenting to the emergency department during the COVID-19 pandemic. *Journal of family medicine and primary care*, 12(4):672–678, 4 2023. ISSN 2249-4863. doi: 10.4103/JFMPC.JFMPC{_}1705{_}22. URL https://pubmed.ncbi.nlm.nih.gov/37312766/.

[174] Jeffrey J. VanWormer, Maria E. Sundaram, Jennifer K. Meece, and Edward A. Belongia. S004: A cross-sectional analysis of symptom severity in adults with influenza and other acute respiratory illness in the outpatient setting. *BMC Infectious Diseases*, 14(1):1–10, 5 2014. ISSN 14712334. doi: 10.1186/1471-2334-14-231/TABLES/3. URL https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-14-231.

[175] E. D'Ortenzio, P. Renault, M. C. Jaffar-Bandjee, B. A. Gaüzère, M. Lagrange-

Xélot, A. Fouillet, P. Poubeau, A. Winer, A. Bourde, F. Staikowsky, P. Morbidelli, E. Rachou, F. Thouillot, A. Michault, and L. Filleul. S005: A review of the dynamics and severity of the pandemic A(H1N1) influenza virus on Réunion island, 2009. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 16(4):309–316, 2010. ISSN 1469-0691. doi: 10.1111/J.1469-0691.2010.03171.X. URL https://pubmed.ncbi.nlm.nih.gov/20121825/.

[176] Seong Hui Kang, Hee Jin Cheong, Joon Young Song, Ji Yun Noh, Ji Ho Jeon, Min Joo Choi, Jacob Lee, Yu Bin Seo, Jin Soo Lee, Seong Heon Wie, Hye Won Jeong, Young Keun Kim, Kyung Hwa Park, Shin Woo Kim, Eun Joo Jeong, Sun Hee Lee, Won Suk Choi, and Woo Joo Kim. S006: Analysis of Risk Factors for Severe Acute Respiratory Infection and Pneumonia and among Adult Patients with Acute Respiratory Illness during 2011-2014 Influenza Seasons in Korea. *Infection & Chemotherapy*, 48(4):294, 2016. ISSN 20926448. doi: 10.3947/IC.2016.48.4.294. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5204008/.

[177] Ibrahim Franklyn Kamara, Ajay M.V. Kumar, Anna Maruta, Bobson Derrick Fofanah, Charles Kuria Njuguna, Steven Shongwe, Francis Moses, Sia Morenike Tengbe, Joseph Sam Kanu, Sulaiman Lakoh, Alie H.D. Mansaray, Kalaiselvi Selvaraj, Mohammed Khogali, and Rony Zachariah. S007: Antibiotic Use in Suspected and Confirmed COVID-19 Patients Admitted to Health Facilities in Sierra Leone in 2020–2021: Practice Does Not Follow Policy. *International Journal of Environmental Research and Public Health*, 19(7):4005, 4 2022. ISSN 16604601. doi: 10.3390/IJERPH19074005. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8998021/.

[178] Dianne Zakaria, Samina Aziz, Sharon Bartholomew, Su Bin Park, Cynthia Robitaille, and Murray Weeks. S008: Associations between chronic conditions and death in hospital among adults (aged 20+ years) during first acute care hospitalizations with a confirmed or suspected COVID-19 diagnosis in Canada. *PloS one*, 18

(1), 1 2023. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0280050. URL https://pubmed.ncbi.nlm.nih.gov/36598923/.

[179] Craig B. Dalton, Sandra J. Carlson, Michelle T. Butler, Elissa Elvidge, and David N. Durrheim. S009: Building influenza surveillance pyramids in near real time, Australia. *Emerging infectious diseases*, 19(11):1863–1865, 11 2013. ISSN 1080-6059. doi: 10.3201/EID1911.121878. URL https://pubmed.ncbi.nlm.nih.gov/24207165/.

[180] Ariana Perez, Joana Y. Lively, Aaron Curns, Geoffrey A. Weinberg, Natasha B. Halasa, Mary Allen Staat, Peter G. Szilagyi, Laura S. Stewart, Monica M. McNeal, Benjamin Clopper, Yingtao Zhou, Brett L. Whitaker, Elizabeth LeMasters, Elizabeth Harker, Janet A. Englund, Eileen J. Klein, Rangaraj Selvarangan, Christopher J. Harrison, Julie A. Boom, Leila C. Sahni, Marian G. Michaels, John V. Williams, Gayle E. Langley, Susan I. Gerber, Angela Campbell, Aron J. Hall, Brian Rha, Meredith McMorrow, Bonnie Strelitz, Kirsten Lacombe, Mary Moffatt, Jennifer Schuster, Chelsea Rohlfs, Miranda Howard, Yesenia Romero, James Chappell, Pedro A. Piedra, Vasanthi Avadhanula, Wende Fregoe, Christina Albertin, Robert H. Hickey, and Judith M Martin. S010: Respiratory Virus Surveillance Among Children with Acute Respiratory Illnesses — New Vaccine Surveillance Network, United States, 2016–2021. *MMWR. Morbidity and Mortality Weekly Report*, 71(40): 1253–1259, 10 2022. ISSN 0149-21951545-861X. doi: 10.15585/MMWR.MM7140A1. URL https://www.cdc.gov/mmwr/volumes/71/wr/mm7140a1.htm.

[181] Johannes Leiner, Sven Hohenstein, Vincent Pellissier, Sebastian König, Claudia Winklmair, Irit Nachtigall, Andreas Bollmann, and Ralf Kuhlen. S011: COVID-19 and Severe Acute Respiratory Infections: Monitoring Trends in 421 German Hospitals During the First Four Pandemic Waves. *S011: Infection and drug resistance*, 16:2775–2781, 2023. ISSN 1178-6973. doi: 10.2147/IDR.S402313. URL https://pubmed.ncbi.nlm.nih.gov/37187482/.

[182] Jana Seligová, Andrea Čulmanová, Zuzana Krištufková, Lydia Čisláková, and

Henrieta Hudečková. S012: Changes in surveillance of acute respiratory infections including influenza in the slovak republic during 1993-2008. *Central European Journal of Public Health*, 19(1):20–25, 2011. ISSN 12107778. doi: 10.21101/CEJPH.A3593.

[183] Emily A. Lees, Enitan D. Carrol, Christine Gerrard, Fiona Hardiman, Gareth Howel, Alison Timmis, Kent Thorburn, Malcolm Guiver, and Paul S. McNamara. S013: Characterisation of acute respiratory infections at a United Kingdom paediatric teaching hospital: observational study assessing the impact of influenza A (2009 pdmH1N1) on predominant viral pathogens. *BMC infectious diseases*, 14(1), 6 2014. ISSN 1471-2334. doi: 10.1186/1471-2334-14-343. URL https://pubmed.ncbi.nlm.nih.gov/24948099/.

[184] Gerardo Chowell, Santiago Echevarría-Zuno, Cécile Viboud, Lone Simonsen, James Tamerius, Mark A. Miller, and Víctor H. Borja-AburtoVí. S014: Characterizing the Epidemiology of the 2009 Influenza A/H1N1 Pandemic in Mexico. *PLoS Medicine*, 8(5):e1000436, 5 2011. ISSN 15491277. doi: 10.1371/JOURNAL.PMED.1000436. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3101203/.

[185] David Luque-Paz, Pierre Tattevin, Paul Loubet, François Bénézit, Vincent Thibault, Fabrice Lainé, Philippe Vanhems, Selilah Amour, Bruno Lina, Xavier Duval, Anne Sophie L'Honneur, Nadhira Fidouh, Christine Vallejo, Sophie Alain, Florence Galtier, Vincent Foulongne, Gisèle Lagathu, Nezha Lenzi, Zineb Lesieur, Odile Launay, Stéphane Jouneau, O. Launay, N. Lenzi, Z. Lesieur, P. Loulergue, S. Momcilovic, J. P. Mira, N. Marin, J. Charpentier, A. Regent, R. Kanaan, F. Dumas, B. Doumenc, A. S. L'Honneur, M. Lachatre, T. Szwebel, J. Kansao, Y. Costa, X. Duval, J. F. Alexandra, H. Becheur, K. Belghalem, J. Bernard, A. Bleibtreu, M. Boisseau, R. Bories, O. Brugiere, F. Brunet, C. Burdet, E. Casalino, M. Caseris, C. Chansiaux, M. Chauchard, P. Chavance, C. Choquet, A. Cloppet-Fontaine, L. Colosi, B. Couset, B. Crestani, F. Crocket, A. Debit, Delanoe, V. Descamps,

P. Dieude, A. Dossier, N. Douron, E. Dupeyrat, N. Emeyrat, C. Fernet, T. Goulenok, S. Harent, R. Jouenne, A. Justet, M. Lachatre, A. Leleu, I. Lerat, M. Lilamand, H. Mal, A. Marceau, A. C. Metivier, K. Oplelatora, T. Papo, A. L. Pelletier, L. Pereira, P. Pradere, Prommier, P. Ralainnazava, M. Ranaivoision, A. Raynaud-Simon, C. Rioux, K. Sacre, V. Verry, V. Vuong, Y. Yazdapanah, N. Houhou, F. Galtier, P. Géraud, V. Driss, V. Maugueret, L. Crantelle, C. Agostini, M. Ray, F. Letois, T. Mura, C. Serrand, C. Agostini, S. Noslier, A. Giordano, H. Chevassus, E. Nyiramigisha, C. Merle, A. Bourdin, A. Konaté, X. Capdevilla, G. Du Cailar, A. Terminet, H. Blain, M. S. Leglise, A. Le Quellec, P. Corne, L. Landreau, K. Klouche, A. Bourgeois, M. Sebbane, G. Mourad, H. Leray, V. Foulongne, D. Postil, S. Alcolea, E. Couve-Deacon, S. Rogez, S. Amour, P. Vanhems, L. Argaud, M. Cour, R. Hernu, M. Simon, T. Baudry, K. Tazarourte, C. Bui-Xuan, J. Fattoum, B. Lina, M. Valette, F. Lainé, V. Thibault, S. Rochas, S. Cochennec, E. Thébault, G. Lagathu, S. Jouneau, M. Revest, F. Bénézit, M. Sébillotte, A. Le Bot, M. Baldeyrou, S. Patrat-Delon, M. Cailleaux, C. Pronier, and P. Tattevin. S015: Chronic use of inhaled corticosteroids in patients admitted for respiratory virus infections: a 6-year prospective multicenter study. *Scientific Reports*, 12(1): 4199, 12 2022. ISSN 20452322. doi: 10.1038/S41598-022-08089-0. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8913614/.

[186] Preeti Singh, Karanvir Attri, Deonath Mahto, Virendra Kumar, Dipti Kapoor, Anju Seth, Varinder Singh, Harish Pemde, Praveen Kumar, Ravitanaya Sodani, and Ankita Goel. S016: Clinical Profile of COVID-19 Illness in Children—Experience from a Tertiary Care Hospital. *Indian Journal of Pediatrics*, 89(1):45, 1 2021. ISSN 09737693. doi: 10.1007/S12098-021-03822-5. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8313877/.

[187] Felipe Cotrim De Carvalho, Erica Tatiane Da Silva, Walquiria Aparecida Ferreira De Almeida, Matheus Almeida Maroneze, Jaqueline de Araujo Schwartz, João Pedro Vieira Jardim, and Henry Maia Peixoto. S017: Clinical and epidemiological aspects of severe acute respiratory infection: before and during the first year of

the COVID-19 pandemic in Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 117(3):161–173, 3 2023. ISSN 1878-3503. doi: 10.1093/TRSTMH/TRAC074. URL https://pubmed.ncbi.nlm.nih.gov/35929810/.

[188] Grégory Quéromès, Emilie Frobert, Elena Burtseva, Anca Drăgănescu, Paravaiz A. Koul, Andrey Komissarov, V. Alberto Laguna-Torres, Jason Leblanc, F. Xavier López-Labrador, Snežana Medić, Alla Mironenko, Nancy A. Otieno, Guillermo M. Ruiz-Palacios, Tanriover MD, NGS team - Lyon, GIHSN collaborators, Laurence Josset, and Bruno Lina. S018: Clinical and phylogenetic influenza dynamics for the 2019-20 season in the global influenza hospital surveillance network (GIHSN) - Pilot study. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, 152, 7 2022. ISSN 1873-5967. doi: 10.1016/J.JCV.2022.105184. URL https://pubmed.ncbi.nlm.nih.gov/35594785/.

[189] Youri Yordanov, Aurélien Dinh, Alexandre Bleibtreu, Arthur Mensch, François Xavier Lescure, Erwan Debuc, Patrick Jourdain, Luc Jaulmes, and Agnes Dechartres. S019: Clinical characteristics and factors associated with hospital admission or death in 43 103 adult outpatients with coronavirus disease 2019 managed with the Covidom telesurveillance solution: a prospective cohort study. *Clinical Microbiology and Infection*, 27(8):1158, 8 2021. ISSN 14690691. doi: 10.1016/J.CMI.2021.04.010. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8076762/.

[190] Yang Huai, Xuhua Guan, Shali Liu, Timothy M. Uyeki, Hui Jiang, John Klena, Jigui Huang, Maoyi Chen, Youxing Peng, Hui Yang, Jun Luo, Jiandong Zheng, Zhibin Peng, Xixiang Huo, Lin Xiao, Hui Chen, Yuzhi Zhang, Xuesen Xing, Luzhao Feng, Dale J. Hu, Hongjie Yu, Faxian Zhan, and Jay K. Varma. S020: Clinical characteristics and factors associated with severe acute respiratory infection and influenza among children in Jingzhou, China. *Influenza and other respiratory*

viruses, 11(2):148–156, 3 2017. ISSN 1750-2659. doi: 10.1111/IRV.12419. URL https://pubmed.ncbi.nlm.nih.gov/27465959/.

[191] P. Loubet, N. Lenzi, M. Valette, V. Foulongne, A. Krivine, N. Houhou, G. Lagathu, S. Rogez, S. Alain, X. Duval, F. Galtier, D. Postil, P. Tattevin, P. Vanhems, F. Carrat, B. Lina, O. Launay, K. Seddik, Z. Lesieur, I. Bonmarin, P. Loulergue, H. Bodilis, M. Servera-Miyalou, I. Sadler, S. Momcilovic, R. Kanaan, N. Coolent, K. Tan Boun, P. Blanche, J. Charpentier, F. Daviaud, N. Mongardon, A. Bretagnol, Y. E. Claessens, F. Rozenberg, Y. Yazdanpanah, C. Burdet, S. Harent, M. Lachatre, C. Rioux, A. Bleibtreu, E. Casalino, C. Choquet, A. Leleu, K. Belghalem, L. Colosi, M. Ranaivoson, V. Verry, L. Pereira, E. Dupeyrat, J. Bernard, N. Emeyrat, P. Chavance, A. Debit, M. Aubier, P. Pradere, A. Justet, H. Mal, O. Brugiere, T. Papo, T. Goulenok, M. Boisseau, R. Jouenne, J. F. Alexandra, A. Raynaud-Simon, M. Lilamand, A. Cloppet-Fontaine, K. Becheur, A. L. Pelletier, N. Fidouh, P. Ralaimazava, F. Beaumale, Y. Costa, E. Munier, F. Betend, S. Amour, S. Loeffert, K. Francourt, C. Merle, F. Letois, P. Géraud, V. Driss, S. Noslier, M. Ray, M. Sebbane, A. Konaté, A. Bourdin, K. Klouche, M. S. Léglise, E. Couve-Deacon, D. Fruit, C. Fenerol, C. Vallejo, S. Jouneau, F. Lainé, E. Thébault, P. Fillatre, C. Le Pape, L. Beuzit, F. Chau, F. Carrat, F. Chau, and I. Goderel. S021: Clinical characteristics and outcome of respiratory syncytial virus infection among adults hospitalized with influenza-like illness in France. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 23(4):253–259, 4 2017. ISSN 1469-0691. doi: 10.1016/J.CMI.2016.11.014. URL https://pubmed.ncbi.nlm.nih.gov/27903461/.

[192] Arturo Galindo-Fraga, Ana A. Ortiz-Hernández, Alejandra Ramírez-Venegas, Rafael Valdez Vázquez, Sarbelio Moreno-Espinosa, Beatriz Llamosas-Gallardo, Santiago Pérez-Patrigeon, Maggie Salinger, Laura Freimanis, Chiung yu Huang, Wenjuan Gu, M. Lourdes Guerrero, John Beigel, and Guillermo M. Ruiz-Palacios. S022: Clinical characteristics and outcomes of influenza and other influenza-like illnesses in Mexico City. *International journal of infectious diseases :*

IJID : official publication of the International Society for Infectious Diseases, 17(7), 7 2013. ISSN 1878-3511. doi: 10.1016/J.IJID.2013.01.006. URL https://pubmed.ncbi.nlm.nih.gov/23416208/.

[193] Ashraf Hatem, Sherif Mohamed, Usama E. Abu Elhassan, Eman A.M. Ismael, Magda S. Rizk, Amany El-Kholy, and Mohamed El-Harras. S023: Clinical characteristics and outcomes of patients with severe acute respiratory infections (SARI): results from the Egyptian surveillance study 2010–2014. *Multidisciplinary Respiratory Medicine*, 14(1):11, 4 2019. ISSN 20496958. doi: 10.1186/S40248-019-0174-7. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6442424/.

[194] Ankur Sharma, Nikhil Kothari, Akhil Dhanesh Goel, Balakrishnan Narayanan, Shilpa Goyal, Pradeep Bhatia, Deepak Kumar, Gopal Krishna Bohra, Nishant Kumar Chauhan, Ramniwas Jalandra, Naveen Dutt, Pankaj Bhardwaj, Mahendra Kumar Garg, and Sanjeev Misra. S024: Clinical features and mortality in COVID-19 SARI versus non COVID-19 SARI cases from Western Rajasthan, India. *Journal of Family Medicine and Primary Care*, 10(9):3240, 9 2021. ISSN 2249-4863. doi: 10.4103/JFMPC.JFMPC{_}14{_}21. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8565113/.

[195] Iván de Jesús Ascencio-Montiel, Oscar David Ovalle-Luna, Ramón Alberto Rascón-Pacheco, Victor Hugo Borja-Aburto, and Gerardo Chowell. S025: Comparative epidemiology of five waves of COVID-19 in Mexico, March 2020–August 2022. *BMC Infectious Diseases*, 22(1):1–11, 12 2022. ISSN 14712334. doi: 10.1186/S12879-022-07800-W/TABLES/3. URL https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-022-07800-w.

[196] Emily Rowlinson, Lisa Peters, Adel Mansour, Hoda Mansour, Nahed Azazzy, Mayar Said, Sahar Samy, Eman Abbas, Hanaa Abu Elsood, Manal Fahim, Alaa Eid, Erik Reaves, Chris Van Beneden, Sarah Hamid, Sonja Olsen, Julia Fitzner, and Erica Dueger. S026: Comparison of common acute respiratory infection case definitions for identification of hospitalized influenza cases at a population-based surveillance

site in Egypt. *PloS one*, 16(3), 3 2021. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0248563. URL https://pubmed.ncbi.nlm.nih.gov/33765010/.

[197] Zaid Haddadin, James Chappell, Rendie McHenry, Claudia Guevara Pulido, Herdi Rahman, Wenying Gu, Danielle A. Rankin, Rana Talj, Leigh M. Howard, John V. Williams, Samir Faouri, Asem Shehabi, Najwa Khuri-Bulos, and Natasha B. Halasa. S027: Coronavirus Surveillance in a Pediatric Population in Jordan From 2010 to 2013: A Prospective Viral Surveillance Study. *The Pediatric infectious disease journal*, 40(1):E12–E17, 1 2021. ISSN 1532-0987. doi: 10.1097/INF.0000000000002965. URL https://pubmed.ncbi.nlm.nih.gov/33165274/.

[198] P Efstathiou, M Tseroni, and A Baka. S028: Deaths and Hospitalizations Related to 2009 Pandemic Influenza A (H1N1) — Greece, May 2009–February 2010. URL https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5922a2.htm.

[199] Tomoe Shimada, Tomimasa Sunagawa, Kiyosu Taniguchi, Yuichiro Yahata, Hajime Kamiya, Kumi Ueno Yamamoto, Yoshinori Yasui, and Nobuhiko Okabe. S029: Description of hospitalized cases of influenza A(H1N1)pdm09 infection on the basis of the national hospitalized-case surveillance, 2009-2010, Japan. *Japanese journal of infectious diseases*, 68(2):151–158, 11 2015. ISSN 1884-2836. doi: 10.7883/YOKEN.JJID.2014.125. URL https://pubmed.ncbi.nlm.nih.gov/25672359/.

[200] Víctor H. Borja-Aburto, Gerardo Chowell, Cécile Viboud, Lone Simonsen, Mark A. Miller, Concepción Grajales-Muñiz, Cesar R. González-Bonilla, Jose A. Diaz-Quiñonez, and Santiago Echevarría-Zuno. S039: Epidemiological characterization of a fourth wave of pandemic A/H1N1 influenza in Mexico, winter 2011-2012: age shift and severity. *Archives of medical research*, 43(7):563–570, 10 2012. ISSN 1873-5487. doi: 10.1016/J.ARCMED.2012.09.005. URL https://pubmed.ncbi.nlm.nih.gov/23079035/.

[201] Beth K. Thielen, Hannah Friedlander, Sarah Bistodeau, Bo Shu, Brian Lynch, Karen Martin, Erica Bye, Kathryn Como-Sabetti, David Boxrud, Anna K. Strain,

Sandra S. Chaves, Andrea Steffens, Ashley L. Fowlkes, Stephen Lindstrom, and Ruth Lynfield. S031: Detection of Influenza C Viruses Among Outpatients and Patients Hospitalized for Severe Acute Respiratory Infection, Minnesota, 2013–2016. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 66(7):1092, 3 2017. ISSN 15376591. doi: 10.1093/CID/CIX931. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5862734/.

[202] Lan Wei, Wei Liu, Xiao Ai Zhang, En Mei Liu, Yin Wo, Benjamin J. Cowling, and Wu Chun Cao. S032: Detection of Viral and Bacterial Pathogens in Hospitalized Children With Acute Respiratory Illnesses, Chongqing, 2009–2013. *Medicine*, 94 (16):e742, 4 2015. ISSN 15365964. doi: 10.1097/MD.0000000000000742. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4602679/.

[203] John Paul Cauchi, Maria Louise Borg, Aušra Džiugytė, Jessica Attard, Tanya Melillo, Graziella Zahra, Christopher Barbara, Michael Spiteri, Allan Drago, Luke Zammit, Joseph Debono, Jorgen Souness, Steve Agius, Sharon Young, Alan Dimech, Ian Chetcuti, Mark Camenzuli, Ivan Borg, Neville Calleja, Lorraine Tabone, Charmaine Gauci, Pauline Vassallo, and Joaquin Baruch. S033: Digitalizing and Upgrading Severe Acute Respiratory Infections Surveillance in Malta: System Development. *JMIR public health and surveillance*, 8(12), 12 2022. ISSN 2369-2960. doi: 10.2196/37669. URL https://pubmed.ncbi.nlm.nih.gov/36227157/.

[204] Marika K. Iwane, Sandra S. Chaves, Peter G. Szilagyi, Kathryn M. Edwards, Caroline B. Hall, Mary A. Staat, Cedric J. Brown, Marie R. Griffin, Geoffrey A. Weinberg, Katherine A. Poehling, Mila M. Prill, John V. Williams, and Carolyn B. Bridges. S034: Disparities Between Black and White Children in Hospitalizations Associated With Acute Respiratory Illness and Laboratory-confirmed Influenza and Respiratory Syncytial Virus in 3 US Counties—2002–2009. *American Journal of Epidemiology*, 177(7):656–665, 4 2013. ISSN 0002-9262. doi: 10.1093/AJE/KWS299. URL https://dx.doi.org/10.1093/aje/kws299.

[205] Ellen B. Fragaszy, Charlotte Warren-Gash, Peter J. White, Maria Zambon, William J. Edmunds, Jonathan S. Nguyen-Van-Tam, and Andrew C. Hayward. S035: Effects of seasonal and pandemic influenza on health-related quality of life, work and school absence in England: Results from the Flu Watch cohort study. *Influenza and Other Respiratory Viruses*, 12(1):171–182, 1 2018. ISSN 1750-2659. doi: 10.1111/IRV.12506. URL https://onlinelibrary.wiley.com/doi/full/10.1111/irv.12506https://onlinelibrary.wiley.com/doi/abs/10.1111/irv.12506https://onlinelibrary.wiley.com/doi/10.1111/irv.12506.

[206] Matthew J. Cummings, Barnabas Bakamutumaho, John Kayiwa, Timothy Byaruhanga, Nicholas Owor, Barbara Namagambo, Allison Wolf, Joseph F. Wamala, Stephen S. Morse, Julius J. Lutwama, and Max R. O'Donnell. S036: Epidemiologic and Spatiotemporal Characterization of Influenza and Severe Acute Respiratory Infection in Uganda, 2010-2015. *Annals of the American Thoracic Society*, 13(12):2159–2168, 12 2016. ISSN 2325-6621. doi: 10.1513/ANNALSATS.201607-561OC. URL https://pubmed.ncbi.nlm.nih.gov/27612095/.

[207] Nathália Mariana Santos Sansone, Matheus Negri Boschiero, and Fernando Augusto Lima Marson. S037: Epidemiologic Profile of Severe Acute Respiratory Infection in Brazil During the COVID-19 Pandemic: An Epidemiological Study. *Frontiers in microbiology*, 13, 7 2022. ISSN 1664-302X. doi: 10.3389/FMICB.2022.911036. URL https://pubmed.ncbi.nlm.nih.gov/35854935/.

[208] Gerardo Chowell, Santiago Echevarría-Zuno, Cécile Viboud, Lone Simonsen, Mark A. Miller, Irma Fernández-Gárate, Cesar González-Bonilla, and Víctor H. Borja-Aburto. S038: Epidemiological characteristics and underlying risk factors for mortality during the autumn 2009 pandemic wave in Mexico. *PloS one*, 7 (7), 7 2012. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0041069. URL https://pubmed.ncbi.nlm.nih.gov/22815917/.

[209] Zhiguo Liu, Liping Gao, Chuizhao Xue, Chunchun Zhao, Tiezhu Liu, Alie Tia,

Lili Wang, Junling Sun, Zhenjun Li, and Doris Harding. S040: Epidemiological Trends of Coronavirus Disease 2019 in Sierra Leone From March 2020 to October 2021. *Frontiers in public health*, 10, 6 2022. ISSN 2296-2565. doi: 10.3389/FPUBH.2022.949425. URL https://pubmed.ncbi.nlm.nih.gov/35844842/.

[210] B. Dwibedi, J. Sabat, S. Dixit, S. Rathore, S. Subhadra, S. Panda, S. S. Pati, M. Mandal, L. M. Ho, B. Thakur, and S. K. Kar. S041: Epidemiological and clinical profile of Influenza A(H1N1) pdm09 in Odisha, eastern India. *Heliyon*, 5 (10):e02639, 10 2019. ISSN 2405-8440. doi: 10.1016/J.HELIYON.2019.E02639.

[211] Manal Fahim, Basma AbdElGawad, Hossam Hassan, Amel Naguib, El Sabbah Ahmed, Salma Afifi, Hanaa Abu ElSood, and Amira Mohsen. S042: Epidemiology and outcome of influenza-associated infections among hospitalized patients with acute respiratory infections, Egypt national surveillance system, 2016-2019. *Influenza and other respiratory viruses*, 15(5):589–598, 9 2021. ISSN 1750-2659. doi: 10.1111/IRV.12867. URL https://pubmed.ncbi.nlm.nih.gov/33960675/.

[212] Katherine L. Anders, Hoa L. Nguyen, Nguyet Minh Nguyen, Nguyen Thi Van Thuy, Nguyen Thi Hong Van, Nguyen Trong Hieu, Nguyen Thi Hong Tham, Phan Thi Thanh Ha, Le Bich Lien, Nguyen Van Vinh Chau, Vu Thi Ty Hang, H. Rogier Van Doorn, and Cameron P. Simmons. S043: Epidemiology and virology of acute respiratory infections during the first year of life: a birth cohort study in Vietnam. *The Pediatric infectious disease journal*, 34(4):361–370, 4 2015. ISSN 1532-0987. doi: 10.1097/INF.0000000000000643. URL https://pubmed.ncbi.nlm.nih.gov/25674708/.

[213] Hossein Faramarzi, Razieh Sadat Mousavi-Roknabadi, Abdolrasoul Hemmati, Ali Faramarzi, and Hamid Bakhtiari. S044: Epidemiology of Influenza in Fars Province, Southern Iran; a Population-Based Study (2015-2019). *Archives of Iranian medicine*, 24(3):199–208, 3 2021. ISSN 1735-3947. doi: 10.34172/AIM.2021.31. URL https://pubmed.ncbi.nlm.nih.gov/33878878/.

[214] Antonia Ho, Jane Mallewa, Ingrid Peterson, Miguel San Joaquin, Shikha Garg, Naor

Bar-Zeev, Mavis Menyere, Maaike Alaerts, Gugulethu Mapurisa, Moses Chilombe, Mulinda Nyirenda, David G. Lalloo, Camilla Rothe, Marc Alain Widdowson, Meredith McMorrow, Neil French, Dean Everett, and Robert S. Heyderman. S045: Epidemiology of Severe Acute Respiratory Illness and Risk Factors for Influenza Infection and Clinical Severity among Adults in Malawi, 2011–2013. *The American Journal of Tropical Medicine and Hygiene*, 99(3):772, 2018. ISSN 00029637. doi: 10.4269/AJTMH.17-0905. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6169174/.

[215] Joan Puig-Barberà, Angels Natividad-Sancho, Svetlana Trushakova, Anna Sominina, Maria Pisareva, Meral A. Ciblak, Selim Badur, Hongjie Yu, Benjamin J. Cowling, Clotilde El Guerche-Séblain, Ainara Mira-Iglesias, Lidiya Kisteneva, Kirill Stolyarov, Kubra Yurtcu, Luzhao Feng, Xavier López-Labrador, Elena Burtseva, V. Afanasieva, F. Aktaş, S. Borekci, A. Buigues-Vila, Z. Buzitskaya, J. Cai, B. Çakir, M. Carballido-Fernández, C. Carratalá-Munuera, C. Chai, E. Chen, S. Çelebi, Y. Cui, D. B. Deniz, H. Dong, X. Dong, M. Durusu, A. Fadeev, S. Feng, E. Garina, S. Gencer, V. Gil-Guillén, M. Hacimustafaoğlu, S. Hancerli, L. Huang, D. K. Ip, L. Kolobukhina, K. Krasnoslobotsev, C. Li, R. Limón-Ramírez, C. Mahé, L. Merkulova, J. Mollar Maseres, E. Mukasheva, L. Ozisik, M. C. Otero-Reigada, S. Özer, Y. Qin, A. Eren-Şensoy, E. Smorodintseva, V. Sukhovetskaya, G. Sun, Y. Tang, A. Tormos, F. X. López-Labrador, M. Tortajada-Girbés, R. Vartanyan, L. Voloshchuk, Q. Wang, D. Wen, P. Wu, P. Yang, B. Yi, S. Zhang, Y. Zhang, and J. Zheng. S046: Epidemiology of Hospital Admissions with Influenza during the 2013/2014 Northern Hemisphere Influenza Season: Results from the Global Influenza Hospital Surveillance Network. *PloS one*, 11(5), 5 2016. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0154970. URL https://pubmed.ncbi.nlm.nih.gov/27196667/.

[216] Marina Beretta Duarte, Tatiana Schäffer Gregianini, Letícia G. Martins, and Ana Beatriz G. Veiga. S047: Epidemiology of influenza B infection in the state of Rio Grande do Sul, Brazil, from 2003 to 2019. *Journal of medical virology*, 93

(8):4756–4762, 8 2021. ISSN 1096-9071. doi: 10.1002/JMV.26822. URL https://pubmed.ncbi.nlm.nih.gov/33501655/.

[217] Cheryl Cohen, Sibongile Walaza, Jocelyn Moyes, Michelle Groome, Stefano Tempia, Marthi Pretorius, Orienka Hellferscee, Halima Dawood, Summaya Haffejee, Ebrahim Variava, Kathleen Kahn, Akhona Tshangela, Anne Von Gottberg, Nicole Wolter, Adam L. Cohen, Babatyi Kgokong, Marietjie Venter, and Shabir A. Madhi. S048: Epidemiology of severe acute respiratory illness (SARI) among adults and children aged 5 years in a high HIV-prevalence setting, 2009-2012. *PloS one*, 10 (2), 2 2015. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0117716. URL https://pubmed.ncbi.nlm.nih.gov/25706880/.

[218] Manal Fahim, Wael H. Roshdy, Ola Deghedy, Reham Kamel, Amel Naguib, Shymaa Showky, Nancy Elguindy, Mohammad Abdel Fattah, Salma Afifi, Amira Mohsen, Amr Kandeel, and Khaled Abdelghaffar. S049: Epidemiology, Disease Severity and Outcome of Severe Acute Respiratory Syndrome Coronavirus 2 and Influenza Viruses Coinfection Seen at Egypt Integrated Acute Respiratory Infections Surveillance, 2020-2022. *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale*, 2022, 2022. ISSN 1712-9532. doi: 10.1155/2022/7497500. URL https://pubmed.ncbi.nlm.nih.gov/36437892/.

[219] Pritimoy Das, Zubair Akhtar, Syeda Mah-E-Muneer, Md Ariful Islam, Mohammed Ziaur Rahman, Mustafizur Rahman, Mahmudur Rahman, Mahbubur Rahman, Mallick Masum Billah, A. S.M. Alamgir, Meerjady Sabrina Flora, Tahmina Shirin, Sayera Banu, and Fahmida Chowdhury. S050: Establishing a sentinel surveillance system for the novel COVID-19 in a resource-limited country: methods, system attributes and early findings. *BMJ open*, 11(12), 12 2021. ISSN 2044-6055. doi: 10.1136/BMJOPEN-2021-055169. URL https://pubmed.ncbi.nlm.nih.gov/34857579/.

[220] S. Buda, K. Tolksdorf, E. Schuler, R. Kuhlen, and W. Haas. S051: Establishing

an ICD-10 code based SARI-surveillance in Germany - description of the system and first results from five recent influenza seasons. *BMC public health*, 17(1), 6 2017. ISSN 1471-2458. doi: 10.1186/S12889-017-4515-1. URL https://pubmed.ncbi.nlm.nih.gov/28666433/.

[221] Kaitlyn Vette, Christina Bareja, Robert Clark, and Aparna Lal. S052: Establishing thresholds and parameters for pandemic influenza severity assessment, Australia. *Bulletin of the World Health Organization*, 96(8):558, 2018. ISSN 15640604. doi: 10.2471/BLT.18.211508. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6083389/.

[222] Astride Jules, Carlos G. Grijalva, Yuwei Zhu, Keipp H. Talbot, John V. Williams, William D. Dupont, Kathryn M. Edwards, William Schaffner, David K. Shay, and Marie R. Griffin. S053: Estimating age-specific influenza-related hospitalization rates during the pandemic (H1N1) 2009 in Davidson Co, TN. *Influenza and other respiratory viruses*, 6(3), 5 2012. ISSN 1750-2659. doi: 10.1111/J.1750-2659.2012.00343.X. URL https://pubmed.ncbi.nlm.nih.gov/22360812/.

[223] Liselotte Van Asten, Angie Luna Pinzon, Dylan W. De Lange, Evert De Jonge, Frederika Dijkstra, Sierk Marbus, Gé A. Donker, Wim Van Der Hoek, and Nicolette F. De Keizer. S054: Estimating severity of influenza epidemics from severe acute respiratory infections (SARI) in intensive care units. *Critical care (London, England)*, 22(1), 12 2018. ISSN 1466-609X. doi: 10.1186/S13054-018-2274-8. URL https://pubmed.ncbi.nlm.nih.gov/30567568/.

[224] Doaa M. Abdel-Hady, Rima M. Al Balushi, Badr A. Al Abri, Seif S. Al Abri, Hanan S. Al Kindi, Amina K. Al-Jardani, Fatma M. Al Yaqubi, and Idris S. Al Abaidani. S055: Estimating the burden of influenza-associated hospitalization and deaths in Oman (2012-2015). *Influenza and other respiratory viruses*, 12(1):146–152, 1 2018. ISSN 1750-2659. doi: 10.1111/IRV.12500. URL https://pubmed.ncbi.nlm.nih.gov/29205882/.

[225] Paola del Carmen Guerra-de Blas, Ana M. Ortega-Villa, Ana A. Ortiz-Hernández, Alejandra Ramírez-Venegas, Sarbelio Moreno-Espinosa, Beatriz Llamosas-Gallardo, Santiago Pérez-Patrigeon, Sally Hunsberger, Martín Magaña, Rafael Valdez-Vázquez, Laura Freimanis, Juan Francisco Galán-Herrera, M. Lourdes Guerrero-Almeida, John H. Powers, Guillermo M. Ruiz-Palacios, John Beigel, and Arturo Galindo-Fraga. S056: Etiology, clinical characteristics, and risk factors associated with severe influenza-like illnesses in Mexican adults. *IJID regions*, 6: 152–158, 3 2023. ISSN 2772-7076. doi: 10.1016/J.IJREGI.2023.01.012. URL https://pubmed.ncbi.nlm.nih.gov/36865993/.

[226] Suchitra Rao, Emad Yanni, Angela Moss, Molly M. Lamb, Anne Schuind, Rafik Bekkat-Berkani, Bruce L. Innis, Jillian Cotter, Rakesh D. Mistry, and Edwin J. Asturias. S057: Evaluation of a New Clinical Endpoint for Moderate to Severe Influenza Disease in Children: A Prospective Cohort Study. *Journal of the Pediatric Infectious Diseases Society*, 9(4):460, 2019. ISSN 20487207. doi: 10.1093/JPIDS/PIZ075. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7495912/.

[227] Gideon Loevinsohn, Mutinta Hamahuwa, Pamela Sinywimaanzi, Katherine Z.J. Fenstermacher, Kathryn Shaw-Saliba, Andrew Pekosz, Mwaka Monze, Richard E. Rothman, Edgar Simulundu, Philip E. Thuma, and Catherine G. Sutcliffe. S058: Facility-based surveillance for influenza and respiratory syncytial virus in rural Zambia. *BMC Infectious Diseases*, 21(1):1–15, 12 2021. ISSN 14712334. doi: 10.1186/S12879-021-06677-5/TABLES/4. URL https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-021-06677-5.

[228] Kamilla Lelis Rodrigues de Araujo, Érika Carvalho de Aquino, Lara Lívia Santos da Silva, and Yves Mauro Fernandes Ternes. S059: Factors associated with Severe Acute Respiratory Syndrome in a Brazilian central region. *Ciencia & saude coletiva*, 25(suppl 2):4121–4130, 10 2020. ISSN 1678-4561. doi: 10.1590/1413-812320202510.2.26802020. URL https://pubmed.ncbi.nlm.nih.gov/33027348/.

[229] Afaf Merza Mohamed, Adel Al Sayyad, Ebrahim Matar, Hasan M. Isa, Wafa Fawzi

Hasan, Nawra Sayed Jalal Yusuf Hashim, Bayan Abduljalil Alajaimi, and Qatrmeer Aldolabi. S060: Factors associated with poor outcomes in patients with severe acute respiratory infections in Bahrain. *Influenza and other respiratory viruses*, 17(4), 4 2023. ISSN 1750-2659. doi: 10.1111/IRV.13133. URL https://pubmed.ncbi.nlm.nih.gov/37123813/.

[230] Melissa S. Stockwell, Carrie Reed, Celibell Y. Vargas, Liqun Wang, Luis R. Alba, Haomiao Jia, Philip Larussa, Elaine L. Larson, and Lisa Saiman. S062: Five-Year Community Surveillance Study for Acute Respiratory Infections Using Text Messaging: Findings From the MoSAIC Study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 75(6):987–995, 9 2022. ISSN 1537-6591. doi: 10.1093/CID/CIAC027. URL https://pubmed.ncbi.nlm.nih.gov/35037056/.

[231] Osman Abdalla, Mutaz Mohammed, Ahmed Mohammed Hakawi, Alanoud Aljifri, Mohamed Abdalla, Sara Eltigani, Sahibzada Azhar Mujib, and Abdullah Assiri. S063: Hospital-based surveillance of influenza A(H1N1)pdm09 virus in Saudi Arabia, 2010-2016. *Annals of Saudi medicine*, 40(1):1–6, 2 2020. ISSN 0975-4466. doi: 10.5144/0256-4947.2020.1. URL https://pubmed.ncbi.nlm.nih.gov/32026719/.

[232] Camille Pelat, Andrea Lasserre, Ana Xavier, Clément Turbelin, Thierry Blanchon, and Thomas Hanslik. S064: Hospitalization of influenza-like illness patients recommended by general practitioners in France between 1997 and 2010. *Influenza and other Respiratory Viruses*, 7(1):74–84, 1 2013. ISSN 17502640. doi: 10.1111/J.1750-2659.2012.00356.X.

[233] Z. Sakkou, F. Stripeli, N. G. Papadopoulos, E. Critselis, V. Georgiou, M. Mavrikou, P. Drossatou, A. Constantopoulos, D. Kafetzis, and M. Tsolia. S066: Impact of influenza infection on children's hospital admissions during two seasons in Athens, Greece. *Vaccine*, 29(6):1167–1172, 2 2011. ISSN 1873-2518. doi: 10.1016/J.VACCINE.2010.12.014. URL https://pubmed.ncbi.nlm.nih.gov/21172380/.

[234] Ru Ning Guo, Hui Zhen Zheng, Chun Quan Ou, Li Qun Huang, Yong Zhou, Xin Zhang, Can Kun Liang, Jin Yan Lin, Hao Jie Zhong, Tie Song, and Hui Ming Luo. S067: Impact of Influenza on Outpatient Visits, Hospitalizations, and Deaths by Using a Time Series Poisson Generalized Additive Model. *PLOS ONE*, 11(2): e0149468, 2 2016. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0149468. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0149468.

[235] Rachael Pung and Vernon Jian Ming Lee. S068: Implementing the World Health Organization Pandemic Influenza Severity Assessment framework-Singapore's experience. *Influenza and other respiratory viruses*, 14(1):3–10, 1 2020. ISSN 1750-2659. doi: 10.1111/IRV.12680. URL https://pubmed.ncbi.nlm.nih.gov/31622034/.

[236] Patrick E. Obermeier, Lea D. Seeber, Maren Alchikh, Brunhilde Schweiger, and Barbara A. Rath. S069: Incidence, Disease Severity, and Follow-Up of Influenza A/A, A/B, and B/B Virus Dual Infections in Children: A Hospital-Based Digital Surveillance Program. *Viruses*, 14(3), 3 2022. ISSN 1999-4915. doi: 10.3390/V14030603. URL https://pubmed.ncbi.nlm.nih.gov/35337010/.

[237] Keisuke Yoshihara, Minh Nhat Le, Michiko Toizumi, Hien Anh Nguyen, Hien Minh Vo, Takato Odagiri, Seiichiro Fujisaki, Koya Ariyoshi, Hiroyuki Moriuchi, Masahiro Hashizume, Duc Anh Dang, and Lay Myint Yoshida. S070: Influenza B associated paediatric acute respiratory infection hospitalization in central vietnam. *Influenza and other respiratory viruses*, 13(3):248–261, 5 2019. ISSN 1750-2659. doi: 10.1111/IRV.12626. URL https://pubmed.ncbi.nlm.nih.gov/30575288/.

[238] Mohammad Al-Abdallat, Patrick Dawson, Aktham Jeries Haddadin, Waleed El-Shoubary, Erica Dueger, Tarek Al-Sanouri, Mayar M. Said, and Maha Talaat. S071: Influenza hospitalization epidemiology from a severe acute respiratory infection surveillance system in Jordan, January 2008-February 2014. *Influenza and other*

respiratory viruses, 10(2):91–97, 3 2016. ISSN 1750-2659. doi: 10.1111/IRV.12354. URL https://pubmed.ncbi.nlm.nih.gov/26505620/.

[239] Maria Luiza Moretti, Verônica Sinkoc, Luis Gustavo de Oliveira Cardoso, Gema Jesus de Camargo, Luis Felipe Bachur, Christian Cruz Hofling, Rodrigo Angerami, Plínio Trabasso, Márcia Teixeira Garcia, and Mariângela Ribeiro Resende. S072: Lessons from the epidemiological surveillance program, during the influenza A (H1N1) virus epidemic, in a reference university hospital of Southeastern Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, 44(4):405–411, 7 2011. ISSN 1678-9849. doi: 10.1590/S0037-86822011005000048. URL https://pubmed.ncbi.nlm.nih.gov/21789355/.

[240] Ana Lúcia Escobar, Tomás Daniel Menéndez Rodriguez, and Janne Cavalcante Monteiro. S073: Lethality and characteristics of deaths due to COVID-19 in Rondônia: an observational study. *Epidemiologia e servicos de saude : revista do Sistema Unico de Saude do Brasil*, 30(1), 2020. ISSN 2237-9622. doi: 10.1590/S1679-49742021000100019. URL https://pubmed.ncbi.nlm.nih.gov/33331602/.

[241] Amr Kandeel, Patrick Dawson, Manal Labib, Mayar Said, Samir El-Refai, Amani El-Gohari, and Maha Talaat. S074: Morbidity, Mortality, and Seasonality of Influenza Hospitalizations in Egypt, November 2007-November 2014. *PLoS ONE*, 11(9):e0161301, 9 2016. ISSN 19326203. doi: 10.1371/JOURNAL.PONE.0161301. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5015910/.

[242] Cheryl Cohen, Jocelyn Moyes, Stefano Tempia, Michelle Groome, Sibongile Walaza, Marthi Pretorius, Halima Dawood, Meera Chhagan, Summaya Haffejee, Ebrahim Variava, Kathleen Kahn, Anne Von Gottberg, Nicole Wolter, Adam L. Cohen, Babatyi Malope-Kgokong, Marietjie Venter, and Shabir A. Madhi. S075: Mortality amongst Patients with Influenza-Associated Severe Acute Respiratory Illness, South Africa, 2009-2013. *PLoS ONE*, 10(3):e0118884, 3

2015. ISSN 19326203. doi: 10.1371/JOURNAL.PONE.0118884. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4365037/.

[243] Eduardo Azziz-Baumgartner, Ana María Cabrera, Loretta Chang, Rogelio Calli, Gabriela Kusznierz, Clarisa Baez, Pablo Yedlin, Ana María Zamora, Romina Cuezzo, Elena Beatriz Sarrouf, Andrea Uboldi, Juan Herrmann, Elsa Zerbini, Osvaldo Uez, Pedro Osvaldo Rico Cordeiro, Pollyanna Chavez, George Han, Julián Antman, Fatima Coronado, Joseph Bresee, Marina Kosacoff, Marc Alain Widdowson, and Horacio Echenique. S076: Mortality, Severe Acute Respiratory Infection, and Influenza-Like Illness Associated with Influenza A(H1N1)pdm09 in Argentina, 2009. *PLOS ONE*, 7(10):e47540, 10 2012. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0047540. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0047540.

[244] Ting Chia Weng, Han Yi Robert Chiu, Shey Ying Chen, Fuh Yuan Shih, Chwan Chuen King, and Cheng Chung Fang. S077: National retrospective cohort study to identify age-specific fatality risks of comorbidities among hospitalised patients with influenza-like illness in Taiwan. *BMJ open*, 9(6), 6 2019. ISSN 2044-6055. doi: 10.1136/BMJOPEN-2018-025276. URL https://pubmed.ncbi.nlm.nih.gov/31239301/.

[245] François Bénézit, Paul Loubet, Florence Galtier, Charlotte Pronier, Nezha Lenzi, Zineb Lesieur, Stéphane Jouneau, Gisèle Lagathu, Anne Sophie L'Honneur, Vincent Foulongne, Christine Vallejo, Sophie Alain, Xavier Duval, Nawal Houhou, Yolande Costa, Philippe Vanhems, Sélilah Amour, Fabrice Carrat, Bruno Lina, Odile Launay, Pierre Tattevin, P. Loulergue, S. Momcilovic, J. P. Mira, N. Marin, A. Regent, Kanaan, F. Dumas, B. Doumenc, J. F. Alexandra, H. Becheur, K. Belghalem, J. Bernard, A. Bleitreu, M. Boisseau, R. Bories, O. Brugiere, F. Brunet, C. Burdet, E. Casalino, M. Caseris, Chansiaux, M. Chauchard, P. Chavance, C. Choquet, Cloppet-Fontaine, L. Colosi, B. Couset, B. Crestani, F. Crocket, A. Debit, Delanoe, V. Descamps, P. Dieude, A. Dossier, N. Douron, E. Dupeyrat,

N. Emeyrat, Fernet, T. Goulenok, S. Harent, Jouenne, A. Justet, M. Lachatre, A. Leleu, I. Lerat, M. Lilamand, H. Mal, A. Marceau, A. C. Metivier, K. Oplelatora, T. Papo, A. L. Pelletier, L. Pereira, P. Pradere, Prommier, P. Ralainnazava, M. Ranaivoision, A. Raynaud-Simon, C. Rioux, K. Sacre, V. Verry, V. Vuong, Y. Yazdapanah, P. Géraud, V. Driss, V. Maugueret, M. Ray, F. Letois, T. Mura, C. Merle, A. Bourdin, A. Konaté, X. Capdevilla, G. Du Cailar, A. Terminet, H. Blain, M. S. Leglise, A. Le Quellec, P. Corne, L. Landreau, K. Klouche, A. Bourgeois, M. Sebbane, G. Mourad, H. Leray, M. Maarouf, D. Postil, S. Alcolea, E. Couve Deacon, S. Rogez, L. Argaud, K. Tazarourte, R. Hernu, M. Cour, M. Simon, T. Baudry, L. Jacquin, F. Lainé, B. Laviolle, J. S. Allain, N. Belhomme, V. Thibault, S. Rochas, S. Cochennec, E. Ouamara-Digue, C. Lepape, M. Revest, S. Simon, J. Fouchard, C. Gautier, N. Nouredine, and E. Thébault. S078: Non-influenza respiratory viruses in adult patients admitted with influenza-like illness: a 3-year prospective multicenter study. *Infection*, 48(4):489–495, 8 2020. ISSN 1439-0973. doi: 10.1007/S15010-019-01388-1. URL https://pubmed.ncbi.nlm.nih.gov/32056143/.

[246] Yolanda Miroballi, J. Scott Baird, Sheemon Zackai, Jean Marie Cannon, Maria Messina, Thyyar Ravindranath, Robert Green, Phyllis Della-Latta, Stephen Jenkins, Bruce M. Greenwald, E. Yoko Furuya, Philip L. Graham, F. Meridith Sonnett, Shari Platt, Patricia DeLaMora, and Lisa Saiman. S079: Novel influenza A(H1N1) in a pediatric health care facility in New York City during the first wave of the 2009 pandemic. *Archives of pediatrics & adolescent medicine*, 164(1):24–30, 1 2010. ISSN 1538-3628. doi: 10.1001/ARCHPEDIATRICS.2009.259. URL https://pubmed.ncbi.nlm.nih.gov/20048238/.

[247] Juan Pablo Castillo-Palencia, Lucie Laflamme, and Joel Monárrez-Espino. S080: Occurrence of AH1N1 viral infection and clinical features in symptomatic patients who received medical care during the 2009 influenza pandemic in Central Mexico. *BMC Infectious Diseases*, 12, 12 2012. ISSN 14712334. doi: 10.1186/1471-2334-12-363/METRICS.

[248] Collective Chilean Task Force for study of Pandemic Influenza A (H1N1), E Pedroni, M García, V Espínola, A Guerrero, C González, A Olea, M Calvo, B Martorell, M Winkler, M V Carrasco, J A Vergara, J Ulloa, A M Carrazana, O Mujica, J E Villarroel, M Labraña, M Vargas, P González, L Cáceres, C G Zamorano, R Momberg, G Muñoz, J Rocco, V Bosque, A Gallardo, J Elgueta, and J Vega. S081: Outbreak of 2009 pandemic influenza A(H1N1), Los Lagos, Chile, April-June 2009. *Eurosurveillance*, 15(1), 1 2010. ISSN 1560-7917. doi: 10.2807/ESE.15.01.19456-EN.

[249] Vladimir Petrović, Zorica Šeguljev, Gorana Ćosić, Mioljub Ristić, Jasminka Nedeljković, Nataša Dragnić, and Snežana Ukropina. S084: Overview of the winter wave of 2009 pandemic influenza A(H1N1)v in Vojvodina, Serbia. *Croatian medical journal*, 52(2):141–150, 2011. ISSN 1332-8166. doi: 10.3325/CMJ.2011.52.141. URL https://pubmed.ncbi.nlm.nih.gov/21495196/.

[250] W. Oliveira, E. Carmo, G. Penna, R. Kuchenbecker, H. Santos, W. Araujo, R. Malaguti, B. Duncan, and M. Schmidt. S085: Pandemic H1N1 influenza in Brazil: analysis of the first 34,506 notified cases of influenza-like illness with severe acute respiratory infection (SARI). *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 14 (42), 2009. ISSN 1560-7917. doi: 10.2807/ESE.14.42.19362-EN. URL https://pubmed.ncbi.nlm.nih.gov/19883548/.

[251] Kelly Osezele Elimian, Chinwe Lucia Ochu, Blessing Ebhodaghe, Puja Myles, Emily E. Crawford, Ehimario Igumbor, Winifred Ukponu, Adobola Olayinka, Olusola Aruna, Chioma Dan-Nwafor, Olatayo Ayodeji Olawepo, Oladipo Ogunbode, Rhoda Atteh, William Nwachukwu, Sudhir Venkatesan, Chijioke Obagha, Samuel Ngishe, Kabir Suleiman, Muhammad Usman, Hakeem Abiola Yusuff, Ifeoma Nwadiuto, Abbas Aliyu Mohammed, Rabi Usman, Nwando Mba, Olaolu Aderinola, Elsie Ilori, John Oladejo, Ibrahim Abubakar, and Chikwe Ihekweazu. S086: Patient characteristics associated with COVID-19 positivity and fatality in Nigeria:

retrospective cohort study. *BMJ open*, 10(12), 12 2020. ISSN 2044-6055. doi: 10.1136/BMJOPEN-2020-044079. URL https://pubmed.ncbi.nlm.nih.gov/33334842/.

[252] Melissa K. Andrew, Sarah MacDonald, Judith Godin, Janet E. McElhaney, Jason LeBlanc, Todd F. Hatchette, William Bowie, Kevin Katz, Allison McGeer, Makeda Semret, and Shelly A. McNeil. S087: Persistent Functional Decline Following Hospitalization with Influenza or Acute Respiratory Illness. *Journal of the American Geriatrics Society*, 69(3):696–703, 3 2021. ISSN 1532-5415. doi: 10.1111/JGS.16950. URL https://pubmed.ncbi.nlm.nih.gov/33294986/.

[253] Ana B. Cazé, Thiago Cerqueira-Silva, Adriele P. Bomfim, Gisley L. de Souza, Amanda C.A. Azevedo, Michelle Q.A. Brasil, Nara R. Santos, Ricardo Khouri, Jennifer Dan, Antonio C. Bandeira, Luciano P.G. Cavalcanti, Manoel Barral-Netto, Aldina Barral, Cynara G. Barbosa, and Viviane S. Boaventura. S088: Prevalence and risk factors for long COVID after mild disease: A cohort study with a symptomatic control group. *Journal of Global Health*, 13:06015, 2023. ISSN 20472986. doi: 10.7189/JOGH.13.06015. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10173895/.

[254] Yuli V. Fuentes, Elsa D. Ibáñez-Prada, Cristian C. Serrano-Mayorga, Carlos G. Pfizenmaier, Marcela Cano, Natalia Boada, Paola Rincon, Esteban García-Gallo, Sara Duque, Andrés F. Ocampo, Alirio Bastidas, Sandra Gomez, Hernán Vargas, and Luis F. Reyes. S089: Prevalence, incidence, and severity associated with viral respiratory tract infections in Colombian adults before the COVID-19 pandemic. *Journal of infection and public health*, 15(12):1381–1387, 12 2022. ISSN 1876-035X. doi: 10.1016/J.JIPH.2022.10.015. URL https://pubmed.ncbi.nlm.nih.gov/36370485/.

[255] Gerardo Chowell, Santiago Echevarría-Zuno, Cecile Viboud, Lone Simonsen, Concepcion Grajales Muñiz, Ramón Alberto Rascón Pacheco, Margot González León, and Víctor Hugo Borja Aburto. S090: Recrudescent wave of pandemic A/H1N1

influenza in Mexico, winter 2011-2012: Age shift and severity. *PLoS currents*, 4: RRN1306, 3 2012. ISSN 2157-3999. doi: 10.1371/CURRENTS.RRN1306. URL https://pubmed.ncbi.nlm.nih.gov/22485199/.

[256] Ann R. Falsey, Janet E. McElhaney, Jiri Beran, Gerrit A. Van Essen, Xavier Duval, Meral Esen, Florence Galtier, Pierre Gervais, Shinn Jang Hwang, Peter Kremsner, Odile Launay, Geert Leroux-Roels, Shelly A. McNeil, Andrzej Nowakowski, Jan Hendrik Richardus, Guillermo Ruiz-Palacios, Suzanne St Rose, Jeanne Marie Devaster, Lidia Oostvogels, Serge Durviaux, and Sylvia Taylor. S091: Respiratory syncytial virus and other respiratory viral infections in older adults with moderate to severe influenza-like illness. *The Journal of infectious diseases*, 209(12): 1873–1881, 6 2014. ISSN 1537-6613. doi: 10.1093/INFDIS/JIT839. URL https://pubmed.ncbi.nlm.nih.gov/24482398/.

[257] Ingrid Peterson, Naor Bar-Zeev, Neil Kennedy, Antonia Ho, Laura Newberry, Miguel A. SanJoaquin, Mavis Menyere, Maaike Alaerts, Gugulethu Mapurisa, Moses Chilombe, Ivan Mambule, David G. Lalloo, Suzanne T. Anderson, Thembi Katangwe, Nigel Cunliffe, Nico Nagelkerke, Meredith McMorrow, Marc Allain Widdowson, Neil French, Dean Everett, and Robert S. Heyderman. S092: Respiratory Virus–Associated Severe Acute Respiratory Illness and Viral Clustering in Malawian Children in a Setting With a High Prevalence of HIV Infection, Malaria, and Malnutrition. *The Journal of Infectious Diseases*, 214(11): 1700, 12 2016. ISSN 15376613. doi: 10.1093/INFDIS/JIW426. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5341080/.

[258] Sibongile Walaza, Stefano Tempia, Anne von Gottberg, Nicole Wolter, Jinal N. Bhiman, Amelia Buys, Daniel Amoako, Fahima Moosa, Mignon du Plessis, Jocelyn Moyes, Meredith L. McMorrow, Halima Dawood, Ebrahim Variava, Gary Reubenson, Jeremy Nel, Heather J. Zar, Mvuyo Makhasi, Susan Meiring, Vanessa Quan, and Cheryl Cohen. S093: Risk Factors for Severe Coronavirus Disease 2019 Among Human Immunodeficiency Virus-Infected and -Uninfected Individuals in South

Africa, April 2020-March 2022: Data From Sentinel Surveillance. *Open forum infectious diseases*, 9(12), 12 2022. ISSN 2328-8957. doi: 10.1093/OFID/OFAC578. URL https://pubmed.ncbi.nlm.nih.gov/36570970/.

[259] Stefano Tempia, Sibongile Walaza, Jocelyn Moyes, Adam L. Cohen, Claire Von Mollendorf, Florette K. Treurnicht, Marietjie Venter, Marthi Pretorius, Orienka Hellferscee, Senzo Mtshali, Mpho Seleka, Akhona Tshangela, Athermon Nguweneza, Johanna M. McAnerney, Nicole Wolter, Anne Von Gottberg, Halima Dawood, Ebrahim Variava, Shabir A. Madhi, and Cheryl Cohen. S094: Risk Factors for Influenza-Associated Severe Acute Respiratory Illness Hospitalization in South Africa, 2012-2015. *Open forum infectious diseases*, 4(1), 1 2017. ISSN 2328-8957. doi: 10.1093/OFID/OFW262. URL https://pubmed.ncbi.nlm.nih.gov/28480255/.

[260] Amani A. El Kholy, Nadia A. Mostafa, Aliaa Adel Ali, Seham A. El-Sherbini, Reem I. Ismail, Rania I. Magdy, May S. Soliman, and Mayar M. Said. S095: Risk factors of prolonged hospital stay in children with viral severe acute respiratory infections. *Journal of infection in developing countries*, 8(10):1285–1293, 10 2014. ISSN 1972-2680. doi: 10.3855/JIDC.4682. URL https://pubmed.ncbi.nlm.nih.gov/25313605/.

[261] Zubair Akhtar, Md Ariful Islam, Mohammad Abdul Aleem, Syeda Mah-E-Muneer, M. Kaousar Ahmmed, Probir K. Ghosh, Mustafizur Rahman, Mohammed Ziaur Rahman, Mariya Kibtiya Sumiya, Md Mahfuzur Rahman, Tahmina Shirin, A. S.M. Alamgir, Sayera Banu, Mahmudur Rahman, and Fahmida Chowdhury. S096: SARS-CoV-2 and influenza virus coinfection among patients with severe acute respiratory infection during the first wave of COVID-19 pandemic in Bangladesh: a hospital-based descriptive study. *BMJ open*, 11(11), 11 2021. ISSN 2044-6055. doi: 10.1136/BMJOPEN-2021-053768. URL https://pubmed.ncbi.nlm.nih.gov/34845073/.

[262] S097: 2022 Acute respiratory illness surveillance report, . URL https://www.es

r.cri.nz/digital-library/2022-acute-respiratory-illness-surveillance-report/.

[263] S098: Sentiworld - Welcome !, . URL https://sentiworld.sentiweb.fr/.

[264] Giorgi Chakhunashvili, Abram L. Wagner, Laura E. Power, Cara B. Janusz, Ann Machablishvili, Irakli Karseladze, Olgha Tarkhan-Mouravi, Khatuna Zakhashvili, Paata Imnadze, Gregory C. Gray, Benjamin Anderson, and Matthew L. Boulton. S099: Severe Acute Respiratory Infection (SARI) sentinel surveillance in the country of Georgia, 2015-2017. *PloS one*, 13(7), 7 2018. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0201497. URL https://pubmed.ncbi.nlm.nih.gov/30059540/.

[265] Mohammad Abdullah Al Amad, Ali Ali Al Mahaqri, Abdulwahed Abdulgabar Al Serouri, and Yousef S. Khader. S100: Severe Acute Respiratory Infections With Influenza and Noninfluenza Respiratory Viruses: Yemen, 2011-2016. *Inquiry : a journal of medical care organization, provision and financing*, 56, 5 2019. ISSN 1945-7243. doi: 10.1177/0046958019850731. URL https://pubmed.ncbi.nlm.nih.gov/31137990/.

[266] James A. Zhou, Jo Ellen Schweinle, Richard Lichenstein, Robert E. Walker, and James C. King. S101: Severe Illnesses Associated With Outbreaks of Respiratory Syncytial Virus and Influenza in Adults. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 70(5):773–779, 3 2020. ISSN 1537-6591. doi: 10.1093/CID/CIZ264. URL https://pubmed.ncbi.nlm.nih.gov/30944930/.

[267] Meredith L. McMorrow, Emile Okitolonda Wemakoy, Joelle Kabamba Tshilobo, Gideon O. Emukule, Joshua A. Mott, Henry Njuguna, Lilian Waiboci, Jean Michel Heraud, Soatianana Rajatonirina, Norosoa H. Razanajatovo, Moses Chilombe, Dean Everett, Robert S. Heyderman, Amal Barakat, Thierry Nyatanyi, Joseph Rukelibuga, Adam L. Cohen, Cheryl Cohen, Stefano Tempia, Juno Thomas, Marietjie

Venter, Elibariki Mwakapeje, Marcelina Mponela, Julius Lutwama, Jazmin Duque, Kathryn Lafond, Ndahwouh Talla Nzussouo, Thelma Williams, and Marc Alain Widdowson. S102: Severe Acute Respiratory Illness Deaths in Sub-Saharan Africa and the Role of Influenza: A Case Series From 8 Countries. *The Journal of infectious diseases*, 212(6):853–860, 9 2015. ISSN 1537-6613. doi: 10.1093/INFDIS/JIV100. URL https://pubmed.ncbi.nlm.nih.gov/25712970/.

[268] Maryanne N. Gachari, Linus Ndegwa, Gideon O. Emukule, Lily Kirui, Rosalia Kalani, Bonventure Juma, Lilian Mayieka, Peter Kinuthia, Marc Alain Widdowson, and Sandra S. Chaves. S103: Severe acute respiratory illness surveillance for influenza in Kenya: Patient characteristics and lessons learnt. *Influenza and other respiratory viruses*, 16(4):740–748, 7 2022. ISSN 1750-2659. doi: 10.1111/IRV.12979. URL https://pubmed.ncbi.nlm.nih.gov/35289078/.

[269] Robert F. Breiman, Leonard Cosmas, M. Kariuki Njenga, John Williamson, Joshua A. Mott, Mark A. Katz, Dean D. Erdman, Eileen Schneider, M. Steven Oberste, John C. Neatherlin, Henry Njuguna, Daniel M. Ondari, Kennedy Odero, George O. Okoth, Beatrice Olack, Newton Wamola, Joel M. Montgomery, Barry S. Fields, and Daniel R. Feikin. S104: Severe acute respiratory infection in children in a densely populated urban slum in Kenya, 2007-2011. *BMC Infectious Diseases*, 15(1):1–11, 12 2015. ISSN 14712334. doi: 10.1186/S12879-015-0827-X/FIGURES/1. URL https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-015-0827-x.

[270] Cheryl Cohen, Jocelyn Moyes, Stefano Tempia, Michelle Groom, Sibongile Walaza, Marthi Pretorius, Halima Dawood, Meera Chhagan, Summaya Haffejee, Ebrahim Variava, Kathleen Kahn, Akhona Tshangela, Anne von Gottberg, Nicole Wolter, Adam L. Cohen, Babatyi Kgokong, Marietjie Venter, and Shabir A. Madhi. S105: Severe influenza-associated respiratory infection in high HIV prevalence setting, South Africa, 2009-2011. *Emerging infectious diseases*, 19(11):1766–1774, 11

2013. ISSN 1080-6059. doi: 10.3201/EID1911.130546. URL https://pubmed.ncbi.nlm.nih.gov/24209781/.

[271] Sana Eybpoosh, Mahdi Afshari, Ali-Akbar Haghdoost, Parvin Afsar Kazerooni, Mohammad Mehdi Gouya, and Katayoon Tayeri. S106: Severity and mortality of COVID-19 infection in HIV-infected individuals: Preliminary findings from Iran. *Medical journal of the Islamic Republic of Iran*, 35, 4 2021. ISSN 1016-1430. doi: 10.47176/MJIRI.35.33. URL https://pubmed.ncbi.nlm.nih.gov/34211935/.

[272] Jérôme O. Wishaupt, Tjeerd van der Ploeg, Ronald de Groot, Florens G.A. Versteegh, and Nico G. Hartwig. S107: Single-and multiple viral respiratory infections in children: Disease and management cannot be related to a specific pathogen. *BMC Infectious Diseases*, 17(1):1–11, 1 2017. ISSN 14712334. doi: 10.1186/S12879-016-2118-6/TABLES/6. URL https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-016-2118-6.

[273] Nicolas Padilla-Raygoza, Cuauhtemoc Sandoval-Salazar, Xóchitl Sofía Ramírez-Gómez, Efraín Navarro-Olivos, María de Jesus Gallardo-Luna, Francisco J. Magos-Vazquez, and Daniel Alberto Diaz-Martinez. S108: Status of novel coronavirus disease and analysis of mortality in Mexico, until June 30th, 2020: An ecological study. *Biomedical and Pharmacology Journal*, 13(4):1781–1790, 12 2020. ISSN 24562610. doi: 10.13005/BPJ/2053.

[274] Javier Dávila, Gerardo Chowell, Víctor H. Borja-Aburto, Cécile Viboud, Concepciòn Grajales Muñiz, and Mark Miller. S109: Substantial Morbidity and Mortality Associated with Pandemic A/H1N1 Influenza in Mexico, Winter 2013-2014: Gradual Age Shift and Severity. *PLoS currents*, 6, 2014. ISSN 2157-3999. doi: 10.1371/CURRENTS.OUTBREAKS.A855A92F19DB1D90CA955F5E908D6631. URL https://pubmed.ncbi.nlm.nih.gov/24744975/.

[275] Fabio Tramuto, Vincenzo Restivo, Claudio Costantino, Giuseppina Maria Elena Colomba, Carmelo Massimo Maida, Alessandra Casuccio, and Francesco Vitale.

S110: Surveillance Data for Eight Consecutive Influenza Seasons in Sicily, Italy. *The American journal of tropical medicine and hygiene*, 101(6):1232–1239, 2019. ISSN 1476-1645. doi: 10.4269/AJTMH.19-0059. URL https://pubmed.ncbi.nlm.nih.gov/31571567/.

[276] Jennifer R. Verani, John McCracken, Wences Arvelo, Alejandra Estevez, Maria Renee Lopez, Lissette Reyes, Juan Carlos Moir, Chris Bernart, Fabiola Moscoso, Jennifer Gray, Sonja J. Olsen, and Kim A. Lindblade. S111: Surveillance for Hospitalized Acute Respiratory Infection in Guatemala. *PLoS ONE*, 8(12):e83600, 12 2013. ISSN 19326203. doi: 10.1371/JOURNAL.PONE.0083600. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3877070/.

[277] Tamara J. Meerhoff, Artan Simaku, Dritan Ulqinaku, Liana Torosyan, Natalia Gribkova, Veronica Shimanovich, Giorgi Chakhunashvili, Irakli Karseladze, Aizhan Yesmagambetova, Ainagul Kuatbayeva, Zuridin Nurmatov, Dinagul Otorbaeva, Emilia Lupulescu, Odette Popovici, Elizaveta Smorodintseva, Anna Sominina, Olga Holubka, Olga Onyshchenko, Caroline S. Brown, and Diane Gross. S112: Surveillance for severe acute respiratory infections (SARI) in hospitals in the WHO European region - an exploratory analysis of risk factors for a severe outcome in influenza-positive SARI cases. *BMC infectious diseases*, 15 (1), 1 2015. ISSN 1471-2334. doi: 10.1186/S12879-014-0722-X. URL https://pubmed.ncbi.nlm.nih.gov/25567701/.

[278] F. Tramuto, C. M. Maida, F. Bonura, A. M. Perna, S. Puzelli, M. A. de Marco, I. Donatelli, L. Aprea, A. Firenze, A. Arcadipane, U. Palazzo, and F. Vitale. S113: Surveillance of hospitalised patients with influenza-like illness during pandemic influenza A(H1N1) season in Sicily, April 2009 - December 2010. *Eurosurveillance*, 16(35):19957, 9 2011. ISSN 15607917. doi: 10.2807/ESE.16.35.19957-EN/CITE/REFWORKS. URL https://www.eurosurveillance.org/content/10.2807/ese.16.35.19957-en.

[279] G. Sigmundsdottir, T. Gudnason, Ö Ólafsson, G. E. Baldvinsdóttir, A. Atladottir,

A. Löve, L. Danon, and H. Briem. S114: Surveillance of influenza in Iceland during the 2009 pandemic. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 15(49):3, 2010. ISSN 1560-7917. doi: 10.2807/ESE.15.49.19742-EN. URL https://pubmed.ncbi.nlm.nih.gov/21163181/.

[280] Salah Al-Awaidy, Sarah Hamid, Idris Al Obaidani, Said Al Baqlani, Suleiman Al Busaidi, Shyam Bawikar, Waleed El-Shoubary, Erica L. Dueger, Mayar M. Said, Emdeldin Elamin, Parag Shah, and Maha Talaat. S115: The Burden of Influenza-Associated Hospitalizations in Oman, January 2008-June 2013. *PLOS ONE*, 10(12): e0144186, 12 2015. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0144186. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144186.

[281] Gideon O. Emukule, Fredrick Otiato, Bryan O. Nyawanda, Nancy A. Otieno, Caroline A. Ochieng, Linus K. Ndegwa, Peter Muturi, Godfrey Bigogo, Jennifer R. Verani, Philip M. Muthoka, Elizabeth Hunsperger, and Sandra S. Chaves. S116: The Epidemiology and Burden of Influenza B/Victoria and B/Yamagata Lineages in Kenya, 2012–2016. *Open Forum Infectious Diseases*, 6(10):ofz421, 10 2019. ISSN 23288957. doi: 10.1093/OFID/OFZ421. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6804754/.

[282] Yasser Sakr, Ricard Ferrer, Konrad Reinhart, Richard Beale, Andrew Rhodes, Rui Moreno, Jean Francois Timsit, Laurent Brochard, B. Taylor Thompson, Ederlon Rezende, Jean Daniel Chiche, IC-GLOSSARI Investigators, and ESICM Trials Group. S117: The Intensive Care Global Study on Severe Acute Respiratory Infection (IC-GLOSSARI): a multicenter, multinational, 14-day inception cohort study. *Intensive Care Medicine*, 42(5):817, 5 2016. ISSN 14321238. doi: 10.1007/S00134-015-4206-2. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7080095/.

[283] André Ricardo Ribas Freitas, Otto Albuquerque Beckedorff, Luciano Pamplona

de Góes Cavalcanti, Andre M. Siqueira, Daniel Barros de Castro, Cristiano Fernandes da Costa, Daniele Rocha Queiróz Lemos, and Eliana N.C. Barros. S118: The emergence of novel SARS-CoV-2 variant P.1 in Amazonas (Brazil) was temporally associated with a change in the age and sex profile of COVID-19 mortality: A population based ecological study. *The Lancet Regional Health - Americas*, 1:0, 9 2021. ISSN 2667193X. doi: 10.1016/j.lana.2021.100021. URL https://www.thelancet.com/action/showFullText?pii=S2667193X21000132https://www.thelancet.com/action/showAbstract?pii=S2667193X21000132https://www.thelancet.com/journals/lanam/article/PIIS2667-193X(21)00013-2/abstract.

[284] Saverio Caini, Doménica de Mora, Maritza Olmedo, Denisses Portugal, María A. Becerra, Marcela Mejía, María C. Pacurucu, Jenny Ojeda, Guglielmo Bonaccorsi, Chiara Lorini, John Paget, and Alfredo Bruno. S119: The epidemiology and severity of respiratory viral infections in a tropical country: Ecuador, 2009-2016. *Journal of infection and public health*, 12(3):357–363, 5 2019. ISSN 1876-035X. doi: 10.1016/J.JIPH.2018.12.003. URL https://pubmed.ncbi.nlm.nih.gov/30573330/.

[285] Rogelio Perez-Padilla, Cecilia Garcia-Sancho, Rosario Fernandez, Francisco Franco-Marina, Hugo Lopez-Gatell, and Ietza Bojorquez. S120: The impact of altitude on hospitalization and hospital mortality from pandemic 2009 influenza A (H1N1) virus pneumonia in Mexico. *Salud publica de Mexico*, 55(1):92–95, 2013. ISSN 1606-7916. doi: 10.1590/S0036-36342013000100013. URL https://pubmed.ncbi.nlm.nih.gov/23370263/.

[286] Albert Jan van Hoek, Anthony Underwood, Mark Jit, Elizabeth Miller, and W. John Edmunds. S121: The impact of pandemic influenza H1N1 on health-related quality of life: a prospective population-based study. *PloS one*, 6(3), 2011. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0017030. URL https://pubmed.ncbi.nlm.nih.gov/21399678/.

[287] Hongjie Yu, Jigui Huang, Yang Huai, Xuhua Guan, John Klena, Shali Liu, Youxing Peng, Hui Yang, Jun Luo, Jiandong Zheng, Maoyi Chen, Zhibin Peng, Nijuan Xiang, Xixiang Huo, Lin Xiao, Hui Jiang, Hui Chen, Yuzhi Zhang, Xuesen Xing, Zhen Xu, Zijian Feng, Faxian Zhan, Weizhong Yang, Timothy M. Uyeki, Yu Wang, and Jay K. Varma. S122: The substantial hospitalization burden of influenza in central China: surveillance for severe, acute respiratory infection, and influenza viruses, 2010-2012. *Influenza and other respiratory viruses*, 8(1):53–65, 2014. ISSN 1750-2659. doi: 10.1111/IRV.12205. URL https://pubmed.ncbi.nlm.nih.gov/24209711/.

[288] Dan Morgenstern-Kaplan, Bruno Buitano-Tang, Mercedes Martínez-Gil, Andrea Zaldívar Pérez Pavón, and Juan O. Talavera. S123: U-shaped-aggressiveness of SARS-CoV-2: Period between initial symptoms and clinical progression to COVID-19 suspicion. A population-based cohort study. *PloS one*, 15(12), 12 2020. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0243268. URL https://pubmed.ncbi.nlm.nih.gov/33270769/.

[289] Srinivas Murthy. S124: Using research to prepare for outbreaks of severe acute respiratory infection. *BMJ Global Health*, 4(1):1061, 2 2019. ISSN 2059-7908. doi: 10.1136/BMJGH-2018-001061. URL https://gh.bmj.com/content/4/1/e001061.

[290] T. Sonia Boender, Wei Cai, Madlen Schranz, Theresa Kocher, Birte Wagner, Alexander Ullrich, Silke Buda, Rebecca Zöllner, Felix Greiner, Michaela Diercke, and Linus Grabenhenrich. S125: Using routine emergency department data for syndromic surveillance of acute respiratory illness, Germany, week 10 2017 until week 10 2021. *Eurosurveillance*, 27(27):2100865, 7 2022. ISSN 15607917. doi: 10.2807/1560-7917.ES.2022.27.27.2100865/CITE/REFWORKS. URL https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2022.27.27.2100865.

[291] Katherine C. Horton, Erica L. Dueger, Amr Kandeel, Mohamed Abdallat, Amani

El-Kholy, Salah Al-Awaidy, Abdul Hakim Kohlani, Hanaa Amer, Abel Latif ElKhal, Mayar Said, Brent House, Guillermo Pimentel, and Maha Talaat. S126: Viral etiology, seasonality and severity of hospitalized patients with severe acute respiratory infections in the Eastern Mediterranean Region, 2007–2014. *PLOS ONE*, 12(7): e0180954, 7 2017. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0180954. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0180954.

Appendices

Appendix A1

Chapter 1 Introduction

A1.1 CASE DEFINITIONS

Table A1.1: Surveillance case definitions for acute respiratory infection (ARI) and related syndromes

Acute Respiratory Infection (ARI) (EU, 2018)

- Sudden onset of symptoms

AND

- At least one of: cough, sore throat, shortness of breath, or coryza

AND

- Clinician's judgement that illness is due to an infection

Influenza-like Illness (ILI) (RCGP-RSC)

Continued on next page

Table A1.1: Surveillance case definitions for acute respiratory infection (ARI) and related syndromes (continued)

- Acute respiratory illness with temperature ≥38°C **and** cough, onset within the past 10 days
- Should not have another more plausible diagnosis
- Systemic upset (myalgia, fatigue, malaise, headache, etc.) is common
- Swabs ideally taken within 7 days of onset

Upper Respiratory Tract Infection (URTI)

An acute infection above and including the larynx, encompassing the nasal mucosa, paranasal sinuses, middle ear, pharynx, and larynx.

Lower Respiratory Tract Infection (LRTI)

An acute infection below the larynx, including bronchus, bronchioles, and alveoli.

Exacerbations of Chronic Lung Disease (ECLD)

A sudden deterioration in respiratory symptoms (e.g., cough, sputum production, breathlessness) in individuals with chronic lung disease.

Appendix A2

Chapter 3 Systematic review

A2.1 SEARCH STRATEGIES

A2.1.1 Global Health (2009 to 2023 Week 23)

Table A2.1: Global Health database search strategy

#	Search Strategy	Results
1	surveillance/ or sentinel surveillance/ or syndromic surveillance/	29,423
2	(ARI or SARI or ILI).mp.	4,924
3	"surveillance indicator*".mp. or surveillance.in.	10,225
4	("public health effect*" or "epidemiological effect*").mp.	364
5	1 or 2 or 3 or 4	43,049
6	exp influenza/ or influenza.mp.	47,257
7	SARS-CoV-2/ or COVID-19/	111,070
8	(corona* adj1 (virus* or viral*)).mp.	2,093
9	CoV not (...) .ti,ab.	46,438
10	(coronavirus* or 2019nCoV* ...).mp.	126,133
11	"severe acute respiratory syndrome".mp.	110,444
12	7 or 8 or 9 or 10 or 11	126,595
13	limit 12 to yr="2020-current"	120,492
14	6 or 13	163,906
15	("hospitali?ation rat*" or ...).mp.	184,226
16	(sever* or "seriousness").ti.	39,616
17	("clinical* sever*" or "clinical* serious*").mp.	2,278
18	((sever* or serious*) adj2 ...).mp.	71,038
19	15 or 16 or 17 or 18	270,359
20	5 and 14 and 19	1,732
21	exp animals/ not humans/	720,712
22	20 not 21	1,701
23	limit 22 to yr="2009-current"	1,616
24	limit 23 to english language	1,486

A2.1.2 Embase (1974 to present)

Table A2.2: Embase database search strategy

#	Search Strategy	Results
1	community based surveillance/ or epidemiological surveillance/	1,613
2	(ARI or SARI or ILI).ti,ab,kf.	11,313
3	"surveillance indicator*".ti,ab,kf. or surveillance.in.	32,987
4	("public health effect*" or "epidemiological effect*").ti,ab,kf.	719
5	1 or 2 or 3 or 4	46,300
6	exp *influenza/	56,430
7	influenza.ti,ab.	129,765
8	(corona* adj1 (virus* or viral*)).ti,ab,kf.	6,062
9	CoV not (...) .ti,ab,kf.	151,936
10	(coronavirus* or 2019nCoV* ...).ti,ab,kf.	456,528
11	exp *severe acute respiratory syndrome/ or ...	383,741
12	8 or 9 or 10 or 11	486,040
13	limit 12 to yr="2020-current"	469,303
14	6 or 7 or 13	597,126
15	("hospitali?ation rat*" or ...).ti,ab,kf.	321,556
16	(severe or severity).ti.	323,591
17	("clinical* sever*" or "clinical* serious*").ti,ab,kf.	19,920
18	((sever* or serious*) adj2 ...).ti,ab,kf.	351,980
19	15 or 16 or 17 or 18	944,751
20	5 and 14 and 19	1,299
21	exp animal/ not human/	5,333,940
22	20 not 21	1,288
23	limit 22 to yr="2009-current"	1,260
24	limit 23 to english language	1,213

A2.1.3 MEDLINE (1946 to present)

Table A2.3: MEDLINE database search strategy

#	Search Strategy	Results
1	Sentinel Surveillance/ or Public Health Surveillance/	12,077
2	(ARI or SARI or ILI).ti,ab,kf.	8,064
3	"surveillance indicator*".ti,ab,kf. or surveillance.in.	21,946
4	("public health effect*" or "epidemiological effect*").ti,ab,kf.	678
5	1 or 2 or 3 or 4	41,720
6	influenza.ti,ab,kf. or Influenza, Human/	122,865
7	SARS-CoV-2/ or COVID-19/	233,242
8	(corona* adj1 (virus* or viral*)).ti,ab,kf.	4,442
9	CoV not (...) .ti,ab.	106,972
10	(coronavirus* or 2019nCoV* ...).ti,ab,kf.	370,567
11	"severe acute respiratory syndrome".ti,ab,kf.	42,896
12	7 or 8 or 9 or 10 or 11	380,183
13	limit 12 to yr="2020-current"	366,869
14	6 or 13	481,854
15	("hospitali?ation rat*" or ...).ti,ab,kf.	224,265
16	(sever* or "seriousness").ti.	269,411
17	("clinical* sever*" or "clinical* serious*").ti,ab,kf.	13,163
18	((sever* or serious*) adj2 ...).ti,ab,kf.	234,861
19	15 or 16 or 17 or 18	693,052
20	5 and 14 and 19	1,015
21	exp animals/ not humans/	5,130,390
22	20 not 21	1,008
23	limit 22 to yr="2009-current"	984
24	limit 23 to english language	954

A2.1.4 Grey Literature Search (15 March 2024)

To complement database searching, five public health web portals were searched for surveillance reports published between 1 January 2018 and 31 December 2023 that contained human respiratory-virus surveillance data.

1. **United States Centers for Disease Control and Prevention (CDC)**:
   "Respiratory virus surveillance reports" query via site search; all returned titles and snippets screened until saturation.
   `https://search.cdc.gov/search/?query=Respiratory%20virus%20surveillance%20reports&dpage=1`

2. **World Health Organization (WHO) Disease Outbreak News**:
   First 300 notices reviewed for routine human respiratory-virus surveillance data.
   `https://www.who.int/emergencies/disease-outbreak-news`

3. **UK Health Security Agency (UKHSA)**:
   GOV.UK search for respiratory-virus surveillance bulletins sorted by relevance.
   `https://www.gov.uk/search/all?keywords=Respiratory+virus+surveillance+reports`

4. **European Centre for Disease Prevention and Control (ECDC)**:
   Screened the 20 most recent monitoring updates for human respiratory-virus content.
   `https://www.ecdc.europa.eu/en/publications-data/monitoring/epidemiological-updates`

5. **Sentiworld sentinel-network portal**:
   Accessed systems from `https://www.sentiworld.info` and browsed publications for relevant surveillance bulletins.

For every source, titles, executive summaries, and (where available) PDF bulletins were examined. Items that met the inclusion criteria were exported as PDF.

A2.2 DATA ITEMS EXTRACTED FROM STUDIES

Table A2.4: Data items collected for systematic review

Table A2.4: Data items collected for systematic review

Category	Data Element
Study Details	1. Study title 2. Study aim 3. Publication type and date 4. Study period (start and end date) 5. Geographical scope (country(s) and region(s))
Case Information	**Case types:** 1. acute respiratory infection (ARI) 2. influenza-like illness (ILI) 3. severe acute respiratory infection (SARI) 4. Suspected COVID 5. Other **Case definitions used:** 1. World Health Organization (WHO) 2. United States Centers for Disease Control and Prevention (CDC) 3. Other **Number of cases:** 1. > 500 **Case recruitment type:** 1. Hospitalised 2. Treatment-seeking (e.g., primary care) 3. Non-treatment-seeking (e.g., community survey) 4. intensive care unit (ICU) 5. Mixed
Severity Markers	**Severity marker type:** 1. Severe outcomes (e.g., complication, death) 2. Predictors of severe outcomes (e.g., clinical signs)

A2.3 REFERENCES FROM SYSTEMATIC REVIEW

Table A2.5: List of included studies relevant to acute respiratory infection surveillance (Appendix)

Study ID	Study title
S001 [171]	A Comparison of the Epidemiology and Clinical Presentation of Seasonal Influenza A and 2009 Pandemic Influenza A (H1N1) in Guatemala
S002 [172]	A Comprehensive Descriptive Epidemiological and ClinicalAnalysis of SARS-CoV-2 in West-Mexico during COVID-19Pandemic 2020
S003 [173]	A clinico-epidemiological profile, coinfections and outcome of patients with Influenza Like Illnesses (ILI) presenting to the emergency department during the COVID-19 pandemic
S004 [174]	A cross-sectional analysis of symptom severity in adults with influenza and other acute respiratory illness in the outpatient setting.
S005 [175]	A review of the dynamics and severity of the pandemic A(H1N1) influenza virus on R union island, 2009
S006 [176]	Analysis of Risk Factors for Severe Acute Respiratory Infection and Pneumonia and among Adult Patients with Acute Respiratory Illness during 2011-2014 Influenza Seasons in Korea
S007 [177]	Antibiotic Use in Suspected and Confirmed COVID-19 Patients Admitted to Health Facilities in Sierra Leone in 2020–2021: Practice Does Not Follow Policy
S008 [178]	Associations between chronic conditions and death in hospital among adults (aged 20+years) during first acute care hospitalizations with a confirmed or suspected COVID-19 diagnosis in Canada
S009 [179]	Building influenza surveillance pyramids in near real time, Australia
S010 [180]	CDC-Respiratory Virus Surveillance Among Children with Acute Respiratory Illnesses – New Vaccine Surveillance Network, United States, 2016–2021
S011 [181]	COVID-19 and Severe Acute Respiratory Infections: Monitoring Trends in 421 German Hospitals During the First Four Pandemic Waves

Table A2.5: Table continued

Study ID	Study title
S012 [182]	Changes in Surveillance of Acute Respiratory Infections Including Influenza in Slovak Republic during 1993–2008
S013 [183]	Characterisation of acute respiratory infections at a United Kingdom paediatric teaching hospital: observational study assessing the impact of influenza A (2009 pdmH1N1) on predominant viral pathogens
S014 [184]	Characterizing the Epidemiology of the 2009 Influenza A/H1N1 Pandemic in Mexico
S015 [185]	Chronic use of inhaled corticosteroids in patients admitted for respiratory virus infections: a 6-year prospective multicenter study
S016 [186]	Clinical Profile of COVID-19 Illness in Children—Experience from a Tertiary Care Hospital
S017 [187]	Clinical and epidemiological aspects of severe acute respiratory infection: before and during the first year of the COVID-19 pandemic in Brazil
S018 [188]	Clinical and phylogenetic influenza dynamics for the 2019-20 season in the global influenza hospital surveillance network (GIHSN) – Pilot study
S019 [189]	Clinical characteristics and factors associated with hospital admission or death in 43 103 adult outpatients with coronavirus disease 2019 managed with the Covidom telesurveillance solution: a prospective cohort study
S020 [190]	Clinical characteristics and factors associated with severe acute respiratory infection and influenza among children in Jingzhou, China
S021 [191]	Clinical characteristics and outcome of respiratory syncytial virus infection among adults hospitalized with influenza-like illness in France
S022 [192]	Clinical characteristics and outcomes of influenza and other influenza-like illnesses in Mexico City
S023 [193]	Clinical characteristics and outcomes of patients with severe acute respiratory infections (SARI): results from the Egyptian surveillance study 2010–2014

Table A2.5: Table continued

Study ID	Study title
S024 [194]	Clinical features and mortality in COVID-19 SARI versus non-COVID-19 SARI cases from Western Rajasthan, India
S025 [195]	Comparative epidemiology of five waves of COVID-19 in Mexico, March 2020–August 2022
S026 [196]	Comparison of common acute respiratory infection case definitions for identification of hospitalized influenza cases at a population-based surveillance site in Egypt
S027 [197]	Coronavirus Surveillance in a Pediatric Population in Jordan From 2010 to 2013: A Prospective Viral Surveillance Study
S028 [198]	Deaths and hospitalizations related to 2009 pandemic influenza A (H1N1) – Greece, May 2009–February 2010
S029 [199]	Description of Hospitalized Cases of Influenza A(H1N1)pdm09 Infection on the Basis of the National Hospitalized-Case Surveillance, 2009–2010, Japan
S030 [200]	Descriptive Epidemiology of Novel Influenza A (H1N1), Andhra Pradesh 2009–2010
S031 [201]	Detection of Influenza C Viruses Among Outpatients and Patients Hospitalized for Severe Acute Respiratory Infection, Minnesota, 2013–2016
S032 [202]	Detection of Viral and Bacterial Pathogens in Hospitalized Children With Acute Respiratory Illnesses, Chongqing, 2009–2013
S033 [203]	Digitalizing and upgrading severe acute respiratory infections surveillance in Malta: system development
S034 [204]	Disparities Between Black and White Children in Hospitalizations Associated With Acute Respiratory Illness and Laboratory-confirmed Influenza and Respiratory Syncytial Virus in 3 US Counties – 2002–2009

Table A2.5: Table continued

Study ID	Study title
S035 [205]	Effects of seasonal and pandemic influenza on health-related quality of life, work and school absence in England: Results from the Flu Watch cohort study
S036 [206]	Epidemiologic and spatiotemporal characterization of influenza and severe acute respiratory infection in Uganda, 2010–2015
S037 [207]	Epidemiologic profile of severe acute respiratory infection in Brazil during the COVID-19 pandemic: an epidemiological study
S038 [208]	Epidemiological Characteristics and Underlying Risk Factors for Mortality during the Autumn 2009 Pandemic Wave in Mexico
S039 [200]	Epidemiological Characterization of a Fourth Wave of Pandemic A/H1N1 Influenza in Mexico, Winter 2011–2012: Age Shift and Severity
S040 [209]	Epidemiological Trends of Coronavirus Disease 2019 in Sierra Leone From March 2020 to October 2021
S041 [210]	Epidemiological and clinical profile of Influenza A(H1N1) pdm09 in Odisha, eastern India
S042 [211]	Epidemiology and outcome of influenza-associated infections among hospitalized patients with acute respiratory infections, Egypt national surveillance system, 2016–2019
S043 [212]	Epidemiology and virology of acute respiratory infections during the first year of life: a birth cohort study in Vietnam
S044 [213]	Epidemiology of Influenza in Fars Province, Southern Iran; a Population-Based Study (2015–2019)
S045 [214]	Epidemiology of Severe Acute Respiratory Illness and Risk Factors for Influenza Infection and Clinical Severity among Adults in Malawi, 2011–2013

Table A2.5: Table continued

Study ID	Study title
S046 [215]	Epidemiology of hospital admissions with influenza during the 2013/2014 Northern Hemisphere influenza season: results from the Global Influenza Hospital Surveillance Network
S047 [216]	Epidemiology of influenza B infection in the state of Rio Grande do Sul, Brazil, from 2003 to 2019
S048 [217]	Epidemiology of severe acute respiratory illness (SARI) among adults and children aged ≤ 5 years in a high HIV-prevalence setting, 2009–2012
S049 [218]	Epidemiology, disease severity and outcome of Severe acute respiratory syndrome coronavirus 2 and influenza viruses coinfection seen at Egypt integrated acute respiratory infections surveillance, 2020–2022
S050 [219]	Establishing a sentinel surveillance system for the novel COVID-19 in a resource-limited country: methods, system attributes and early findings
S051 [220]	Establishing an ICD-10 code based SARI-surveillance in Germany – description of the system and first results from five recent influenza seasons
S052 [221]	Establishing thresholds and parameters for pandemic influenza severity assessment, Australia
S053 [222]	Estimating age-specific influenza-related hospitalization rates during the pandemic (H1N1) 2009 in Davidson Co, TN
S054 [223]	Estimating severity of influenza epidemics from severe acute respiratory infections (SARI) in intensive care units
S055 [224]	Estimating the burden of influenza-associated hospitalization and deaths in Oman (2012–2015)
S056 [225]	Etiology, clinical characteristics, and risk factors associated with severe influenza-like illnesses in Mexican adults
S057 [226]	Evaluation of a new clinical endpoint for moderate to severe influenza disease in children: A prospective cohort study
S058 [227]	Facility-based surveillance for influenza and respiratory syncytial virus in rural Zambia

Table A2.5: Table continued

Study ID	Study title
S059 [228]	Factors associated with Severe Acute Respiratory Syndrome in a Brazilian central region
S060 [229]	Factors associated with poor outcomes in patients with severe acute respiratory infections in Bahrain
S061 [68]	First-year results of the Global Influenza Hospital Surveillance Network: 2012–2013 northern hemisphere influenza season
S062 [230]	Five-year community surveillance study for acute respiratory infections using text messaging: findings from the MoSAIC study
S063 [231]	Hospital-based surveillance of influenza A(H1N1)pdm09 virus in Saudi Arabia, 2010–2016
S064 [232]	Hospitalization of influenza-like illness patients recommended by general practitioners in France between 1997 and 2010
S065 [69]	Impact of 2009 pandemic influenza among Vietnamese children based on a population-based prospective surveillance from 2007 to 2011
S066 [233]	Impact of influenza infection on children's hospital admissions during two seasons in Athens, Greece
S067 [234]	Impact of influenza on outpatient visits, hospitalizations, and deaths by using a time series Poisson generalized additive model
S068 [235]	Implementing the World Health Organization Pandemic Influenza Severity Assessment framework—Singapore's experience
S069 [236]	Incidence, disease severity, and follow-up of influenza A/A, A/B, and B/B virus dual infections in children: a hospital-based digital surveillance program
S070 [237]	Influenza B associated paediatric acute respiratory infection hospitalization in central Vietnam
S071 [238]	Influenza hospitalization epidemiology from a severe acute respiratory infection surveillance system in Jordan, January 2008–February 2014

Table A2.5: Table continued

Study ID	Study title
S072 [239]	Lessons from the epidemiological surveillance program, during the influenza A (H1N1) virus epidemic, in a reference university hospital of southeastern Brazil
S073 [240]	Lethality and characteristics of deaths due to COVID-19 in Rondonia: an observational study
S074 [241]	Morbidity, mortality, and seasonality of influenza hospitalizations in Egypt, November 2007–November 2014
S075 [242]	Mortality amongst patients with influenza-associated severe acute respiratory illness, South Africa, 2009–2013
S076 [243]	Mortality, Severe Acute Respiratory Infection, and Influenza-Like Illness Associated with Influenza A(H1N1)pdm09 in Argentina, 2009
S077 [244]	National retrospective cohort study to identify age-specific fatality risks of comorbidities among hospitalised patients with influenza-like illness in Taiwan
S078 [245]	Non-influenza respiratory viruses in adult patients admitted with influenza-like illness: a 3-year prospective multicenter study
S079 [246]	Novel influenza A(H1N1) in a pediatric health care facility in New York City during the first wave of the 2009 pandemic
S080 [247]	Occurrence of AH1N1 viral infection and clinical features in symptomatic patients who received medical care during the 2009 influenza pandemic in Central Mexico
S081 [248]	Outbreak of 2009 pandemic influenza A(H1N1), Los Lagos, Chile, April–June 2009
S082 [70]	Outcomes of influenza A(H1N1)pdm09 virus infection: results from two international cohort studies

Table A2.5: Table continued

Study ID	Study title
S083 [70]	Outcomes of patients with Severe Acute Respiratory Infections (SARI) admitted to the intensive care unit: results from the Egyptian Surveillance Study 2010–2014
S084 [249]	Overview of the winter wave of 2009 pandemic influenza A(H1N1)v in Vojvodina, Serbia
S085 [250]	Pandemic H1N1 influenza in Brazil: Analysis of the first 34 506 notified cases of influenza-like illness with severe acute respiratory infection (SARI)
S086 [251]	Patient characteristics associated with COVID-19 positivity and fatality in Nigeria: retrospective cohort study
S087 [252]	Persistent Functional Decline Following Hospitalization with Influenza or Acute Respiratory Illness
S088 [253]	Prevalence and risk factors for long COVID after mild disease: A cohort study with a symptomatic control group
S089 [254]	Prevalence, incidence, and severity associated with viral respiratory tract infections in Colombian adults before the COVID-19 pandemic
S090 [255]	Recrudescent wave of pandemic A/H1N1 influenza in Mexico, winter 2011–2012: Age shift and severity
S091 [256]	Respiratory syncytial virus and other respiratory viral infections in older adults with moderate to severe influenza-like illness
S092 [257]	Respiratory virus-associated severe acute respiratory illness and viral clustering in Malawian children in a setting with a high prevalence of HIV infection, malaria, and malnutrition
S093 [258]	Risk Factors for Severe Coronavirus Disease 2019 Among Human Immunodeficiency Virus-Infected and -Uninfected Individuals in South Africa, April 2020–March 2022: Data From Sentinel Surveillance
S094 [259]	Risk factors for influenza-associated severe acute respiratory illness hospitalization in South Africa, 2012–2015

Table A2.5: Table continued

Study ID	Study title
S095 [260]	Risk factors of prolonged hospital stay in children with viral severe acute respiratory infections
S096 [261]	SARS-CoV-2 and influenza virus coinfection among patients with severe acute respiratory infection during the first wave of COVID-19 pandemic in Bangladesh: a hospital-based descriptive study
S097 [262]	Sentiworld-2022 Acute Respiratory Illness Surveillance Report
S098 [263]	Sentiworld-Weekly Report on Severe Acute Respiratory Infection (SARI), Week 20 2023 (week ending 21/05/2023)
S099 [264]	Severe Acute Respiratory Infection(SARI) sentinel surveillance in the country of Georgia, 2015–2017
S100 [265]	Severe Acute Respiratory Infections With Influenza and Noninfluenza Respiratory Viruses: Yemen, 2011–2016
S101 [266]	Severe Illnesses Associated With Outbreaks of Respiratory Syncytial Virus and Influenza in Adults
S102 [267]	Severe acute respiratory illness deaths in sub-Saharan Africa and the role of influenza: a case series from 8 countries
S103 [268]	Severe acute respiratory illness surveillance for influenza in Kenya: Patient characteristics and lessons learnt
S104 [269]	Severe acute respiratory infection in children in a densely populated urban slum in Kenya, 2007–2011
S105 [270]	Severe influenza-associated respiratory infection in high HIV prevalence setting, South Africa, 2009–2011
S106 [271]	Severity and mortality of COVID-19 infection in HIV-infected individuals: Preliminary findings from Iran
S107 [272]	Single- and multiple viral respiratory infections in children: disease and management cannot be related to a specific pathogen
S108 [273]	Status of novel coronavirus disease and analysis of mortality in Mexico, until June 30th, 2020: An ecological study

Table A2.5: Table continued

Study ID	Study title
S109 [274]	Substantial Morbidity and Mortality Associated with Pandemic A/H1N1 Influenza in Mexico, Winter 2013-2014: Gradual Age Shift and Severity
S110 [275]	Surveillance data for eight consecutive influenza seasons in Sicily, Italy
S111 [276]	Surveillance for hospitalized acute respiratory infection in Guatemala
S112 [277]	Surveillance for severe acute respiratory infections (SARI) in hospitals in the WHO European region – an exploratory analysis of risk factors for a severe outcome in influenza-positive SARI cases
S113 [278]	Surveillance of hospitalised patients with influenza-like illness during pandemic influenza A(H1N1) season in Sicily, April 2009–December 2010
S114 [279]	Surveillance of influenza in Iceland during the 2009 pandemic
S115 [280]	The Burden of Influenza-Associated Hospitalizations in Oman, January 2008–June 2013
S116 [281]	The Epidemiology and Burden of Influenza B/Victoria and B/Yamagata Lineages in Kenya, 2012–2016
S117 [282]	The Intensive Care Global Study on Severe Acute Respiratory Infection (IC-GLOSSARI): a multicenter, multinational, 14-day inception cohort study
S118 [283]	The emergence of novel SARS-CoV-2 variant P.1 in Amazonas (Brazil) was temporally associated with a change in the age and sex profile of COVID-19 mortality: A population based ecological study
S119 [284]	The epidemiology and severity of respiratory viral infections in a tropical country: Ecuador, 2009–2016
S120 [285]	The impact of altitude on hospitalization and hospital mortality from pandemic 2009 influenza A (H1N1) virus pneumonia in Mexico
S121 [286]	The impact of pandemic influenza H1N1 on health-related quality of life: A prospective population-based study

Table A2.5: Table continued

Study ID	Study title
S122 [287]	The substantial hospitalization burden of influenza in central China: surveillance for severe, acute respiratory infection, and influenza viruses, 2010–2012
S123 [288]	U-shaped-aggressiveness of SARS-CoV-2: Period between Initial Symptoms and Clinical Progression to COVID-19 Suspicion. A Population-Based Cohort Study
S124 [289]	Using research to prepare for outbreaks of severe acute respiratory infection
S125 [290]	Using routine emergency department data for syndromic surveillance of acute respiratory illness, Germany, week 10 2017 until week 10 2021
S126 [291]	Viral etiology, seasonality and severity of hospitalized patients with severe acute respiratory infections in the Eastern Mediterranean Region, 2007–2014

A2.4 STUDY DETAILS

A2.4.1 Study Title and Aims

This section provides study codes, titles, and aims for all included studies.

DETAILS study_id	DETAILS study_title	DETAILS study_aim
S001	A Comparison of the Epidemiology and Clinical Presentation of Seasonal Influenza A and 2009 Pandemic Influenza A (H1N1) in Guatemala	We compare the epidemiology and clinical presentation of seasonal influenza A (H1N1 and H3N2) and 2009 pandemic influenza A (H1N1) (pH1N1) using a prospective surveillance system for acute respiratory disease in Guatemala.
S002	A Comprehensive Descriptive Epidemiological and ClinicalAnalysis of SARS-CoV-2 in West-Mexico during COVID-19Pandemic 2020	This study aimed to summarize the epidemiological and clinical characteristics of COVID-19 from Western Mexico people during 2020.
S003	A clinico-epidemiological profile, coinfections and outcome of patients with Influenza Like Illnesses (ILI) presenting to the emergency department during the COVID-19 pandemic	This is a prospective study comparing the demographics, clinical profile, co-infection with other viruses, and predictors of mortality in COVID-19 patients presenting with typical symptoms of an ILI to those with atypical symptoms and would help our primary care physicians in managing such cases with limited resources.
S004	A cross-sectional analysis of symptom severity in adults with influenza and other acute respiratory illness in the outpatient setting.	We examined the association between a subjective symptom severity score, demographic and clinical characteristics, and presence of laboratory-confirmed influenza among central Wisconsin adults who sought care for ARI during four influenza seasons.
S005	A review of the dynamics and severity of the pandemic A(H1N1) influenza virus on R,union island, 2009	This report summarizes the results of this surveillance and describes the dynamics and impact of the influenza pan- demic on Reïunion Island and the characteristics of labora- tory-confirmed cases, including hospitalized, severe and fatal cases.
S006	Analysis of Risk Factors for Severe Acute Respiratory Infection and	In this study, therefore, we aimed to identify risk factors for SARI and pneumonia among adult patients with ARI during the 2011-2014 influenza seasons using HIMM surveillance data in Korea.

	Pneumonia and among Adult Patients with Acute Respiratory Illness during 2011-2014 Influenza Seasons in Korea	
S007	Antibiotic Use in Suspected and Confirmed COVID-19 Patients Admitted to Health Facilities in Sierra Leone in 2020?2021: Practice Does Not Follow Policy	In this study, we report on the prevalence of antibiotic use and its associated factors among suspected and confirmed COVID-19 patients admitted to 35 health facilities in Sierra Leone from March 2020?March 2021.
S008	Associations between chronic conditions and death in hospital among adults (aged 20+ years) during first acute care hospitalizations with a confirmed or suspected COVID-19 diagnosis in Canada	We aimed to quantify life course-specific associations between death in hospital and 30 chronic conditions, and comorbidity among them, in adults (aged 20+ years) during their first acute care hospitalization with a confirmed or suspected COVID-19 diagnosis in Canada.
S009	Building influenza surveillance pyramids in near real time, Australia	Flutracking data for 2011 and 2012 were used to investigate whether a near real time severity measure for circulating influenza strains could be determined.
S010	CDC-Respiratory Virus Surveillance Among Children with Acute Respiratory Illnesses ? New Vaccine Surveillance Network, United States, 2016?2021	This report describes demographic characteristics of enrolled children who received care in these settings, and yearly circulation of influenza, RSV, HMPV, HPIV1?3, adenovirus, human rhinovirus and enterovirus (RV/EV),* and SARS-CoV-2 during December 2016?August 2021.
S011	COVID-19 and Severe Acute Respiratory Infections: Monitoring Trends in 421 German Hospitals During the First Four Pandemic Waves	In addition to statutory quality assurance, more than 500 hospitals in Germany and Switzerland are voluntarily organised in the ?Initiative of Quality Medicine? (IQM, Berlin, Germany).7 Utilizing a large-scale administrative dataset derived from the IQM, we conducted an extended analysis of SARI and COVID-19 cases covering four pandemic waves to report on patient characteristics and respective outcomes and compare pre-pandemic with pandemic time period

S012	Changes in Surveillance of Acute Respiratory Infections Including Influenza in Slovak Republic during 1993-2008	The authors evaluated surveillance of acute respiratory infections (ARI), influenza and influenza-like illnesses (ILI) in the Slovak Republic (SR). They analyze morbidity, age-specific morbidity, complications, mortality, number of influenza viruses isolations and vaccination coverage rates in the SR in the years 1993?2008.
S013	Characterisation of acute respiratory infections at a United Kingdom paediatric teaching hospital: observational study assessing the impact of influenza A (2009 pdmH1N1) on predominant viral pathogens	there is limited published data on which respiratory viral pathogens cause ARI in children in the UK. Our study has attempted to address this knowledge gap by describing viral pathogen prevalence, occurrence of co-infection, diagnostic yield of sampling methods and presence of co-morbidity in patients with ARI caused by 2009 pdmH1N1 and other respiratory viruses, in a large paediatric teaching hospital in the North West of England over a year between April 2010 and March 2011.
S014	Characterizing the Epidemiology of the 2009 Influenza A/H1N1 Pandemic in Mexico	Here, we analyze the age- and state-specific incidence of influenza morbidity and mortality in 32 Mexican States, on the basis of reports to the Mexican Institute for Social Security (IMSS), a private medical system that covers 40% of the Mexican population. We also quantify the association between local influenza transmission rates, school cycles, and demographic factors.
S015	Chronic use of inhaled corticosteroids in patients admitted for respiratory virus infections: a 6-year prospective multicenter study	We aimed to compare the characteristics and outcome of respiratory virus infections in adults hospitalized for influenza-like illness (ILI) with, or without, chronic use of ICS.
S016	Clinical Profile of COVID-19 Illness in Children?Experience from a Tertiary Care Hospital	The present study aims to analyze the clinical features and outcome in children infected with SARS-CoV-2 in tertiary care pediatric teaching hospital in Northern India..
S017	Clinical and epidemiological aspects of severe acute respiratory infection: before and during the first year of the COVID-19 pandemic in Brazil	Our objective was to describe cases and deaths from severe acute respiratory infection (SARI) in Brazil over the past 8 y as well as changes in the distribution and risk of illness and death from SARI before and in the first year of the coronavirus disease 2019 (COVID-19) pandemic (FYP).
S018	Clinical and phylogenetic influenza dynamics	We first aimed to report the epidemiological dynamics of the 2019/20 season at the GIHSN scale for the influenza-associated hospitalisation of

	for the 2019-20 season in the global influenza hospital surveillance network (GIHSN) - Pilot study	patients aged over 5 years. The second objective was to describe the phylogenetic characterization of the viruses responsible for these hospitalisations for the first time within our network, with a specific focus on the representativeness of these cases compared to other medically attended influenza surveillance networks.
S019	Clinical characteristics and factors associated with hospital admission or death in 43 103 adult outpatients with coronavirus disease 2019 managed with the Covidom telesurveillance solution: a prospective cohort study	Studies on coronavirus disease 2019 (COVID-19) have mainly focused on hospitalized patients or those with severe disease. We aim to assess the clinical characteristics, outcomes and factors associated with hospital admission or death in adult outpatients with COVID-19.
S020	Clinical characteristics and factors associated with severe acute respiratory infection and influenza among children in Jingzhou, China	We described the clinical and epidemiological characteristics of children with influenza and analyzed the association between potential risk factors and SARI patients with influenza.
S021	Clinical characteristics and outcome of respiratory syncytial virus infection among adults hospitalized with influenza-like illness in France	The aim of this study was to analyse characteristics and outcome of respiratory syncytial virus (RSV) infection in adults hospitalized with influenza-like illness (ILI).
S022	Clinical characteristics and outcomes of influenza and other influenza-like illnesses in Mexico City	Recognizing the need to investigate both influenza and non-influenza ILIs in the Mexican population, La Red implemented a study with the objective of describing the etiology, symptoms, and outcomes of subjects presenting with ILI in Mexico City.
S023	Clinical characteristics and outcomes of patients with severe acute respiratory infections (SARI):	This study describes the clinical features and outcomes of patients with severe acute respiratory infections (SARI) in hospitalized patients in Egypt.

	results from the Egyptian surveillance study 2010?2014	
S024	Clinical features and mortality in COVID-19 SARI versus non COVID-19 SARI cases from Western Rajasthan, India	The objective of the present study was to examine the clinical and laboratory as well as comorbidities and outcomes of the SARI patients admitted in the COVID-19 suspect ICU.
S025	Comparative epidemiology of five waves of COVID-19 in Mexico, March 2020?August 2022	In this study, we investigated the epidemiological patterns of COVID-19 infections, hospitalizations, deaths, and factors associated with severe disease outcomes during the five epidemic waves in Mexico. To that end, we used data reported by the Mexican Institute of Social Security (IMSS) surveillance system, the most extensive Latin-American social security system, and Mexico?s leading health institution [6].
S026	Comparison of common acute respiratory infection case definitions for identification of hospitalized influenza cases at a population-based surveillance site in Egypt	Multiple case definitions are used to identify hospitalized patients with community-acquired acute respiratory infections (ARI). We evaluated several commonly used hospitalized ARI case definitions to identify influenza cases.
S027	Coronavirus Surveillance in a Pediatric Population in Jordan From 2010 to 2013: A Prospective Viral Surveillance Study	In our prior viral surveillance studies in Amman Jordan, over 80% of young hospitalized children with fever and/or ARI had at least one respiratory virus identified; however, testing for the 4 common HCoVs was not performed.17 Furthermore, viral surveillance studies in the Middle East and North Africa region assessing the incidence, seasonality, and clinical characteristics of individual HCoV species are limited. Therefore, this study aimed to assess the clinical characteristics and distribution of HCoV infections in hospitalized children less than 2 years over 3 respiratory seasons.
S028	Deaths and hospitalizations related to 2009 pandemic influenza a (H1N1) - Greece, May 2009-February 2010	Editorial Note This is the first report to summarize the epidemiology of 2009 H1N1 in Greece. During July--August 2009, Greece experienced a moderate wave of transmission, followed by a stronger wave beginning in October and peaking during November 23--29.
S029	Description of Hospitalized Cases of Influenza A(H1N1)pdm09	The aim of the present study was toclarify the characteristics of hospitalized cases ofA(H1N1)pdm09 infection on the basis of analyses of

	Infection on theBasis of the National Hospitalized-Case Surveillance, 2009?2010, Japan	thesurveillance data collected during the 2009/2010 season.
S030	Descriptive Epidemiology of Novel Influenza A (H1N1), Andhra Pradesh 2009-2010	In this paper, we describe the descriptive epidemiology of A (H1N1) cases reported in Andhra Pradesh during May 2009 to December 2010
S031	Detection of Influenza C Viruses Among Outpatients and Patients Hospitalized for Severe Acute Respiratory Infection, Minnesota, 2013?2016	Thus, beginning in May 2013, the Minnesota Department of Health (MDH) with the support of the Centers for Disease Control and Prevention (CDC) incorporated molecular testing for influenza C into existing sentinel surveillance systems for outpatient and inpatient ARI, allowing us to study the epidemiology of influenza C virus infection.
S032	Detection of Viral and Bacterial Pathogens in Hospitalized Children With Acute Respiratory Illnesses, Chongqing, 2009?2013	In the present study, we aimed to describe the detection of viruses and bacteria in hospitalized children with ARI in a subtropical city of mainland China, investigate the simultaneous detection pattern of multiple viruses and bacteria, and assess their association with clinical outcomes and severity.
S033	Digitalizing and upgrading severe acute respiratory infections surveillance in Malta: system development	We sought to describe the process of digitizing and upgrading SARI surveillance in Malta, an island country with a centralized health system, during the COVID-19 pandemic from February to November 2021. We described the characteristics of people included in the surveillance system and compared different SARI case definitions, including their advantages and disadvantages. This study also discusses the process, output, and future for SARI and other public health surveillance opportunities.
S034	Disparities Between Black and White Children in Hospitalizations Associated With Acute Respiratory Illness and Laboratory-confirmed Influenza and Respiratory Syncytial Virus in 3 US Counties?2002?2009	The primary goal of this prospective, population-based study was to determine whether racial disparities existed between black and white children under 5 years of age for hospitalizations associated with community-acquired ARI and laboratory-confirmed seasonal influenza and RSV illness in 3 large US counties. Secondarily, we assessed racial differences in access to care, influenza vaccination, and other characteristics that may affect racial disparities in hospitalization rates.

S035	Effects of seasonal and pandemic influenza on health-related quality of life, work and school absence in England: Results from the Flu Watch cohort study	Objectives To measure quality-adjusted life days and years (QALDs and QALYs) lost and work/school absences among community cases of acute respiratory infections (ARI), ILI and influenza A and B and to estimate community burden of QALY loss and absences from influenza.
S036	Epidemiologic and spatiotemporal characterization of influenza and severe acute respiratory infection in Uganda, 2010-2015	To characterize the epidemiology and transmission dynamics of influenza and risk factors for influenza-associated severe respiratory infection in Uganda.
S037	Epidemiologic profile of severe acute respiratory infection in Brazil during the COVID-19 pandemic: an epidemiological study	we evaluated the features (demographic data,hospitalization information, and outcomes) of hospitalizedpatients with SARI in Brazil, during the COVID-19 pandemic,according to the following groups: SARI due to Influenza virusinfection, SARI due to other respiratory viruses? infection,SARIdue to other known etiologic agents (OEAs), SARI due toSARS-CoV-2 infection (patients with COVID-19), and SARI dueto an undefined etiological agent
S038	Epidemiological Characteristics and Underlying Risk Factors for Mortality during the Autumn 2009 Pandemic Wave in Mexico	In this article we fill this gap and carry out an analysis of clinical features at presentation, hospital admission delays, medical conditions, and receipt of seasonal vaccine on the risk of A/H1N1-related death among hospitalized patients. We use individual-level data from a prospective surveillance system implemented by the largest Mexican Social Security medical system spanning August-December, 2009.
S039	Epidemiological Characterization of a Fourth Wave of Pandemic A/H1N1 Influenza in Mexico, Winter 2011?2012: Age Shift and Severity	Here we report on the epidemiology of a recrudescent (fourth) wave of pandemic A/H1N1 influenza activity in Mexico from December 1, 2011?March 20, 2012 and update our preliminary findings based on data up to February 10, 2012 (10).
S040	Epidemiological Trends of Coronavirus Disease 2019 in Sierra Leone From March 2020 to October 2021	However, characterization of the epidemiological features of COVID-19 is crucial for the development and implementation of effective control strategies to reduce the socioeconomic effects of the COVID-19 pandemic. Here, we report the results of a descriptive, exploratory analysis of all of the cases diagnosed between March 2020 to October 2021 in Sierra Leone to better understand the epidemic's progression and to formulate targeted strategies to contain current and future viral outbreaks.

S041	Epidemiological and clinical profile of Influenza A(H1N1) pdm09 in Odisha, eastern India	We carried out a retrospective analysis of the available information, in order to study the clinico-epidemiological features and establish the magnitude and severity of recent Influenza A(H1N1)pdm09 epidemics in hospitalized patients from the state of Odisha.
S042	Epidemiology and outcome of influenza-associated infections among hospitalized patients with acute respiratory infections, Egypt national surveillance system, 2016-2019	This report presents the results of the national laboratory?based surveillance for hospitalized patients with ARI in 284 hospitals all over Egypt, 2016?2019. The study aims at describing the epidemiology and exploring severity and mortality of influenza?associated infections among hospitalized ARI patients to identify target groups for influenza prevention and control strategies.
S043	Epidemiology and virology of acute respiratory infections during the first year of life: a birth cohort study in Vietnam	Our study aimed to describe the epidemiology and viral etiology of ARI in the first year of life within an ongoing prospective infant cohort in southern Vietnam.
S044	Epidemiology of Influenza in Fars Province, Southern Iran; a Population-Based Study (2015-2019)	To investigate the clinical and epidemiological features of influenza virus A/H1N1, A/H3N2, and B infection in Fars province, southern Iran, in 2015-2019.
S045	Epidemiology of Severe Acute Respiratory Illness and Risk Factors for Influenza Infection and Clinical Severity among Adults in Malawi, 2011?2013	In this high HIV prevalence and malaria-endemic setting, we aimed to describe the epidemiology and viral etiology and factors associated with clinical severity and influenza positivity among individuals aged ? 15 years with severe acute respiratory illness (SARI) during 2011?2013
S046	Epidemiology of hospital admissions with influenza during the 2013/2014 Northern Hemisphere influenza season: results from the Global Influenza Hospital Surveillance Network	Here, we describe the epidemiology of hospital admissions with influenza during the 2013/2014 influenza season in the GIHSN Northern hemisphere participating sites. We also determine the impact of underlying patient characteristics on the risk of hospital admission and complications due to influenza overall and due to influenza A(H1N1)pdm09, A(H3N2), and B/Yamagata lineage.
S047	Epidemiology of influenza B infection	The aim of this study was to analyze cases of IBV infection amongcases of ARI and SARI notified

	in the state of Rio Grande do Sul, Brazil, from 2003 to 2019	between 2003 and 2018 in RioGrande do Sul, Southern Brazil, and analyze demographic and clinicaldata, as well as IBV circulation over the years.
S048	Epidemiology of severe acute respiratory illness (SARI) among adults and children aged 5 years in a high HIV-prevalence setting, 2009-2012	There are few published studies describing severe acute respiratory illness (SARI) epidemiology amongst older children and adults from high HIV-prevalence settings. We aimed to describe SARI epidemiology amongst individuals aged ?5 years in South Africa.
S049	Epidemiology, disease severity and outcome of Severe acute respiratory syndrome coronavirus 2 and influenza viruses coinfection seen at Egypt integrated acute respiratory infections surveillance, 2020-2022	This study aims to better describe the epidemiology, disease severity, and outcome of SARS-CoV-2/Flu coinfection to guide the development of effective preventive and control measures including case management and vaccination policy
S050	Establishing a sentinel surveillance system for the novel COVID-19 in a resource-limited country: methods, system attributes and early findings	To establish a hospital-based platform to explore the epidemiological and clinical characteristics of patients screened for COVID-19.
S051	Establishing an ICD-10 code based SARI-surveillance in Germany - description of the system and first results from five recent influenza seasons	We described the establishment of an ICD-10-based inpatient syndromic sentinel system and its application to the analysis of five influenza seasons. We compared the impact of different case definitions on the ability to capture SARI cases, to allow a timely trend analysis of the seasonal epidemic and to reflect the burden caused by influenza when compared to routine outpatient surveillance.
S052	Establishing thresholds and parameters for pandemic influenza severity assessment, Australia	To implement the World Health Organization?s pandemic influenza severity assessment tool in Australia, using multiple sources of data to establish thresholds and measure influenza severity indicators.
S052	Establishing thresholds and parameters for	To implement the World Health Organization?s pandemic influenza severity assessment tool in

	pandemic influenza severity assessment, Australia	Australia, using multiple sources of data to establish thresholds and measure influenza severity indicators.
S053	Estimating age-specific influenza-related hospitalization rates during the pandemic (H1N1) 2009 in Davidson Co, TN	Objectives: To estimate age-specific hospitalization rates associated with laboratory-confirmed A(H1N1)pdm09 virus in Davidson County, TN, from May 2009 to March 2010.
S054	Estimating severity of influenza epidemics from severe acute respiratory infections (SARI) in intensive care units	The most important complication of influenza virus infection is pneumonia (primary viral or secondary bacterial pneumonia) [9]. While costs can be high, the precise burden remains a blind spot [10?15]. Such burden information is crucial for prevention and response considering that vaccination, the main control measure against influenza infection, is aimed at preventing complications. A severe influenza season may also lead to hospital capacity problems, especially in ICUs. In this study we analyze comprehensive retrospective ICU data to fill the current knowledge gap.
S055	Estimating the burden of influenza-associated hospitalization and deaths in Oman (2012-2015)	To estimate the incidence of influenza- associated hospitalizations and in- hospital death in Oman
S056	Etiology, clinical characteristics, and risk factors associated with severe influenza-like illnesses in Mexican adults	Objective The aim of this study was to determine the risk factors associated with severe influenza-like illness (ILI) in Mexican adults that could be useful to clinicians when assessing patients with ILI.
S057	Evaluation of a new clinical endpoint for moderate to severe influenza disease in children: A prospective cohort study	Therefore, the objectives of this study were to evaluate if the proposed definition of moderate to severe influenza in children predicts hospitalization and other clinically relevant healthcare endpoints including recurrent ED visits, use of antimicrobials (including antivirals), school/daycare or parental work absenteeism, and increased healthcare costs during 2 recent influenza seasons.
S058	Facility-based surveillance for influenza and respiratory syncytial virus in rural Zambia	In December 2018, facility-based surveillance for influenza and RSV was established in rural Zambia to evaluate their role in causing respiratory illness and begin to situate rural Zambia in the landscape of regional and global virus transmission. The objective of this analysis was to describe the burden of influenza and RSV disease during the first year of surveillance and explore predictors of severe disease.

S059	Factors associated with Severe Acute Respiratory Syndrome in a Brazilian central region	This study aimed to analyze the epidemiological profile and the factors associated with hospitalization and deaths from SARI in a central Brazilian region from 2013 to 2018.
S060	Factors associated with poor outcomes in patients with severe acute respiratory infections in Bahrain	Therefore, this study was conducted with the main objective of identifying the risk factors associated with poor outcomes, including mortality, ICU admission, and mechanical ventilation among patients admitted with SARI in Bahrain.
S061	First-year results of the Global Influenza Hospital Surveillance Network: 2012-2013 northern hemisphere influenza season	In this report, we evaluated the characteristics of hospitalizations related to influenza and the temporal and geographic distribution of the different influenza viruses in these cases during the 2012?2013 Northern hemisphere influenza season, the program?s first year.
S062	Five-year community surveillance study for acute respiratory infections using text messaging: findings from the MoSAIC study	The objectives of this 5-year community surveillance study were to describe ARI incidence, etiology, and factors associated with infection and care-seeking, as well as to evaluate use of text messaging for longitudinal surveillance in a low-income population.
S063	Hospital-based surveillance of influenza A(H1N1)pdm09 virus in Saudi Arabia, 2010-2016	Describe the data generated by the influenza A(H1N1) pdm09 surveillance in Saudi Arabia from 2010 to 2016.
S064	Hospitalization of influenza-like illness patients recommended by general practitioners in France between 1997 and 2010	In the present study, we compare the PRH of the 2009? 2010 A(H1N1) pandemic with the twelve preceding sea- sonal epidemics, stratifying by age, sex, and viral subtype. We also investigate the reasons why GPs recommended hospitalization and the presence of risk factors for pan- demic A(H1N1) complications. We finally evaluate the use- fulness of this surveillance for public health information by monitoring the precision of the PRH estimate throughout the pandemic.

S065	Impact of 2009 pandemic influenza among Vietnamese children based on a population-based prospective surveillance from 2007 to 2011	To investigate the impact of A(H1N1)pdm09 on pediatric ARI in Vietnam.
S066	Impact of influenza infection on children's hospital admissions during two seasons in Athens, Greece	The aim of the present study was to prospectively evaluate the burden of influenza infection upon pediatric hospitalizations in our area, and to obtain data on the most common clinical manifesta- tions and possible complications.
S067	Impact of influenza on outpatient visits, hospitalizations, and deaths by using a time series Poisson generalized additive model	In this study, the time series Poisson generalized additive model (GAM) was used to quantitatively assess the disease burden of influenza and ILI, by using the influenza surveillance data in Zhuhai City from 2007 to 2009 combined with outpatient, inpatient, and respiratory disease mortality data from the same period.
S068	Implementing the World Health Organization Pandemic Influenza Severity Assessment framework-Singapore's experience	In this paper, we document Singapore's experience in developing and evaluating the PISA indicators and parameters, and this would provide other countries with suggestions that they can use in developing their own indicators.
S069	Incidence, disease severity, and follow-up of influenza A/A, A/B, and B/B virus dual infections in children: a hospital-based digital surveillance program	Within the framework of this inception cohort, we leveraged uniform case classi-fication of ILI and respiratory infection, standardized longitudinal patient assessments,and a comparable disease severity measure in addition to comprehensive IV PCR testingto investigate ?The incidence of IV dual infections in children based on a hospital-based digitalsurveillance program with a known denominator of all ILI cases and ?Disease severity, symptoms, and detailed course of illness in children with IVdual infections
S070	Influenza B associated paediatric acute respiratory infection hospitalization in central Vietnam	In this study, we investigated the incidence and clinical?epidemiological characteristics of paediatric hospitalized influenza B ARI cases in Vietnam.
S071	Influenza hospitalization epidemiology from a	We aim to describe the epidemiology and seasonality of influenza hospitalizations in Jordan over a time period of 6ÿyears (2008?2014).

	severe acute respiratory infection surveillance system in Jordan, January 2008-February 2014	
S072	Lessons from the epidemiological surveillance program, during the influenza a (H1N1) virus epidemic, in a reference university hospital of southeastern Brazil	to evaluate the impact of two definitions used as epidemiological tools, in adults and children, during the influenza A H1N1 epidemic.
S073	Lethality and characteristics of deaths due to COVID-19 in Rondonia: an observational study	Objective: To describe the characteristics of deaths due to COVID-19 in the state of Rond“nia.
S074	Morbidity, mortality, and seasonality of influenza hospitalizations in Egypt, November 2007-November 2014	The aims of this study were to (1) assess the proportion of SARI cases having influenza infection in Egypt; (2) examine the types and subtypes of detected influenza viruses in Egypt; (3) compare demographic and clinical characteristics of influenza-positive SARI cases to those of influenza-negative SARI cases in Egypt; (4) quantify influenza deaths and assess influenza mortality risk factors in Egypt; and (5) establish a defined period of influenza seasonality in Egypt.
S075	Mortality amongst patients with influenza-associated severe acute respiratory illness, South Africa, 2009-2013	We aimed to estimate the incidence of influenza-associated severe acute respiratory illness (SARI) deaths and describe the risk-factors associated with death using data from prospective, hospital-based sentinel surveillance in South Africa.
S076	Mortality, Severe Acute Respiratory Infection, and Influenza-Like Illness Associated with Influenza A(H1N1)pdm09 in Argentina, 2009	In order to explore whether Argentina's influenza A(H1N1)pdm09 burden was higher or similar to the burden documented elsewhere, we use active facility-based influenza surveillance and health utilization surveys from three cities in Argentina to estimate rates of influenza A(H1N1)pdm09-associated mortality, hospitalization, and influenza-like illnesses.
S077	National retrospective cohort study to identify age-specific fatality risks of comorbidities	Objectives This study aimed to examine comprehensively the prognostic impact of underlying comorbidities among hospitalised patients with influenza-like illness (ILI) in

	among hospitalised patients with influenza-like illness in Taiwan	different age groups and provide recommendations targeting the vulnerable patients
S078	Non-influenza respiratory viruses in adult patients admitted with influenza-like illness: a 3-year prospective multicenter study	To describe the burden, and characteristics, of influenza-like illness (ILI) associated with non-influenza respiratory viruses (NIRV).
S079	Novel influenza A(H1N1) in a pediatric health care facility in New York city during the first wave of the 2009 pandemic	Objective To describe the burden of care experienced by our pediatric health care facility in New York, New York, from May 3, 2009, to July 31, 2009, during the novel influenza A(H1N1) pandemic that began in spring 2009.
S080	Occurrence of AH1N1 viral infection and clinical features in symptomatic patients who received medical care during the 2009 influenza pandemic in Central Mexico	This study estimated the AH1N1 infection, hospitalization and mortality rates in SLP during the 2009 pandemic, and aimed at identifying clinical features associated with AH1N1 infection in individuals with flu-like illness who sought medical care.
S081	Outbreak of 2009 pandemic influenza A(H1N1), Los Lagos, Chile, April-June 2009	The aim of this study, as outlined in the introduction, was to investigate an outbreak of the 2009 pandemic influenza A(H1N1) in the Los Lagos region of Chile. Key objectives of the study included:
S082	Outcomes of influenza A(H1N1)pdm09 virus infection: results from two international cohort studies	In this report, we describe outcomes of outpatients and hospitalized patients with influenza A(H1N1)pdm09 virus infection and examine risk factors for progression of their illness.
S083	Outcomes of patients with Severe Acute Respiratory Infections (SARI) admitted to the intensive care unit: results from the Egyptian	We aimed to investigate the role of different respiratory viruses in causing critical illness requiring ICU admission, which pathogens were related to severe outcomes, and to address the impacts of SARI on the clinical outcomes of patients admitted to the ICU, in terms of morbidity and mortality.

	Surveillance Study 2010-2014	
S084	Overview of the winter wave of 2009 pandemic influenza A(H1N1)v in Vojvodina, Serbia	To analyze the epidemiological data for pandemic influenza A(H1N1)v in the Autonomous Province of Vojvodina, Serbia, during the season of 2009/2010 and to assess whether including severe acute respiratory illness (SARI) hospitalization data to the surveillance system gives a more complete picture of the impact of influenza during the pandemic.
S085	Pandemic H1N1 influenza in Brazil: Analysis of the first 34,506 notified cases of influenza-like illness with severe acute respiratory infection (SARI)	The present paper describes the epidemiological profile of influenza-like illness (ILI) with severe acute respiratory infection (SARI), occurred during EW16 to 33 in Brazil. Case-fatality by sex and presence of comorbidity is also presented.
S086	Patient characteristics associated with COVID-19 positivity and fatality in Nigeria: retrospective cohort study	Despite the increasing disease burden, there is a dearth of context-specific evidence on the risk factors for COVID-19 positivity and subsequent death in Nigeria. Thus, the study objective was to identify context-specific factors associated with testing positive for COVID-19 and fatality in Nigeria.
S087	Persistent Functional Decline Following Hospitalization with Influenza or Acute Respiratory Illness	We aimed to investigate persistent functional change in older adults admitted to hospital with influenza and other acute respiratory illness (ARI).
S088	Prevalence and risk factors for long COVID after mild disease: A cohort study with a symptomatic control group	There is limited data on the prevalence and risk factors for long COVID and few prospective studies with appropriate control groups and adequate sample sizes. We performed a prospective study to determine the prevalence and risk factors for long COVID.
S089	Prevalence, incidence, and severity associated with viral respiratory tract infections in Colombian adults before the COVID-19 pandemic	We hypothesize that the most frequent clinical diagnosis of patients admitted due to ARI is CAP. The severity classification used by the surveillance program might identify patients at higher risk of worse clinical outcomes. Using the ARI-based report strategy in Bogot (Colombia), we will attempt to bridge this gap in the literature by identifying the clinical burden of respiratory viral infections, the incidence per year, disease severity, and clinical outcomes of ARI in adults.
S090	Recrudescent wave of pandemic A/H1N1	Here we describe changes in the epidemiological patterns of the ongoing 4th pandemic wave in 2011-

	influenza in Mexico, winter 2011-2012: Age shift and severity	12, relative to the earlier waves in 2009. The analysis is intended to guide public health intervention strategies in near real time.
S091	Respiratory syncytial virus and other respiratory viral infections in older adults with moderate to severe influenza-like illness	We used multiplex reverse transcriptase PCR (RT-PCR) to identify viral respiratory pathogens in nasal and throat swabs from episodes of moderate-to-severe influenza-like illness (ILI) in influenza-vaccinated elderly individuals.
S092	Respiratory virus-associated severe acute respiratory illness and viral clustering in Malawian children in a setting with a high prevalence of HIV infection, malaria, and malnutrition	In the context of a low-income population with multiple drivers of immune compromise (eg, human immunodeficiency virus [HIV] infection, malnutrition, and malaria) [11], we conducted active surveillance at a large urban teaching hospital in Malawi to estimate the incidence of childhood SARI and explore the association of SARI clinical severity with HIV infection and clustering of respiratory viral coinfection.
S093	Risk Factors for Severe Coronavirus Disease 2019 Among Human Immunodeficiency Virus-Infected and -Uninfected Individuals in South Africa, April 2020-March 2022: Data From Sentinel Surveillance	Using a well established syndromic surveillance program for influenza-like-illness (ILI) [11] and severe respiratory illness (SRI) [12?14], we aimed to describe clinical and epidemiological characteristics of persons with laboratory-confirmed COVID-19 and identify factors associated with COVID-19 hospitalization or mortality.
S094	Risk factors for influenza-associated severe acute respiratory illness hospitalization in South Africa, 2012-2015	Risk factors for influenza-associated severe acute respiratory illness hospitalization in South Africa, 2012-2015
S095	Risk factors of prolonged hospital stay in children with viral severe acute respiratory infections	The present study focused on detection of risk factors for prolonged hospital stay among children with viral SARI, including demographic and clinical characteristics of patients, and the type and seasonality of different respiratory viral pathogens causing acute lower respiratory infection (ALRI).
S096	SARS-CoV-2 and influenza virus coinfection among patients with severe acute respiratory	To estimate the proportion of SARS-CoV-2 and influenza virus coinfection among severe acute respiratory infection (SARI) cases-patients during the first wave of COVID-19 pandemic in Bangladesh.

	infection during the first wave of COVID-19 pandemic in Bangladesh: a hospital-based descriptive study	
S097	Sentiworld-2022 ACUTE RESPIRATORY ILLNESS SURVEILLANCE REPORT	This report provides an overview of priority viral respiratory illnesses in New Zealand in 2022, including those causing influenza, respiratory syncytial virus (RSV) illness and COVID-19. Please note, most viral respiratory illnesses are not legally notifiable diseases in New Zealand.
S098	Sentiworld-Weekly Report on Severe Acute Respiratory Infection (SARI), Week 20 2023 (week ending 21/05/2023)	This report includes data on SARI hospitalised cases, aged 15 years and older who were admitted to St. Vincent?s University Hospital (SVUH), Dublin up to week 20 2023.
S099	Severe Acute Respiratory Infection (SARI) sentinel surveillance in the country of Georgia, 2015-2017	This study aimed to characterize the seasonality and epidemiology of SARI in the country of Georgia over two influenza seasons (2015?2016 and 2016?2017), to describe the etiological and clinical patterns observed, and to assess seasonal influenza vaccine effectiveness using a case-test negative design.
S100	Severe Acute Respiratory Infections With Influenza and Noninfluenza Respiratory Viruses: Yemen, 2011-2016	This study aimed to determine the proportions of influenza and noninfluenza virus among SARI patients, and assess the severity of SARI and its associated factors in Yemen.
S101	Severe Illnesses Associated With Outbreaks of Respiratory Syncytial Virus and Influenza in Adults	Our purpose of the present study was to perform an ecological analysis of the relationship between outbreaks of RSV and flu and advanced medical outcomes of adults in a defined geographic region over 12 consecutive years (2001 to 2013).
S102	Severe acute respiratory illness deaths in sub-Saharan Africa and the role of influenza: a case series from 8 countries	Hospital-based influenza surveillance has increased dramatically in Africa in the last decade [9], and we assessed whether the current systems could provide insights into seasonal influenza?associated mortality among persons hospitalized for respiratory disease in Africa, as well as data on other etiologies of respiratory disease?associated mortality in sub-Saharan Africa.
S103	Severe acute respiratory illness surveillance for influenza in Kenya: Patient	This paper aims to describe the epidemiology and clinical features of patients hospitalized with influenza in Kenya and highlight the importance of year?round surveillance for severe acute respiratory infections (SARI), especially in tropical countries.

	characteristics and lessons learnt	
S104	Severe acute respiratory infection in children in a densely populated urban slum in Kenya, 2007-2011	To broaden the knowledge-base and compare etiology and epidemiology, we analyzed data from our population-based infectious disease surveillance (PBIDS) site in Kibera, an urban slum in Nairobi. The rural and urban PBIDS operate with the same study protocol.
S105	Severe influenza-associated respiratory infection in high HIV prevalence setting, South Africa, 2009-2011	We investigated the incidence of hospitalization for influenza-associated acute lower respiratory tract infection (LRTI) and the clinical course of illness in persons with and without HIV infection in South Africa.
S106	Severity and mortality of COVID-19 infection in HIV-infected individuals: Preliminary findings from Iran	This study aims to investigate the course of COVID-19 infection in HIV-infected individuals by characterizing COVID-19 incidence, clinical presentation, severity, and mortality in HIV-infected patients as compared to HIV-negative COVID cases in Iran. Since the study has been performed in the early stage of the COVID-19 pandemic, few similar evidences have been published in this regard, especially in the study area indicating the novelty of the study.
S107	Single- and multiple viral respiratory infections in children: disease and management cannot be related to a specific pathogen	The aim of the current study was to determine if RT-PCR test results are related to clinical data in children with respiratory symptoms. We investigated clinical symptoms, management and outcome in these children and correlated these findings to the specific virus determined by RT-PCR. We additionally investigated clinical differences between single-, multiple-, and RT-PCR negative ARI.
S108	Status of novel coronavirus disease and analysis of mortality in Mexico, until June 30th, 2020: An ecological study	The aim was to analyze the Cause-Specific Mortality Rate (CSMR) for COVID-19, for each Mexican State and the effect of comorbidities on deaths by COVID-19.
S109	Substantial Morbidity and Mortality Associated with Pandemic A/H1N1 Influenza in Mexico, Winter 2013-2014: Gradual Age Shift and Severity	Here we report preliminary findings on the epidemiology of the on-going A/H1N1 outbreak in Mexico from October 2013 to January 2014. Because past influenza pandemics have had substantial morbidity and mortality burden for several seasons after the initial pandemic waves 6 7 8 9 10 , continued vigilance is prudent. We compare the demographic and clinical characteristics of laboratory-confirmed A/H1N1 hospitalizations and deaths in winter 2013-14 with those reported for the preceding 2011-12 A/H1N1 epidemic. Our data highlight a change in the age distribution of A/H1N1 patients and a slightly

		higher reproduction number compared to the 2011-12 A/H1N1 outbreak.
S110	Surveillance data for eight consecutive influenza seasons in Sicily, Italy	This retrospective study aimed to explore the epidemiology of influenza disease and the heterogeneity of circulating strains in Sicily, at the primary care and hospital level, over eight influenza seasons after the 2009 pandemic.
S111	Surveillance for hospitalized acute respiratory infection in Guatemala	The International Emerging Infections Program of the U.S. Centers for Disease Control and Prevention (CDC), in collaboration with the Guatemala Ministry of Public Health and Welfare and the Universidad del Valle de Guatemala (UVG) conducts surveillance for hospitalized ARI in two sites in Guatemala. The surveillance is aimed at measuring the burden of hospitalized ARI in the catchment area and characterizing ARI etiology. We present the findings of surveillance for hospitalized ARI from November 2007 through December 2011.
S112	Surveillance for severe acute respiratory infections (SARI) in hospitals in the WHO European region - an exploratory analysis of risk factors for a severe outcome in influenza-positive SARI cases	This paper describes the characteristics of SARI patients and investigates risk factors for a severe outcome (ICU/fatal) in influenza-positive SARI patients in countries in Central and Eastern Europe.
S113	Surveillance of hospitalised patients with influenza-like illness during pandemic influenza A(H1N1) season in Sicily, April 2009-december 2010	The aim of the present study was to report the influenza surveillance data describing the epidemiological characteristics of patients with ILI symptoms, laboratory-confirmed infections with pandemic influenza A(H1N1)2009, and fatal cases that occurred among hospitalised patients in Sicily from April 2009 through December 2010.
S114	Surveillance of influenza in Iceland during the 2009 pandemic	In this article we report the changes made in the surveillance of influenza in Iceland and describe the data collected during the 2009 H1N1pandemic.
S115	The Burden of Influenza-Associated Hospitalizations in Oman, January 2008-June 2013	This report describes the establishment of a SARI surveillance system in Oman and presents the epidemiology and seasonality of influenza during the period of January 2008 to June 2013.
S116	The Epidemiology and Burden of Influenza B/Victoria and B/Yamagata	Here, we describe the epidemiology and clinical presentation associated with influenza B virus lineages (B/Victoria and B/Yamagata) among

	Lineages in Kenya, 2012-2016	medically attended cases of acute respiratory illness (ARI) in Kenya.
S117	The Intensive Care Global Study on Severe Acute Respiratory Infection (IC-GLOSSARI): a multicenter, multinational, 14-day inception cohort study	In this prospective, multicenter, 14-day inception cohort study, we investigated the epidemiology, patterns of infections, and outcome in patients admitted to the intensive care unit (ICU) as a result of severe acute respiratory infections (SARIs).
S118	The emergence of novel SARS-CoV-2 variant P.1 in Amazonas (Brazil) was temporally associated with a change in the age and sex profile of COVID-19 mortality: A population based ecological study	With the objective to describe and to identify possible changes in the mortality profile associated temporally to the emergence of the P1 strain in the state of Amazonas, we used public data of COVID-19 cases registered at the national epidemiological surveillance system. Two distinct epidemiological periods were considered in our analysis: the peak of the first wave, between April and May 2020, and January 2021 (the second wave), the month in which the new variant came to predominate.
S119	The epidemiology and severity of respiratory viral infections in a tropical country: Ecuador, 2009-2016	Respiratory viral infections (RVI) are a leading cause of mortality worldwide. We compared the epidemiology and severity of RVI in Ecuador during 2009?2016.
S120	The impact of altitude on hospitalization and hospital mortality from pandemic 2009 influenza A (H1N1) virus pneumonia in Mexico	In this study, our objective was to examine the association between altitude of residence and rates of hospitaliza-tion and mortality, in cases of Influenza?like illness (ILI) and severe acute respiratory illness (SARI), during the first months of the 2009 pandemic influenza A H1N1 virus, to examine if residents of high altitude had more frequently these adverse outcomes.
S121	The impact of pandemic influenza H1N1 on health-related quality of life: A prospective population-based study	The health-related quality of life detriment from a population-based sample of confirmed H1N1v patients was prospectively measured and compared to controls who were investigated because they had influenza like illness (ILI), but were not laboratory confirmed as H1N1v. The aims were: 1) to quantify the burden of H1N1v for individual patients and investigate factors, such as age and treatment with antivirals, that may affect this; 2) compare the severity of the 2009 strain to other infections that cause ILI and previous estimates of the severity of influenza from a systematic literature review; and 3) to estimate the overall burden attributed to H1N1v in the population. The findings can then be used to inform effectiveness and cost-effectiveness analyses

		on policy decisions related to the control of future waves of this or related viruses.
S122	The substantial hospitalization burden of influenza in central China: surveillance for severe, acute respiratory infection, and influenza viruses, 2010-2012	The study aimed to conduct surveillance of Severe Acute Respiratory Infection (SARI) in central China and estimate the rates of hospitalization due to SARI that could be attributed to influenza virus infections. This involved analyzing the prevalence and impact of different influenza virus types/subtypes in hospitalized SARI patients.
S123	U-shaped-aggressiveness of SARS-CoV-2: Period between Initial Symptoms and Clinical Progression to COVID-19 Suspicion. A Population-Based Cohort Study	To determine the aggressiveness of SARS-CoV-2 by analyzing symptom progression in COVID-19 patients.
S124	Using research to prepare for outbreaks of severe acute respiratory infection	We present combined results from the first two seasons of data collection for this programme during 2016?2017, where the primary circu-lating respiratory viruses were influenza A (H3N2) and A(H1N1)pdm09.
S125	Using routine emergency department data for syndromic surveillance of acute respiratory illness, Germany, week 10 2017 until week 10 2021	In this work, we aimed to describe emergency-department attendances for acute respiratory illness in Germany over time, for the purpose of developing and implementing syndromic surveillance.
S126	Viral etiology, seasonality and severity of hospitalized patients with severe acute respiratory infections in the Eastern Mediterranean Region, 2007-2014	Describe viral aetiology of ARI in Eastern Med 2007 - 14

A2.4.2 Publication Details

Details of publication type, publication date, and study dates.

DETAILS study_id	DETAILS publication_type	DETAILS publication_date	DETAILS study_period start_date	DETAILS study_period end_date
S001	Peer-reviewed article	30/12/2010	01/01/2008	31/12/2009
S002	Peer-reviewed article	11/10/2021	01/04/2020	31/12/2020
S003	Peer-reviewed article	17/04/2023	01/04/2020	31/08/2020
S004	Peer-reviewed article	01/05/2014	01/10/2007	31/05/2011
S005	Peer-reviewed article	28/01/2010	05/07/2009	04/10/2009
S006	Peer-reviewed article	22/11/2016	01/10/2011	07/05/2014
S007	Peer-reviewed article	28/03/2022	01/03/2021	31/03/2022
S008	Peer-reviewed article	01/04/2023	01/09/2020	31/03/2021
S009	Peer-reviewed article	11/11/2013	01/01/2011	01/01/2012
S010	Surveillance report	07/10/2022	01/12/2016	31/08/2021
S011	Peer-reviewed article	08/05/2023	01/01/2019	31/12/2021
S012	Peer-reviewed article	01/03/2011	01/01/1993	01/01/2008
S013	Peer-reviewed article	19/06/2014	01/04/2010	31/03/2011
S014	Peer-reviewed article	24/05/2011	01/04/2009	31/12/2009
S015	Peer-reviewed article	10/03/2022	01/12/2012	31/03/2018
S016	Peer-reviewed article	27/07/2021	01/04/2020	31/10/2020
S017	Peer-reviewed article	05/08/2022	01/01/2013	31/12/2020
S018	Peer-reviewed article	14/05/2022	02/09/2019	23/05/2020
S019	Peer-reviewed article	27/04/2021	09/03/2020	11/08/2020
S020	Peer-reviewed article	20/09/2016	05/04/2010	08/04/2012
S021	Peer-reviewed article	27/11/2016	01/11/2012	31/03/2015
S022	Peer-reviewed article	14/02/2013	11/04/2010	10/04/2011
S023	Peer-reviewed article	01/04/2019	01/02/2010	01/02/2014
S024	Peer-reviewed article	30/09/2021	07/03/2020	20/07/2020
S025	Peer-reviewed article	31/10/2022	29/03/2020	27/08/2022
S026	Peer-reviewed article	25/03/2021	01/06/2008	31/12/2012
S027	Peer-reviewed article	01/01/2021	01/03/2010	31/03/2013
S028	Surveillance report	11/06/2010	18/05/2009	28/02/2010
S029	Peer-reviewed article	24/12/2014	24/07/2009	05/09/2010
S030	Peer-reviewed article	01/07/2013	01/05/2009	31/12/2010
S031	Peer-reviewed article	23/10/2017	01/05/2013	31/12/2016
S032	Peer-reviewed article	24/04/2015	01/06/2009	31/10/2013
S033	Peer-reviewed article	05/12/2022	01/02/2021	30/11/2021
S034	Peer-reviewed article	22/02/2013	01/10/2002	31/05/2009
S035	Peer-reviewed article	09/10/2017	01/10/2006	31/12/2011
S036	Peer-reviewed article	01/12/2016	01/10/2010	01/06/2015
S037	Peer-reviewed article	01/07/2022	29/12/2019	31/12/2020
S038	Peer-reviewed article	16/07/2012	01/08/2009	31/12/2009
S039	Peer-reviewed article	16/10/2012	01/12/2011	10/02/2012
S040	Peer-reviewed article	29/06/2022	01/03/2020	31/10/2021
S041	Peer-reviewed article	15/10/2019	01/01/2009	31/12/2017
S042	Peer-reviewed article	07/05/2021	01/01/2016	31/12/2019
S043	Peer-reviewed article	24/03/2015	01/07/2009	31/12/2013
S044	Peer-reviewed article	01/03/2021	22/12/2015	22/09/2019

S045	Peer-reviewed article	23/07/2018	01/01/2011	31/12/2013
S046	Peer-reviewed article	19/05/2016	01/12/2013	30/06/2014
S047	Peer-reviewed article	26/01/2021	01/01/2003	31/12/2019
S048	Peer-reviewed article	23/02/2015	01/02/2009	31/12/2012
S049	Peer-reviewed article	17/11/2022	01/01/2020	30/04/2022
S050	Peer-reviewed article	02/12/2021	10/06/2020	31/08/2020
S051	Peer-reviewed article	30/06/2017	01/01/2012	15/05/2016
S052	Peer-reviewed article	25/06/2018	01/01/2012	31/12/2017
S053	Peer-reviewed article	23/02/2012	01/05/2009	31/03/2010
S054	Peer-reviewed article	19/12/2018	01/01/2007	31/12/2016
S055	Peer-reviewed article	05/12/2017	01/01/2012	31/12/2015
S056	Peer-reviewed article	01/02/2023	11/04/2010	10/04/2014
S057	Peer-reviewed article	14/11/2019	01/01/2017	30/04/2018
S058	Peer-reviewed article	21/09/2021	10/12/2018	09/12/2019
S059	Peer-reviewed article	12/08/2020	01/01/2013	01/01/2018
S060	Peer-reviewed article	26/04/2023	01/01/2018	31/12/2021
S061	Peer-reviewed article	05/06/2014	01/01/2013	31/05/2013
S062	Peer-reviewed article	17/01/2022	01/10/2013	30/09/2017
S063	Peer-reviewed article	06/02/2020	01/01/2010	31/12/2016
S064	Peer-reviewed article	22/03/2012	01/10/1997	31/03/2010
S065	Peer-reviewed article	07/03/2014	01/02/2007	30/03/2011
S066	Peer-reviewed article	18/12/2010	01/12/2002	31/01/2005
S067	Peer-reviewed article	19/02/2016	01/01/2007	31/12/2011
S068	Peer-reviewed article	17/10/2019	01/01/2007	31/12/2017
S069	Peer-reviewed article	14/03/2022	01/12/2009	30/04/2015
S070	Peer-reviewed article	28/02/2019	01/02/2007	30/06/2013
S071	Peer-reviewed article	29/01/2016	01/01/2008	01/02/2014
S072	Peer-reviewed article	22/07/2011	28/04/2009	31/12/2009
S073	Peer-reviewed article	01/12/2022	01/01/2020	20/08/2020
S074	Peer-reviewed article	08/09/2016	01/11/2007	01/11/2014
S075	Peer-reviewed article	18/03/2015	01/02/2009	31/12/2013
S076	Peer-reviewed article	31/10/2012	01/04/2009	31/12/2009
S077	Peer-reviewed article	24/06/2019	01/01/2005	31/12/2010
S078	Peer-reviewed article	13/02/2020	01/12/2012	31/03/2015
S079	Peer-reviewed article	04/01/2010	03/05/2009	31/07/2009
S080	Peer-reviewed article	20/12/2012	15/03/2009	30/10/2009
S081	Surveillance report	07/01/2010	01/04/2008	31/05/2009
S082	Peer-reviewed article	08/07/2014	01/10/2009	31/12/2009
S083	Peer-reviewed article	09/06/2020	01/02/2012	01/02/2014
S084	Peer-reviewed article	01/04/2011	01/06/2009	30/09/2010
S085	Surveillance report	22/10/2009	19/04/2009	21/08/2009
S086	Peer-reviewed article	17/12/2020	27/02/2020	08/06/2020
S087	Peer-reviewed article	08/12/2020	01/10/2011	31/03/2012
S088	Peer-reviewed article	12/05/2023	01/09/2020	30/04/2021
S089	Peer-reviewed article	01/12/2022	01/02/2013	31/08/2019
S090	Peer-reviewed article	26/03/2012	01/12/2011	10/02/2012
S091	Peer-reviewed article	29/01/2014	15/11/2008	30/04/2010

S092	Peer-reviewed article	13/09/2016	01/01/2011	31/12/2014
S093	Peer-reviewed article	02/11/2022	01/04/2020	31/03/2022
S094	Peer-reviewed article	10/02/2017	01/05/2012	30/04/2015
S095	Peer-reviewed article	15/10/2014	01/02/2010	31/05/2011
S096	Peer-reviewed article	29/11/2021	01/03/2020	31/12/2020
S097	Surveillance report	24/07/2023	01/01/2022	31/12/2023
S098	Surveillance report	21/05/2023	15/05/2023	21/05/2023
S099	Peer-reviewed article	30/07/2018	29/09/2015	15/03/2017
S100	Peer-reviewed article	23/04/2019	01/01/2011	31/12/2016
S101	Peer-reviewed article	04/04/2019	01/07/2001	29/06/2013
S102	Peer-reviewed article	01/10/2019	01/01/2009	31/12/2012
S103	Peer-reviewed article	14/03/2022	01/01/2014	31/12/2018
S104	Peer-reviewed article	25/02/2015	01/03/2007	28/02/2011
S105	Peer-reviewed article	01/11/2013	01/02/2009	31/12/2011
S106	Peer-reviewed article	10/03/2021	19/02/2020	08/04/2020
S107	Peer-reviewed article	11/01/2017	01/11/2007	31/03/2009
S108	Peer-reviewed article	30/12/2020	13/01/2020	27/06/2020
S109	Peer-reviewed article	26/03/2014	01/10/2013	31/01/2014
S110	Peer-reviewed article	30/09/2019	01/10/2010	30/04/2018
S111	Peer-reviewed article	31/12/2013	01/11/2007	31/12/2011
S112	Peer-reviewed article	08/01/2015	01/01/2009	31/12/2012
S113	Surveillance report	01/09/2011	01/04/2009	31/12/2010
S114	Surveillance report	09/12/2010	29/06/2009	27/12/2009
S115	Peer-reviewed article	07/12/2015	01/01/2008	30/06/2013
S116	Peer-reviewed article	30/09/2019	01/01/2012	31/12/2016
S117	Peer-reviewed article	15/02/2016	03/11/2013	26/01/2014
S118	Peer-reviewed article	18/07/2021	01/04/2020	31/01/2021
S119	Peer-reviewed article	17/12/2018	01/05/2009	31/12/2016
S120	Peer-reviewed article	01/01/2013	01/06/2009	31/10/2009
S121	Peer-reviewed article	02/03/2011	05/07/2009	18/07/2009
S122	Peer-reviewed article	10/11/2013	05/04/2010	08/04/2012
S123	Peer-reviewed article	03/12/2020	01/01/2020	30/04/2020
S124	Peer-reviewed article	13/02/2019	01/01/2016	31/08/2017
S125	Peer-reviewed article	07/07/2022	06/03/2017	13/03/2021
S126	Peer-reviewed article	13/07/2017	01/02/2007	28/02/2014

A2.4.3 Geographical Scope and Case Types

Study details for geographical scope, case types, and recruitment type. As 13 studies reported severity markers for 2 case types, the total reporting instances is 139 (126 + 13).

Reporting instance	DETAILS study_id	DETAILS geographical scope	DETAILS which_country	POPULATION case_type	RECRUITMENT type
1	S001	Regional	Guatemala	ili	treatment_seeking
2	S001	Regional	Guatemala	Severe pneumonia	hospitalised
3	S002	Regional	Mexico	sus_covid	treatment_seeking
4	S003	Regional	India	sus_covid	treatment_seeking
5	S004	Regional	USA	ari	treatment_seeking hospitalised
6	S005	National	La Reunion island	ari	treatment_seeking hospitalised
7	S006	National	Korea	ari	treatment_seeking
8	S007	National	Sierra Leone	sus_covid	hospitalised
9	S008	National	Canada	sus_covid	hospitalised
10	S009	National	Australia	ili	non_treatment_seeking
11	S010	National	USA	ari	treatment_seeking hospitalised
12	S011	National	Germany	sari_sus_covid	hospitalised
13	S012	National	Slovak Republic	ari_ili	treatment_seeking
14	S013	Regional	UK	ari	hospitalised
15	S014	National	Mexico	ili	treatment_seeking hospitalised
16	S015	National	France	ili	hospitalised
17	S016	Regional	India	sus_covid	non_treatment_seeking
18	S017	National	Brazil	sari	hospitalised
19	S018	Multinational	Brazil, Canada, China, France, India, Ivory Coast, Kenya, Lebanon, Mexico, Nepal, Peru, Romania, Russia, Serbia, Spain, Turkey, Ukraine	ili	hospitalised
20	S019	Regional	France	sus_covid	treatment_seeking
21	S020	Regional	China	sari	hospitalised
22	S021	National	France	ili	hospitalised
23	S022	Regional	Mexico	ili	treatment_seeking
24	S023	Regional	Egypt	sari	hospitalised
25	S024	Regional	India	sari	hospitalised
26	S025	National	Mexico	sus_covid	treatment_seeking hospitalised
27	S026	Regional	Egypt	ari	hospitalised
28	S027	Regional	Jordan	ari	hospitalised
29	S028	National	Greece	ili	treatment_seeking
30	S029	National	Japan	ili	treatment_seeking
31	S029	National	Japan	ili	hospitalised

32	S030	Regional	India	ili	treatment_seeking hospitalised
33	S031	Regional	USA	ari_ili	treatment_seeking
34	S031	Regional	USA	sari	hospitalised
35	S032	Regional	China	ari	hospitalised
36	S033	National	Malta	sari	hospitalised
37	S034	Regional	USA	ari	hospitalised
38	S035	National	England	ari	non_treatment_seeking
39	S035	National	England	ili	non_treatment_seeking
40	S036	National	Uganda	ili	treatment_seeking
41	S036	National	Uganda	sari	hospitalised
42	S037	National	Brazil	sari	hospitalised
43	S038	National	Mexico	ari	hospitalised
44	S039	National	Mexico	sari	hospitalised
45	S040	National	Sierra Leone	sus_covid	unknown
46	S041	Regional	India	ili	hospitalised
47	S042	National	Egypt	sari	hospitalised
48	S043	Regional	Vietnam	sari	hospitalised
49	S044	Regional	Iran	ili	treatment_seeking
50	S044	Regional	Iran	sari	hospitalised
51	S045	Regional	Malawi	sari	treatment_seeking
52	S046	Multinational	Russia, Turkey, China, Spain	ili	hospitalised
53	S047	Regional	Brazil	ari	treatment_seeking
54	S047	Regional	Brazil	sari	treatment_seeking
55	S048	National	South Africa	sari	hospitalised
56	S049	National	Egypt	ili_sari	treatment_seeking hospitalised
57	S050	National	Bangladesh	sus_covid	treatment_seeking hospitalised
58	S051	National	Germany	sari	hospitalised
59	S052	National	Australia	ili	treatment_seeking
60	S052	National	Australia	ili	non_treatment_seeking
61	S053	Regional	USA	ari	hospitalised
62	S054	National	Netherlands	sari	icu
63	S055	National	Oman	sari	hospitalised
64	S056	Regional	Mexico	ili	treatment_seeking
65	S057	Regional	USA	ili	treatment_seeking
66	S058	Regional	Zambia	ili	treatment_seeking
67	S059	Regional	Brazil	sari	treatment_seeking
68	S060	National	Bahrain	sari	hospitalised
69	S061	Multinational	Russia, Turkey, France, Spain	ili	hospitalised
70	S062	Regional	USA	ari	non_treatment_seeking
71	S063	National	Saudi Arabia	ili	hospitalised
72	S064	National	France	ili	treatment_seeking
73	S065	Regional	Vietnam	ari	hospitalised
74	S066	Regional	Greece	ari	hospitalised

75	S067	Regional	China	ili	treatment_seeking
76	S068	National	Singapore	ili	treatment_seeking
77	S069	Regional	Germany	ili	treatment_seeking
78	S070	Regional	Vietnam	ari	hospitalised
79	S071	National	Jordan	sari	hospitalised
80	S072	Regional	Brazil	ari_ili	treatment_seeking
81	S073	Regional	Brazil	sus_covid	unknown
82	S074	National	Egypt	sari	hospitalised
83	S075	National	South Africa	sari	hospitalised
84	S076	National	Argentina	ili	non_treatment_seeking
85	S077	National	Taiwan	ili	unknown
86	S078	National	France	ili	hospitalised
87	S079	Regional	USA	ili	hospitalised
88	S080	Regional	Mexico	ili	unknown
89	S081	Regional	Chile	sari	treatment_seeking
90	S082	Multinational	Argentina, Australia, Austria, Belgium, Chile, China, Denmark, Estonia, Germany, Greece, Norway, Peru, Poland, Spain, Thailand, United Kingdom, United States	ili	treatment_seeking
91	S082	Multinational	Argentina, Australia, Austria, Belgium, Chile, China, Denmark, Estonia, Germany, Greece, Norway, Peru, Poland, Spain, Thailand, United Kingdom, United States	ili	hospitalised
92	S083	National	Egypt	sari	hospitalised
93	S083	National	Egypt	sari	icu
94	S084	Regional	Serbia	sari	hospitalised
95	S085	National	Brazil	ili	hospitalised

96	S086	National	Nigeria	sus_covid	unknown
97	S087	National	Canada	ari	hospitalised
98	S088	Regional	Brazil	sus_covid	treatment_seeking hospitalised
99	S089	Regional	Colombia	ari	treatment_seeking hospitalised
100	S090	National	Mexico	ari	hospitalised
101	S091	Multinational	Belgium, Canada, Czech Rep., Estonia, France, Germany, Mexico, Norway, Poland, Romania, Russia, Taiwan, Netherlands, UK, USA	ili	non_treatment_seeking
102	S092	Regional	Malawi	sari	treatment_seeking
103	S093	Regional	South Africa	ili_sus_covid	treatment_seeking
104	S093	Regional	South Africa	sari_sus_covid	hospitalised
105	S094	Regional	South Africa	sari	hospitalised
106	S094	Regional	South Africa	ili	treatment_seeking
107	S095	Regional	Egypt	sari	hospitalised
108	S096	National	Bangladesh	sari	hospitalised
109	S097	National	New Zealand	sari	hospitalised
110	S098	Regional	Ireland	sari	hospitalised
111	S099	National	Georgia	sari	hospitalised
112	S100	National	Yemen	sari	hospitalised
113	S101	Regional	USA	maari	hospitalised
114	S102	Multinational	Dem. Rep. Congo, Kenya, Madagascar, Malawi, Rwanda, South Africa, Tanzania, Uganda	sari	hospitalised
115	S103	National	Kenya	sari	hospitalised
116	S104	Regional	Kenya	sari	treatment_seeking
117	S105	National	South Africa	sari	hospitalised
118	S106	National	Iran	sari_sus_covid	hospitalised
119	S107	Regional	Netherlands	ari	treatment_seeking
120	S108	National	Mexico	sus_covid	unknown
121	S109	National	Mexico	sari	hospitalised
122	S110	Regional	Italy	ili	hospitalised
123	S111	Regional	Guatemala	ari	hospitalised

124	S112	Multinational	Albania, Armenia, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Romania, Russia, Ukraine	sari	hospitalised
125	S113	Regional	Italy	ili	hospitalised
126	S114	National	Iceland	ili	treatment_seeking hospitalised
127	S115	National	Oman	sari	hospitalised
128	S116	National	Kenya	ili	treatment_seeking
129	S116	National	Kenya	sari	hospitalised
130	S117	Multinational	Austria, Belgium, Czech Republic, Croatia, Denmark, Finland, France, Georgia, Germany, Greece, Ireland, Italy, Lithuania, Macedonia, Netherlands, Poland, Portugal, Romania, Slovenia, Spain, United Kingdom, Iran, Qatar, Saudi Arabia, Turkey, United Arab Emirates, United States, Argentina, Brazil, Chile, Colombia, Ecuador, Mexico, Peru, Venezuela, Australia, China, Philippines, Vietnam, India, Pakistan, Sri Lanka	sari	icu
131	S118	Regional	Brazil	sari_sus_covid	hospitalised

132	S119	National	Ecuador	ili_sari	treatment_seeking hospitalised
133	S120	National	Mexico	ili_sari	treatment_seeking hospitalised
134	S121	National	UK	ili	treatment_seeking hospitalised
135	S122	Regional	China	sari	hospitalised
136	S123	National	Mexico	sus_covid	treatment_seeking hospitalised
137	S124	Multinational	Canada, USA, Mexico, UK, France, Finland, Saudi Arabia, Malawi, Kenya, India, China, Korea, Vietnam, Cambodia, Australia, NZ, Madagascar, Ireland	sari	icu
138	S125	National	Germany	ari	treatment_seeking
139	S126	Multinational	Egypt, Jordan, Oman, Qatar, Yemen	sari	hospitalised

A2.5 STUDY DATES

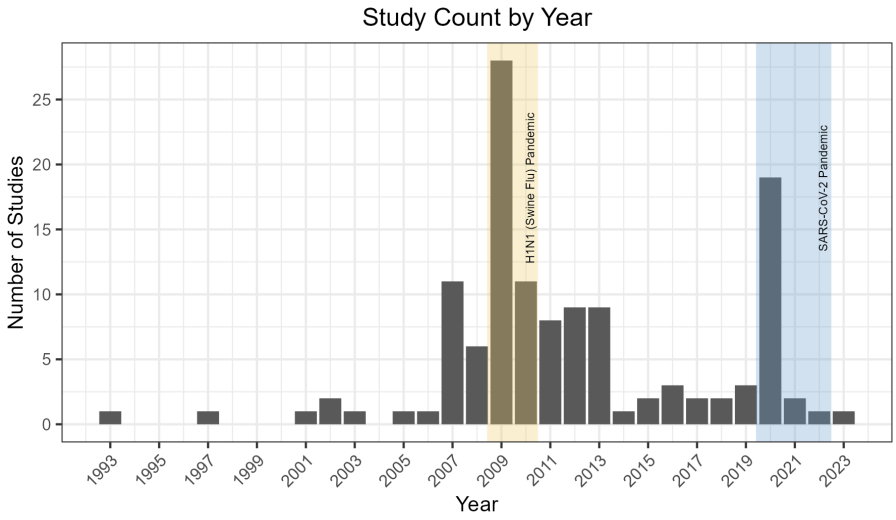


Figure A2.1: Counts of included studies by study data year (not publication date). Yellow shaded area: H1N1 influenza pandemic; blue shaded area: SARS-CoV-2 pandemic. Shaded bars rounded to the nearest year.

A2.6 SEVERITY MARKER REPORTING BY RECRUITMENT TYPE

Table A2.6: Frequency of severity markers by recruitment type

Recruitment type	Group	Severity marker	Frequency
All types			
All types	Death	Death	100 (71.9%)
All types	ICU-related	ICU admission	62 (44.6%)
All types	ICU-related	Ventilation	48 (34.5%)
All types	Symptom	Cough	47 (33.8%)
All types	Symptom	Shortness of breath	42 (30.2%)
All types	Symptom	Fever	41 (29.5%)
All types	Hospital	Hospital length of stay	41 (29.5%)
All types	Hospital	IV fluids	38 (27.3%)
All types	Symptom	Sorethroat	36 (25.9%)
All types	Symptom	Coryza rhinorrea congestion	30 (21.6%)
All types	Symptom	Myalgia arthralgia	30 (21.6%)
All types	Symptom	Diarrhea	27 (19.4%)
All types	Symptom	Nausea vomiting	27 (19.4%)
All types	Hospital	Hospital admission	27 (19.4%)
All types	Symptom	Malaise anorexia	23 (16.6%)
All types	Treatment	Antivirals	21 (15.1%)
All types	Treatment	Antibiotics	19 (13.7%)
All types	Complications	Respiratory complications	18 (12.9%)
All types	Signs	Temperature	17 (12.2%)
All types	Symptom	Chest pain	15 (10.8%)
All types	Signs	O_2 saturation	15 (10.8%)
All types	Signs	Respiratory rate	14 (10.1%)
All types	Signs	Chest signs	13 (9.3%)
All types	Symptom	Productive cough	12 (8.6%)
All types	Investigations	White cell count	11 (7.9%)
All types	Complications	Cardiac complications	11 (7.9%)
All types	Complications	Neurological complications	11 (7.9%)
All types	Symptom	Abdominal pain	10 (7.2%)
All types	Score	Project specific composite	10 (7.2%)
All types	Symptom	Length of illness	9 (6.5%)
All types	Investigations	Chest xray	9 (6.5%)
All types	Symptom	Convulsions	8 (5.8%)
All types	Symptom	Loss of taste or smell	8 (5.8%)

Table A2.6 continued

Recruitment type	Group	Severity marker	Frequency
All types	Symptom	Haemoptysis	7 (5.0%)
All types	Absence	Work absence	7 (5.0%)
All types	Signs	Respiratory distress	7 (5.0%)
All types	ICU-related	ECMO	7 (5.0%)
All types	Signs	Pulse rate	6 (4.3%)
All types	Hospital	Hospital attendance advised	6 (4.3%)
All types	Absence	School absence	5 (3.6%)
All types	Investigations	Inflammatory markers	5 (3.6%)
All types	Complications	Renal complications	5 (3.6%)
All types	Hospital	O_2 therapy	5 (3.6%)
All types	ICU-related	ICU length of stay	5 (3.6%)
All types	Signs	Blood pressure	4 (2.9%)
All types	Signs	Cyanosis	4 (2.9%)
All types	Complications	Organ failure	4 (2.9%)
All types	Complications	Sepsis	4 (2.9%)
All types	Death	HSB	3 (2.2%)
All types	Symptom	Confusion	3 (2.2%)
All types	Symptom	Ear pain	3 (2.2%)
All types	Score	Sofa score	3 (2.2%)
All types	Investigations	Urea electrolytes	3 (2.2%)
All types	Hospital	Duration O_2 therapy	3 (2.2%)
All types	Score	Who severity	2 (1.4%)
All types	Investigations	Liver function tests	2 (1.4%)
All types	Hospital	Hospital attendance	2 (1.4%)
All types	ICU-related	Duration ventilation	2 (1.4%)
All types	ICU-related	Inotropes	2 (1.4%)
All types	Symptom	Irritability	1 (0.7%)
All types	Symptom	Retroocular pain	1 (0.7%)
All types	Score	American academy pediatrics guideline criteria	1 (0.7%)
All types	Score	Apache iv	1 (0.7%)
All types	Score	Barthel index	1 (0.7%)
All types	Score	Euroqol	1 (0.7%)
All types	Score	Gcs	1 (0.7%)
All types	Score	Iss	1 (0.7%)
All types	Score	Mews	1 (0.7%)

Table A2.6 continued

Recruitment type	Group	Severity marker	Frequency
All types	Score	Paediatric chinese medical association	1 (0.7%)
All types	Score	Pediatric risk of mortality iii score	1 (0.7%)
All types	Score	Saps 2 score	1 (0.7%)
All types	Score	ViVI score	1 (0.7%)
All types	Investigations	Arterial blood gas	1 (0.7%)
All types	Investigations	Fibronogen	1 (0.7%)
All types	Investigations	Procalcitonin	1 (0.7%)
All types	Complications	Shock	1 (0.7%)
All types	Treatment	Steroids	1 (0.7%)
Hospitalised			
Hospitalised	Death	Death	63 (87.5%)
Hospitalised	ICU-related	ICU admission	51 (70.8%)
Hospitalised	Hospital	Hospital length of stay	36 (50.0%)
Hospitalised	ICU-related	Ventilation	34 (47.2%)
Hospitalised	Symptom	Cough	23 (31.9%)
Hospitalised	Symptom	Shortness of breath	21 (29.2%)
Hospitalised	Symptom	Fever	19 (26.4%)
Hospitalised	Hospital	Hospital admission	18 (25.0%)
Hospitalised	Symptom	Sorethroat	15 (20.8%)
Hospitalised	Symptom	Coryza rhinorrea congestion	13 (18.1%)
Hospitalised	Treatment	Antibiotics	13 (18.1%)
Hospitalised	Complications	Respiratory complications	12 (16.7%)
Hospitalised	Symptom	Myalgia arthralgia	11 (15.3%)
Hospitalised	Signs	Chest signs	10 (13.9%)
Hospitalised	Signs	Temperature	10 (13.9%)
Hospitalised	Symptom	Diarrhea	9 (12.5%)
Hospitalised	Symptom	Nausea vomiting	9 (12.5%)
Hospitalised	Treatment	Antivirals	9 (12.5%)
Hospitalised	Symptom	Malaise anorexia	8 (11.1%)
Hospitalised	Signs	Respiratory rate	8 (11.1%)
Hospitalised	Investigations	Chest xray	8 (11.1%)
Hospitalised	Complications	Cardiac complications	8 (11.1%)
Hospitalised	Complications	Neurological complications	7 (9.7%)
Hospitalised	Symptom	Productive cough	6 (8.3%)
Hospitalised	Investigations	White cell count	6 (8.3%)
Hospitalised	ICU-related	ECMO	6 (8.3%)

Table A2.6 continued

Recruitment type	Group	Severity marker	Frequency
Hospitalised	Symptom	Chest pain	5 (6.9%)
Hospitalised	Symptom	Convulsions	5 (6.9%)
Hospitalised	Signs	O_2 saturation	5 (6.9%)
Hospitalised	Hospital	Hospital attendance advised	4 (5.6%)
Hospitalised	Symptom	Abdominal pain	3 (4.2%)
Hospitalised	Symptom	Haemoptysis	3 (4.2%)
Hospitalised	Score	Project specific composite	3 (4.2%)
Hospitalised	Complications	Renal complications	3 (4.2%)
Hospitalised	Complications	Sepsis	3 (4.2%)
Hospitalised	Symptom	Length of illness	2 (2.8%)
Hospitalised	Symptom	Loss of taste or smell	2 (2.8%)
Hospitalised	Signs	Cyanosis	2 (2.8%)
Hospitalised	Signs	Pulse rate	2 (2.8%)
Hospitalised	Signs	Respiratory distress	2 (2.8%)
Hospitalised	Hospital	IV fluids	2 (2.8%)
Hospitalised	ICU-related	ICU length of stay	2 (2.8%)
Hospitalised	ICU-related	Duration ventilation	2 (2.8%)
Hospitalised	Signs	Blood pressure	1 (1.4%)
Hospitalised	Score	Barthel index	1 (1.4%)
Hospitalised	Score	GCS	1 (1.4%)
Hospitalised	Score	Pediatric risk of mortality iii score	1 (1.4%)
Hospitalised	Score	Who severity	1 (1.4%)
Hospitalised	Investigations	Inflammatory markers	1 (1.4%)
Hospitalised	Investigations	Urea electrolytes	1 (1.4%)
Hospitalised	Complications	Organ failure	1 (1.4%)
Hospitalised	Complications	Shock	1 (1.4%)
Hospitalised	Treatment	Steroids	1 (1.4%)
Hospitalised	Hospital	Hospital attendance	1 (1.4%)
Intensive care			
Intensive care	Death	Death	4 (100.0%)
Intensive care	ICU-related	ICU length of stay	3 (75.0%)
Intensive care	ICU-related	Ventilation	3 (75.0%)
Intensive care	Symptom	Cough	2 (50.0%)
Intensive care	Symptom	Nausea vomiting	2 (50.0%)
Intensive care	Symptom	Shortness of breath	2 (50.0%)
Intensive care	Score	SOFA score	2 (50.0%)

Table A2.6 continued

Recruitment type	Group	Severity marker	Frequency
Intensive care	Symptom	Convulsions	1 (25.0%)
Intensive care	Symptom	Coryza rhinorrea congestion	1 (25.0%)
Intensive care	Symptom	Productive cough	1 (25.0%)
Intensive care	Symptom	Diarrhea	1 (25.0%)
Intensive care	Symptom	Fever	1 (25.0%)
Intensive care	Symptom	Myalgia arthralgia	1 (25.0%)
Intensive care	Symptom	Sorethroat	1 (25.0%)
Intensive care	Signs	Chest signs	1 (25.0%)
Intensive care	Score	APACHE IV	1 (25.0%)
Intensive care	Score	SAPS 2 score	1 (25.0%)
Intensive care	Investigations	Chest xray	1 (25.0%)
Intensive care	Complications	Organ failure	1 (25.0%)
Intensive care	Complications	Respiratory complications	1 (25.0%)
Intensive care	Treatment	Antibiotics	1 (25.0%)
Non treatment seeking			
Non treatment seeking	Absence	Work absence	5 (62.5%)
Non treatment seeking	Absence	School absence	4 (50.0%)
Non treatment seeking	Symptom	Cough	3 (37.5%)
Non treatment seeking	Symptom	Fever	3 (37.5%)
Non treatment seeking	Symptom	Length of illness	3 (37.5%)
Non treatment seeking	Symptom	Myalgia arthralgia	3 (37.5%)
Non treatment seeking	Death	Death	2 (25.0%)
Non treatment seeking	Death	HSB	2 (25.0%)
Non treatment seeking	Symptom	Coryza rhinorrea congestion	2 (25.0%)
Non treatment seeking	Symptom	Sorethroat	2 (25.0%)
Non treatment seeking	Hospital	IV fluids	2 (25.0%)
Non treatment seeking	Hospital	O_2 therapy	2 (25.0%)
Non treatment seeking	Symptom	Abdominal pain	1 (12.5%)
Non treatment seeking	Symptom	Diarrhea	1 (12.5%)
Non treatment seeking	Symptom	Malaise anorexia	1 (12.5%)
Non treatment seeking	Symptom	Nausea vomiting	1 (12.5%)
Non treatment seeking	Symptom	Shortness of breath	1 (12.5%)
Non treatment seeking	Signs	O_2 saturation	1 (12.5%)
Non treatment seeking	Score	ISS	1 (12.5%)
Non treatment seeking	Score	Paediatric chinese medical association	1 (12.5%)
Non treatment seeking	Investigations	Fibronogen	1 (12.5%)

Table A2.6 continued

Recruitment type	Group	Severity marker	Frequency
Non treatment seeking	Investigations	Inflammatory markers	1 (12.5%)
Non treatment seeking	Investigations	Liver function tests	1 (12.5%)
Non treatment seeking	Investigations	Procalcitonin	1 (12.5%)
Non treatment seeking	Investigations	Urea electrolytes	1 (12.5%)
Non treatment seeking	Investigations	White cell count	1 (12.5%)
Non treatment seeking	Complications	Cardiac complications	1 (12.5%)
Non treatment seeking	Complications	Neurological complications	1 (12.5%)
Non treatment seeking	Complications	Respiratory complications	1 (12.5%)
Non treatment seeking	Complications	Renal complications	1 (12.5%)
Non treatment seeking	Hospital	Hospital admission	1 (12.5%)
Non treatment seeking	Hospital	Duration O_2 therapy	1 (12.5%)
Non treatment seeking	Hospital	Hospital length of stay	1 (12.5%)
Non treatment seeking	ICU-related	Ventilation	1 (12.5%)
Non treatment seeking	ICU-related	Inotropes	1 (12.5%)
Treatment seeking			
Treatment seeking	Hospital	IV fluids	23 (67.7%)
Treatment seeking	Death	Death	14 (41.2%)
Treatment seeking	Symptom	Cough	13 (38.2%)
Treatment seeking	Symptom	Shortness of breath	13 (38.2%)
Treatment seeking	Symptom	Diarrhea	12 (35.3%)
Treatment seeking	Symptom	Fever	12 (35.3%)
Treatment seeking	Symptom	Sorethroat	11 (32.4%)
Treatment seeking	Symptom	Coryza rhinorrea congestion	9 (26.5%)
Treatment seeking	Symptom	Myalgia arthralgia	9 (26.5%)
Treatment seeking	Symptom	Nausea vomiting	9 (26.5%)
Treatment seeking	Signs	O_2 saturation	9 (26.5%)
Treatment seeking	Symptom	Malaise anorexia	8 (23.5%)
Treatment seeking	Hospital	Hospital admission	8 (23.5%)
Treatment seeking	Treatment	Antivirals	7 (20.6%)
Treatment seeking	Symptom	Chest pain	6 (17.6%)
Treatment seeking	Signs	Temperature	6 (17.6%)
Treatment seeking	Score	Project specific composite	6 (17.6%)
Treatment seeking	Signs	Respiratory rate	5 (14.7%)
Treatment seeking	Signs	Respiratory distress	5 (14.7%)
Treatment seeking	Treatment	Antibiotics	5 (14.7%)
Treatment seeking	Symptom	Productive cough	4 (11.8%)

Table A2.6 continued

Recruitment type	Group	Severity marker	Frequency
Treatment seeking	Signs	Pulse rate	4 (11.8%)
Treatment seeking	ICU-related	ICU admission	4 (11.8%)
Treatment seeking	Symptom	Haemoptysis	3 (8.8%)
Treatment seeking	Symptom	Length of illness	3 (8.8%)
Treatment seeking	Symptom	Loss of taste or smell	3 (8.8%)
Treatment seeking	Signs	Blood pressure	3 (8.8%)
Treatment seeking	Investigations	Inflammatory markers	3 (8.8%)
Treatment seeking	Investigations	White cell count	3 (8.8%)
Treatment seeking	Symptom	Abdominal pain	2 (5.9%)
Treatment seeking	Symptom	Confusion	2 (5.9%)
Treatment seeking	Symptom	Convulsions	2 (5.9%)
Treatment seeking	Symptom	Ear pain	2 (5.9%)
Treatment seeking	Signs	Chest signs	2 (5.9%)
Treatment seeking	Complications	Neurological complications	2 (5.9%)
Treatment seeking	Complications	Respiratory complications	2 (5.9%)
Treatment seeking	Hospital	Hospital attendance advised	2 (5.9%)
Treatment seeking	Hospital	O_2 therapy	2 (5.9%)
Treatment seeking	Hospital	Duration O_2 therapy	2 (5.9%)
Treatment seeking	Hospital	Hospital length of stay	2 (5.9%)
Treatment seeking	ICU-related	Ventilation	2 (5.9%)
Treatment seeking	Death	HSB	1 (2.9%)
Treatment seeking	Absence	School absence	1 (2.9%)
Treatment seeking	Absence	Work absence	1 (2.9%)
Treatment seeking	Signs	Cyanosis	1 (2.9%)
Treatment seeking	Score	American academy pediatrics guideline criteria	1 (2.9%)
Treatment seeking	Score	MEWS	1 (2.9%)
Treatment seeking	Score	SOFA score	1 (2.9%)
Treatment seeking	Score	Vivl score	1 (2.9%)
Treatment seeking	Score	Who severity	1 (2.9%)
Treatment seeking	Investigations	Arterial blood gas	1 (2.9%)
Treatment seeking	Investigations	Liver function tests	1 (2.9%)
Treatment seeking	Investigations	Urea electrolytes	1 (2.9%)
Treatment seeking	Complications	Cardiac complications	1 (2.9%)
Treatment seeking	Complications	Organ failure	1 (2.9%)
Treatment seeking	Complications	Renal complications	1 (2.9%)

Table A2.6 continued

Recruitment type	Group	Severity marker	Frequency
Treatment seeking	Complications	Sepsis	1 (2.9%)
Treatment seeking	Hospital	Hospital attendance	1 (2.9%)
Treatment seeking	ICU-related	Inotropes	1 (2.9%)
Treatment seeking hospitalised			
Treatment seeking hospitalised	Death	Death	11 (73.3%)
Treatment seeking hospitalised	Hospital	IV fluids	10 (66.7%)
Treatment seeking hospitalised	ICU-related	Ventilation	7 (46.7%)
Treatment seeking hospitalised	ICU-related	ICU admission	6 (40.0%)
Treatment seeking hospitalised	Symptom	Nausea vomiting	5 (33.3%)
Treatment seeking hospitalised	Symptom	Sorethroat	5 (33.3%)
Treatment seeking hospitalised	Symptom	Coryza rhinorrea congestion	4 (26.7%)
Treatment seeking hospitalised	Symptom	Cough	4 (26.7%)
Treatment seeking hospitalised	Symptom	Fever	4 (26.7%)
Treatment seeking hospitalised	Symptom	Malaise anorexia	4 (26.7%)
Treatment seeking hospitalised	Symptom	Myalgia arthralgia	4 (26.7%)
Treatment seeking hospitalised	Treatment	Antivirals	4 (26.7%)
Treatment seeking hospitalised	Symptom	Abdominal pain	3 (20.0%)
Treatment seeking hospitalised	Symptom	Diarrhea	3 (20.0%)
Treatment seeking hospitalised	Symptom	Shortness of breath	3 (20.0%)

Table A2.6 continued

Recruitment type	Group	Severity marker	Frequency
Treatment seeking hospitalised	Symptom	Chest pain	2 (13.3%)
Treatment seeking hospitalised	Symptom	Loss of taste or smell	2 (13.3%)
Treatment seeking hospitalised	Complications	Respiratory complications	2 (13.3%)
Treatment seeking hospitalised	Hospital	Hospital length of stay	2 (13.3%)
Treatment seeking hospitalised	Symptom	Productive cough	1 (6.7%)
Treatment seeking hospitalised	Symptom	Ear pain	1 (6.7%)
Treatment seeking hospitalised	Symptom	Haemoptysis	1 (6.7%)
Treatment seeking hospitalised	Symptom	Length of illness	1 (6.7%)
Treatment seeking hospitalised	Symptom	Retroocular pain	1 (6.7%)
Treatment seeking hospitalised	Absence	Work absence	1 (6.7%)
Treatment seeking hospitalised	Signs	Temperature	1 (6.7%)
Treatment seeking hospitalised	Score	EUROQOL	1 (6.7%)
Treatment seeking hospitalised	Score	Project specific composite	1 (6.7%)
Treatment seeking hospitalised	Complications	Cardiac complications	1 (6.7%)
Treatment seeking hospitalised	Complications	Neurological complications	1 (6.7%)
Treatment seeking hospitalised	Complications	Organ failure	1 (6.7%)
Treatment seeking hospitalised	Hospital	O_2 therapy	1 (6.7%)

Table A2.6 continued

Recruitment type	Group	Severity marker	Frequency
Treatment seeking hospitalised	ICU-related	ECMO	1 (6.7%)
Unknown			
Unknown	Death	Death	6 (100.0%)
Unknown	Symptom	Chest pain	2 (33.3%)
Unknown	Symptom	Cough	2 (33.3%)
Unknown	Symptom	Fever	2 (33.3%)
Unknown	Symptom	Malaise anorexia	2 (33.3%)
Unknown	Symptom	Myalgia arthralgia	2 (33.3%)
Unknown	Symptom	Shortness of breath	2 (33.3%)
Unknown	Symptom	Sorethroat	2 (33.3%)
Unknown	Symptom	Abdominal pain	1 (16.7%)
Unknown	Symptom	Confusion	1 (16.7%)
Unknown	Symptom	Coryza rhinorrea congestion	1 (16.7%)
Unknown	Symptom	Diarrhea	1 (16.7%)
Unknown	Symptom	Irritability	1 (16.7%)
Unknown	Symptom	Loss of taste or smell	1 (16.7%)
Unknown	Symptom	Nausea vomiting	1 (16.7%)
Unknown	Signs	Cyanosis	1 (16.7%)
Unknown	Signs	Respiratory rate	1 (16.7%)
Unknown	Investigations	White cell count	1 (16.7%)
Unknown	Treatment	Antivirals	1 (16.7%)
Unknown	Hospital	IV fluids	1 (16.7%)
Unknown	ICU-related	ICU admission	1 (16.7%)
Unknown	ICU-related	Ventilation	1 (16.7%)

Notes: Percentages represent the proportion of studies within each recruitment type that reported the specified severity marker. HSB: health service burden; ICU: intensive care unit; ECMO: extracorporeal membrane oxygenation; MEWS: Modified Early Warning Score; SOFA: Sequential Organ Failure Assessment; SAPS: Simplified Acute Physiology Score; GCS: Glasgow Coma Scale; ISS: Injury Severity Score.

Appendix A3

Chapter 4 Data quality assessment

A3.1 RECORDING COMPLETENESS OF CLINICAL SIGNS AND SCORES BY AGE GROUP

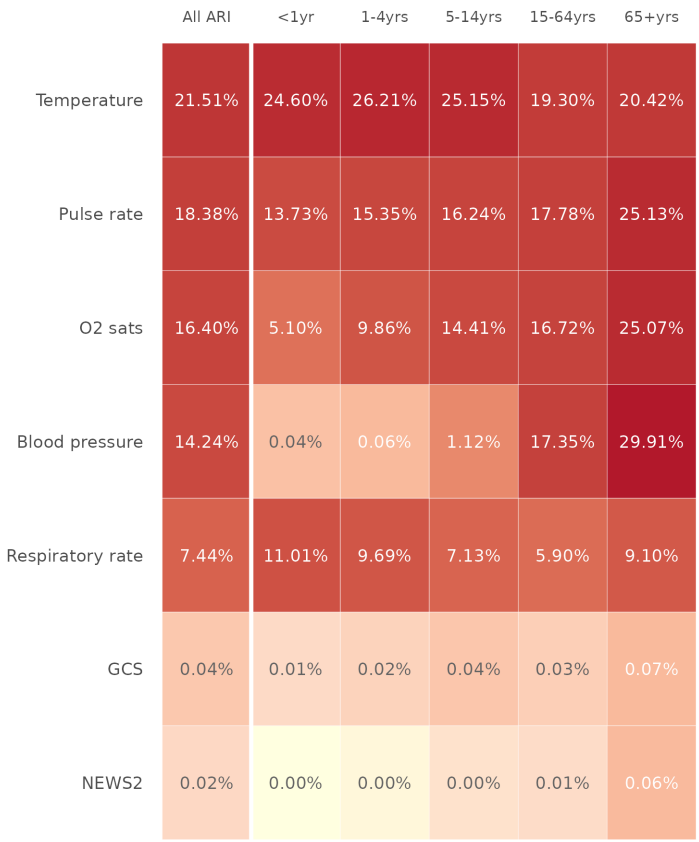


Figure A3.1: Completeness of recording of clinical signs and clinical scores by all ARI and ARI subtype. ARI: Acute respiratory infection; URTI: Upper respiratory tract infection;LRTI: lower respiratory tract infection; ILI: Influenza-like illness, ECLD: Exacerbation of chronic lung disease; Sus COVID: suspected COVID-19; ARI NOS: ARI not otherwise specified; O2 Sats: Periperhal oxygen saturation; GCS: Glasgow Coma Scale; NEWS2: National Early Warning Score 2.

A3.2 CALIBRATION PLOT FOR MULTIPLE-VARIABLE LOGISTIC REGRESSION COMPLETENESS ANALYSIS

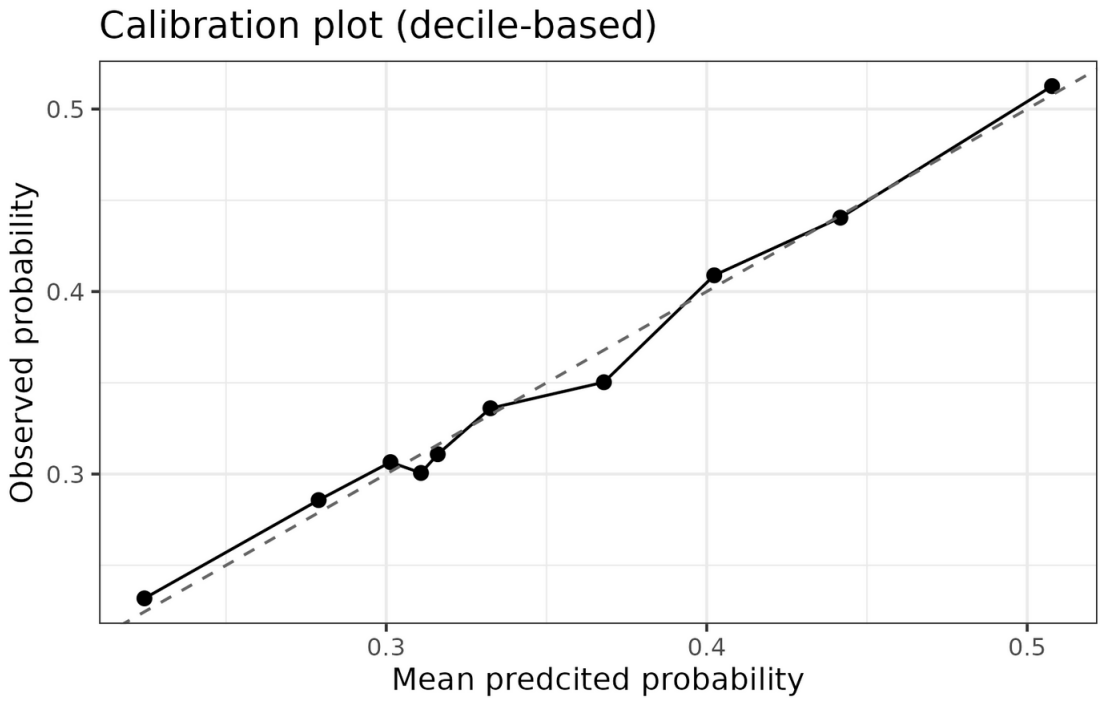


Figure A3.2: Multiple-variable Logistic regression calibration plot highlighting how model's predicted probabilities align with the observed outcomes

A3.3 MEDIAN WEEKLY RECORDING RATES BY STUDY PERIOD

Table A3.1: Median weekly rate (%) and interquartile range (IQR) for each severity marker by period

Severity marker	All	Period 1	Period 2	Period 3	Period 4
Any severe outcome	8.52 (6.57–10.03)	5.68 (4.74–6.69)	8.83 (8.31–9.90)	10.73 (9.64–12.30)	10.26 (9.63–11.46)
Any complication	0.22 (0.04–0.34)	0.04 (0.03–0.05)	0.26 (0.09–0.35)	0.33 (0.28–0.41)	0.35 (0.30–0.39)

Continued on next page

Table A3.1: Median weekly rate (%) and interquartile range (IQR) for each severity marker by period (continued)

Severity marker	All	Period 1	Period 2	Period 3	Period 4
Sepsis	0.20 (0.04–0.30)	0.03 (0.02–0.04)	0.25 (0.08–0.33)	0.28 (0.23–0.34)	0.30 (0.25–0.34)
Acute respiratory failure	0.01 (0.00–0.03)	0.00 (0.00–0.00)	0.01 (0.00–0.01)	0.04 (0.03–0.06)	0.04 (0.03–0.06)
Cardiac arrest	0.01 (0.00–0.01)	0.01 (0.00–0.01)	0.01 (0.00–0.01)	0.01 (0.01–0.02)	0.01 (0.01–0.02)
Any hospital	8.10 (6.09–9.58)	5.27 (4.29–6.25)	8.39 (7.85–9.45)	10.08 (9.10–11.42)	9.79 (9.10–10.84)
Attendance	7.13 (5.37–8.53)	4.50 (3.60–5.56)	7.36 (6.86–8.34)	8.97 (8.19–10.02)	8.80 (8.18–9.62)
Hospital admission	1.71 (1.13–2.02)	1.05 (0.96–1.15)	1.96 (1.73–2.15)	2.11 (1.76–2.72)	1.90 (1.74–2.33)
Any ICU	0.01 (0.00–0.01)	0.00 (0.00–0.01)	0.01 (0.00–0.01)	0.02 (0.01–0.03)	0.01 (0.01–0.02)
ICU admission	0.01 (0.00–0.01)	0.00 (0.00–0.01)	0.01 (0.00–0.01)	0.02 (0.01–0.03)	0.01 (0.01–0.02)
Any death	0.60 (0.52–0.69)	0.52 (0.46–0.59)	0.63 (0.58–0.71)	0.79 (0.64–1.45)	0.64 (0.56–0.84)
Died ≤14 days	0.21 (0.17–0.25)	0.17 (0.14–0.20)	0.23 (0.20–0.26)	0.31 (0.24–0.60)	0.25 (0.21–0.34)
Died ≤28 days	0.36 (0.31–0.42)	0.31 (0.27–0.35)	0.38 (0.35–0.43)	0.51 (0.40–0.92)	0.41 (0.35–0.57)
Died ≤56 days	0.60 (0.52–0.69)	0.52 (0.46–0.59)	0.63 (0.58–0.71)	0.79 (0.64–1.45)	0.64 (0.56–0.84)

Continued on next page

Table A3.1: Median weekly rate (%) and interquartile range (IQR) for each severity marker by period (continued)

Severity marker	All	Period 1	Period 2	Period 3	Period 4
Any predictor	76.73 (72.27–83.02)	71.94 (70.13–73.54)	80.80 (77.59–83.45)	75.31 (71.85–78.66)	81.96 (74.30–86.70)
Any symptom	4.11 (2.54–5.29)	2.31 (2.10–2.57)	4.39 (3.72–4.90)	6.10 (5.06–8.45)	5.85 (5.19–6.72)
Dyspnoea	1.62 (1.28–2.72)	1.21 (1.05–1.34)	1.68 (1.44–2.01)	3.32 (2.66–5.18)	3.32 (2.87–4.20)
Fever	1.45 (0.68–1.94)	0.54 (0.47–0.70)	1.88 (1.50–2.15)	2.02 (1.60–3.06)	1.61 (1.44–2.20)
Malaise	0.86 (0.51–1.07)	0.46 (0.41–0.53)	1.02 (0.85–1.19)	0.97 (0.84–1.34)	1.00 (0.91–1.12)
Haemoptysis	0.08 (0.06–0.10)	0.06 (0.05–0.07)	0.08 (0.06–0.09)	0.11 (0.09–0.14)	0.12 (0.10–0.14)
Confusion	0.06 (0.04–0.09)	0.04 (0.03–0.05)	0.06 (0.05–0.07)	0.12 (0.09–0.18)	0.12 (0.09–0.15)
Parental concern	0.05 (0.04–0.07)	0.04 (0.03–0.05)	0.05 (0.04–0.07)	0.08 (0.05–0.15)	0.06 (0.04–0.10)
Any health seeking	1.15 (0.09–4.63)	0.07 (0.06–0.10)	1.25 (0.78–1.98)	6.09 (4.92–7.44)	5.08 (4.75–6.31)
NHS 111	1.01 (0.01–3.72)	0.00 (0.00–0.01)	1.06 (0.70–1.58)	4.91 (4.12–6.31)	4.18 (3.89–5.35)
Ambulance encounter	0.14 (0.07–1.01)	0.07 (0.05–0.08)	0.22 (0.09–0.46)	1.26 (1.07–1.54)	1.22 (1.06–1.39)
Any absenteeism	3.55 (2.61–4.32)	2.32 (1.94–2.73)	3.72 (3.36–4.15)	4.93 (4.01–5.78)	4.58 (3.96–5.32)

Continued on next page

Table A3.1: Median weekly rate (%) and interquartile range (IQR) for each severity marker by period (continued)

Severity marker	All	Period 1	Period 2	Period 3	Period 4
Work absenteeism	3.55 (2.61–4.32)	2.32 (1.94–2.73)	3.72 (3.36–4.15)	4.93 (4.01–5.78)	4.58 (3.96–5.32)
Any sign	31.95 (18.76–51.81)	17.56 (16.14–20.98)	48.58 (38.24–56.23)	26.14 (20.57–32.64)	41.50 (24.75–55.31)
Temperature	18.72 (6.02–37.91)	3.54 (2.90–7.00)	32.26 (23.30–39.78)	16.98 (12.53–22.52)	30.17 (15.40–42.69)
Pulse rate	14.00 (5.21–32.34)	3.91 (2.69–5.47)	25.94 (15.49–34.16)	16.23 (11.51–21.69)	28.01 (15.06–40.48)
Blood pressure	13.61 (11.56–17.06)	12.18 (11.36–12.98)	17.12 (14.39–19.87)	9.87 (8.75–11.54)	13.25 (9.60–17.03)
O2 saturation	12.99 (3.22–29.30)	1.82 (1.08–3.51)	24.22 (14.83–31.87)	13.51 (9.71–19.13)	24.75 (12.47–36.77)
Respiratory rate	5.63 (1.37–12.70)	0.82 (0.60–1.53)	9.42 (4.92–13.01)	7.76 (5.65–9.99)	12.45 (7.25–17.79)
Chest signs	1.62 (1.40–2.10)	1.42 (1.34–1.52)	2.01 (1.78–2.27)	1.36 (1.21–1.55)	1.67 (1.34–2.31)
Capillary refill	1.39 (0.36–3.86)	0.21 (0.12–0.40)	3.07 (1.38–4.54)	2.41 (1.26–3.02)	3.37 (1.89–4.54)
Work of breathing	0.01 (0.01–0.02)	0.01 (0.00–0.01)	0.01 (0.01–0.02)	0.02 (0.01–0.03)	0.02 (0.01–0.02)
Cyanosis	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.01)	0.00 (0.00–0.01)
Any score	2.05 (0.19–3.24)	0.15 (0.11–0.21)	2.62 (1.00–3.45)	2.48 (2.22–3.04)	3.17 (2.40–4.60)

Continued on next page

Table A3.1: Median weekly rate (%) and interquartile range (IQR) for each severity marker by period (continued)

Severity marker	All	Period 1	Period 2	Period 3	Period 4
AVPU	0.96 (0.18–1.96)	0.14 (0.10–0.19)	1.20 (0.30–2.04)	1.44 (1.22–1.97)	1.94 (1.39–2.19)
Centor	0.74 (0.00–1.12)	0.00 (0.00–0.00)	1.03 (0.68–1.35)	0.87 (0.73–1.16)	0.99 (0.81–1.21)
NICE paediatric traffic light	0.02 (0.01–0.04)	0.00 (0.00–0.01)	0.05 (0.02–0.07)	0.02 (0.01–0.03)	0.03 (0.02–0.04)
CRB65	0.02 (0.00–0.07)	0.00 (0.00–0.01)	0.07 (0.03–0.11)	0.01 (0.01–0.03)	0.03 (0.01–0.06)
GCS	0.01 (0.00–0.04)	0.00 (0.00–0.00)	0.01 (0.01–0.02)	0.05 (0.03–0.08)	0.11 (0.05–0.17)
NEWS2	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.02 (0.00–0.04)	0.05 (0.02–0.10)
Westley	0.01 (0.00–0.02)	0.00 (0.00–0.00)	0.01 (0.01–0.02)	0.02 (0.01–0.03)	0.02 (0.01–0.03)
feverPAIN	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–1.36)
Any investigation	7.36 (6.36–8.41)	6.38 (5.80–6.87)	8.08 (7.38–8.74)	7.45 (6.31–8.68)	7.99 (6.92–9.06)
White blood cell	5.61 (4.91–6.33)	4.98 (4.54–5.37)	6.16 (5.57–6.65)	5.74 (4.82–6.58)	5.97 (5.23–6.70)
CRP	2.21 (1.46–2.76)	1.31 (1.10–1.54)	2.54 (2.16–2.90)	2.70 (2.28–3.22)	2.70 (2.35–3.05)
Chest X-ray	2.14 (1.69–2.52)	1.62 (1.41–1.88)	2.39 (2.17–2.60)	2.07 (1.74–2.60)	2.31 (1.91–2.77)

Continued on next page

Table A3.1: Median weekly rate (%) and interquartile range (IQR) for each severity marker by period (continued)

Severity marker	All	Period 1	Period 2	Period 3	Period 4
Any prescription	62.76 (59.84–64.78)	63.93 (62.30–65.32)	61.93 (59.56–63.66)	57.40 (54.19–60.21)	62.10 (56.92–65.56)
Amoxicillin	34.35 (31.86–37.08)	37.42 (36.13–38.64)	34.07 (32.80–35.42)	26.26 (23.13–28.70)	30.03 (25.51–32.29)
Penicillin V	11.50 (9.27–13.31)	10.16 (8.59–11.81)	10.89 (9.02–13.09)	12.44 (11.23–14.77)	13.29 (11.83–15.20)
Macrolide	9.17 (7.90–10.17)	10.09 (9.49–10.60)	9.39 (8.68–10.18)	7.14 (6.64–7.90)	7.48 (6.87–8.03)
Steroid	8.06 (6.70–9.15)	6.26 (5.69–6.82)	8.59 (7.86–9.31)	8.75 (8.30–9.82)	9.03 (8.33–9.89)
Doxycycline	7.51 (5.58–10.37)	4.96 (4.15–5.78)	7.70 (6.83–8.62)	11.16 (10.12–12.20)	12.10 (10.80–13.16)
Co-amoxiclav	1.82 (1.45–2.72)	2.82 (2.62–3.07)	1.59 (1.45–1.95)	1.52 (1.32–1.85)	1.45 (1.31–1.63)
Cephalosporin	0.47 (0.34–1.03)	1.47 (0.97–2.08)	0.43 (0.36–0.53)	0.32 (0.30–0.39)	0.32 (0.29–0.37)
Oseltamivir	0.01 (0.00–0.02)	0.01 (0.00–0.03)	0.01 (0.00–0.04)	0.00 (0.00–0.00)	0.00 (0.00–0.01)
Zanamivir	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)
Molnupiravir	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)
Paxlovid	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)

Continued on next page

Table A3.1: Median weekly rate (%) and interquartile range (IQR) for each severity marker by period (continued)

Severity marker	All	Period 1	Period 2	Period 3	Period 4
Sotrovimab	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)
Any hospital (referral)	0.13 (0.09–0.17)	0.07 (0.06–0.09)	0.15 (0.12–0.17)	0.17 (0.14–0.22)	0.17 (0.15–0.20)
Emergency hospital referral	0.13 (0.09–0.17)	0.07 (0.06–0.09)	0.15 (0.12–0.17)	0.17 (0.14–0.22)	0.17 (0.15–0.20)

A3.4 ADDITIONAL TIME SERIES FIGURES

A3.4.1 Hospital

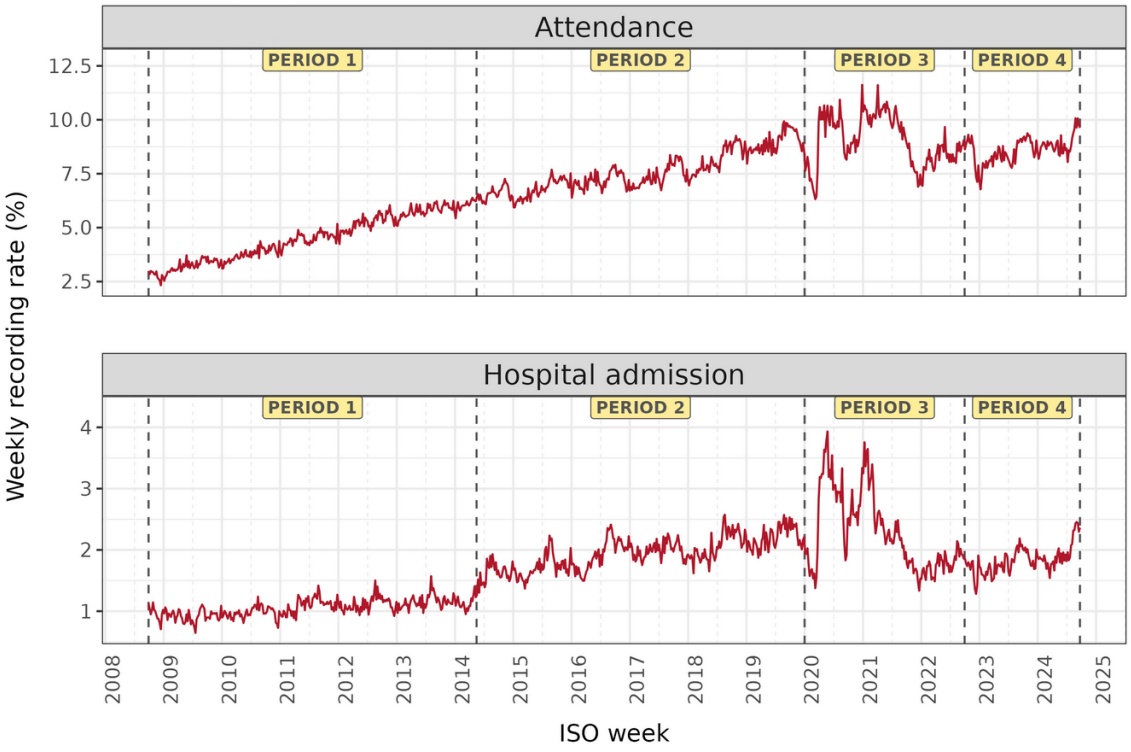


Figure A3.3: Time series of weekly hospital attendances (Top panel) and admissions (Bottom panel) recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO week: International Organization for Stadardardization.

A3.4.2 Intensive Care

The only severity marker in the ICU outcomes group was ICU admission, with a low overall median weekly recording rate of 0.01% (interquartile range 0.00–0.01). The recording rate of ICU admission saw a 1.79-fold increase from Period 1 to Period 2, followed by a 2.68-fold increase from Period 2 to Period 3 (COVID-19 pandemic). As for hospital admission 2 spikes in ICU admission can be seen corresponding to the first 2 waves of COVID. However, ICU weekly rates much more volatile due to the low overall recording rates.

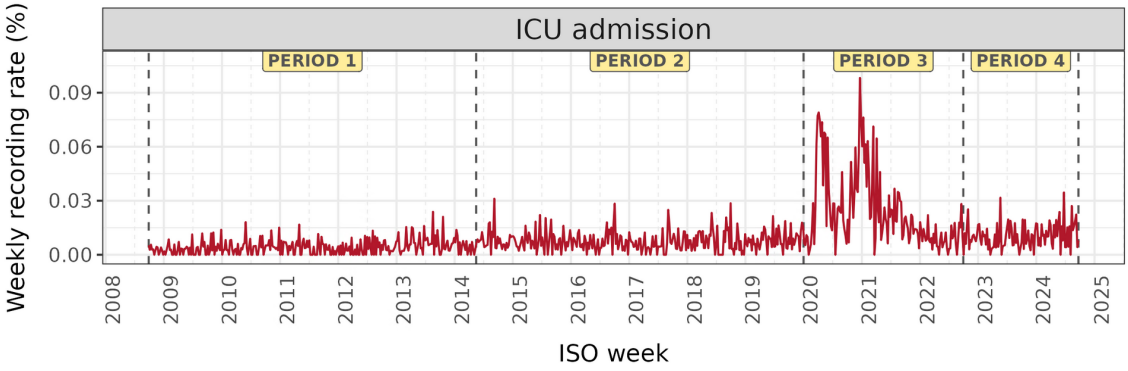


Figure A3.4: Time series of weekly Intensive Care Unit (ICU) admission recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO week: International Organization for Stadardardization.

A3.4.3 Complications

The overall median weekly recording rate for complications (any complication) was 0.22% (interquartile range 0.04–0.34). Within this group, sepsis (0.20%, IQR: 0.04–0.30) had the highest median weekly rate, followed by acute respiratory failure (0.01%, IQR: 0.00–0.03) and cardiac arrest (0.01%, IQR: 0.00–0.01). The recording rate of any complication saw a 6.71-fold increase from Period 1 to Period 2. The time series visualisation shows that the main driver of this change was an increase in sepsis recording, which rose sharply midway through Period 2, around 2016. There was a smaller 1.32-fold increase in overall complication recording from Period 2 to Period 3 (COVID-19 pandemic). In the post-pandemic period, recording rates appear to have returned to pre-pandemic levels.

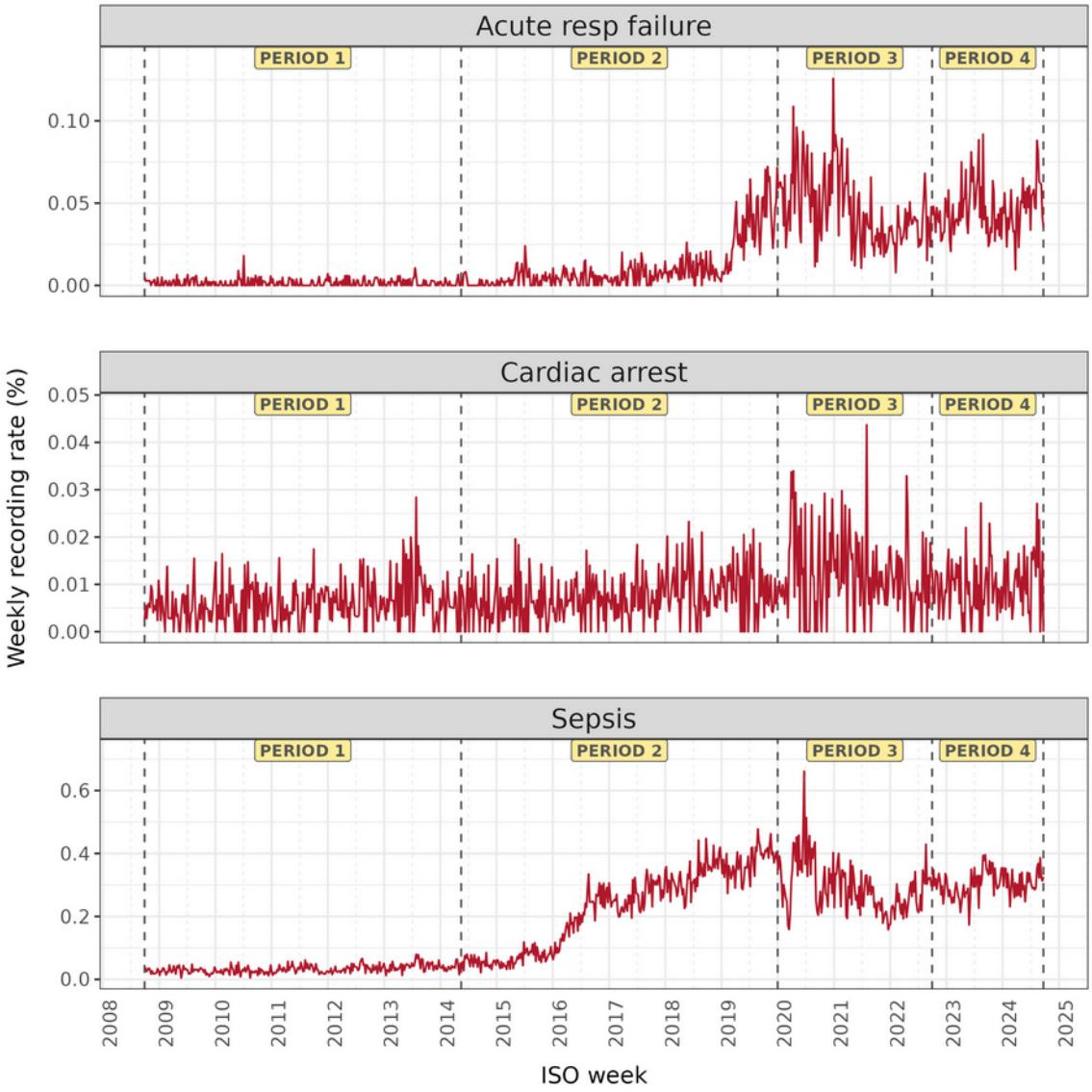


Figure A3.5: Time series of weekly complication (Acute respiratory failure- top panel, cardiac arrest- middle panel, sepsis- bottom panel) recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO week: International Organization for Stadardardization.

A3.4.4 Symptoms

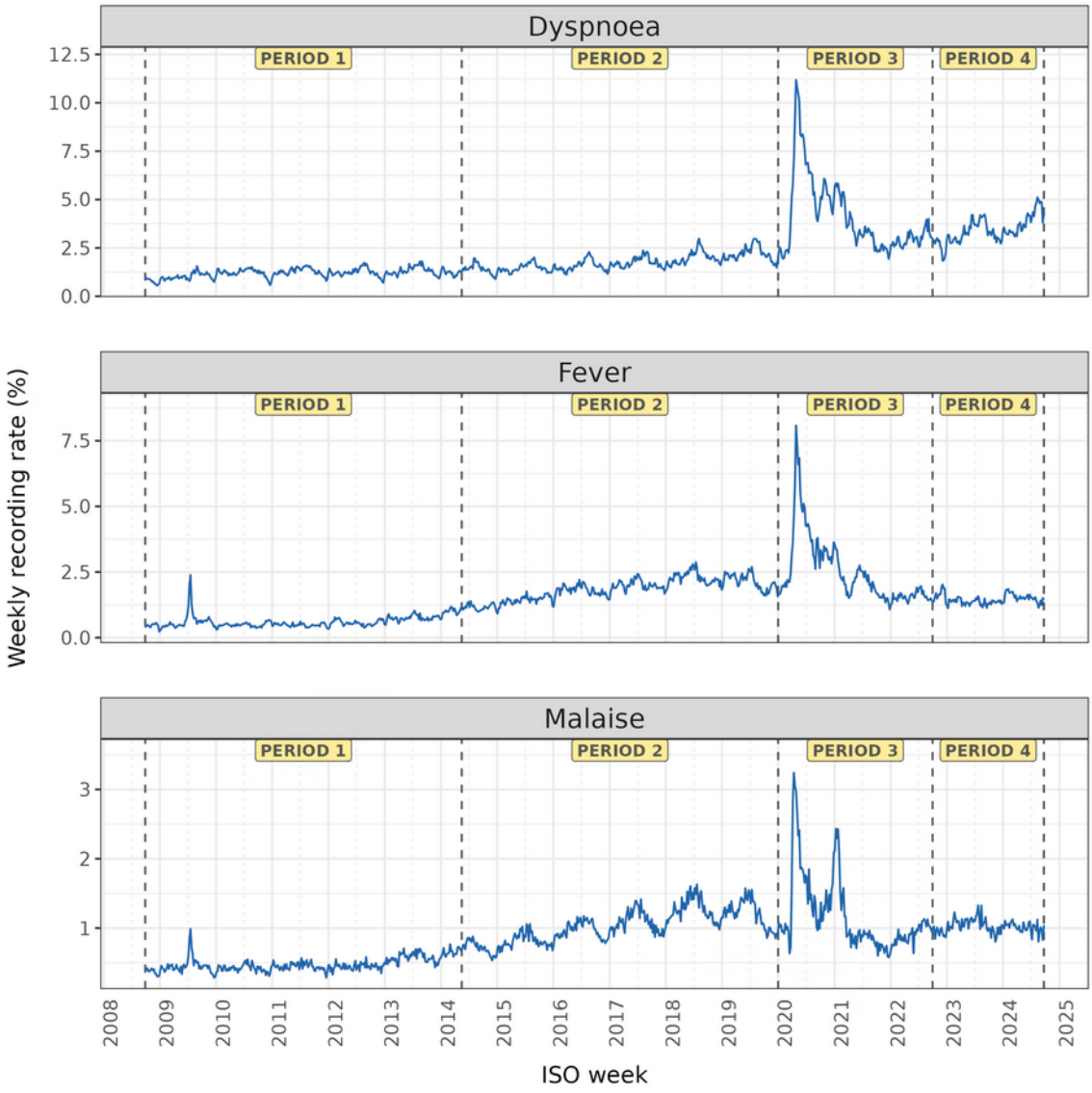


Figure A3.6: Time series of weekly symptom (Dyspneoa- top panel, fever- middle panel, malaise- bottom panel) recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO week: International Organization for Stadardardization.

A3.4.5 Signs

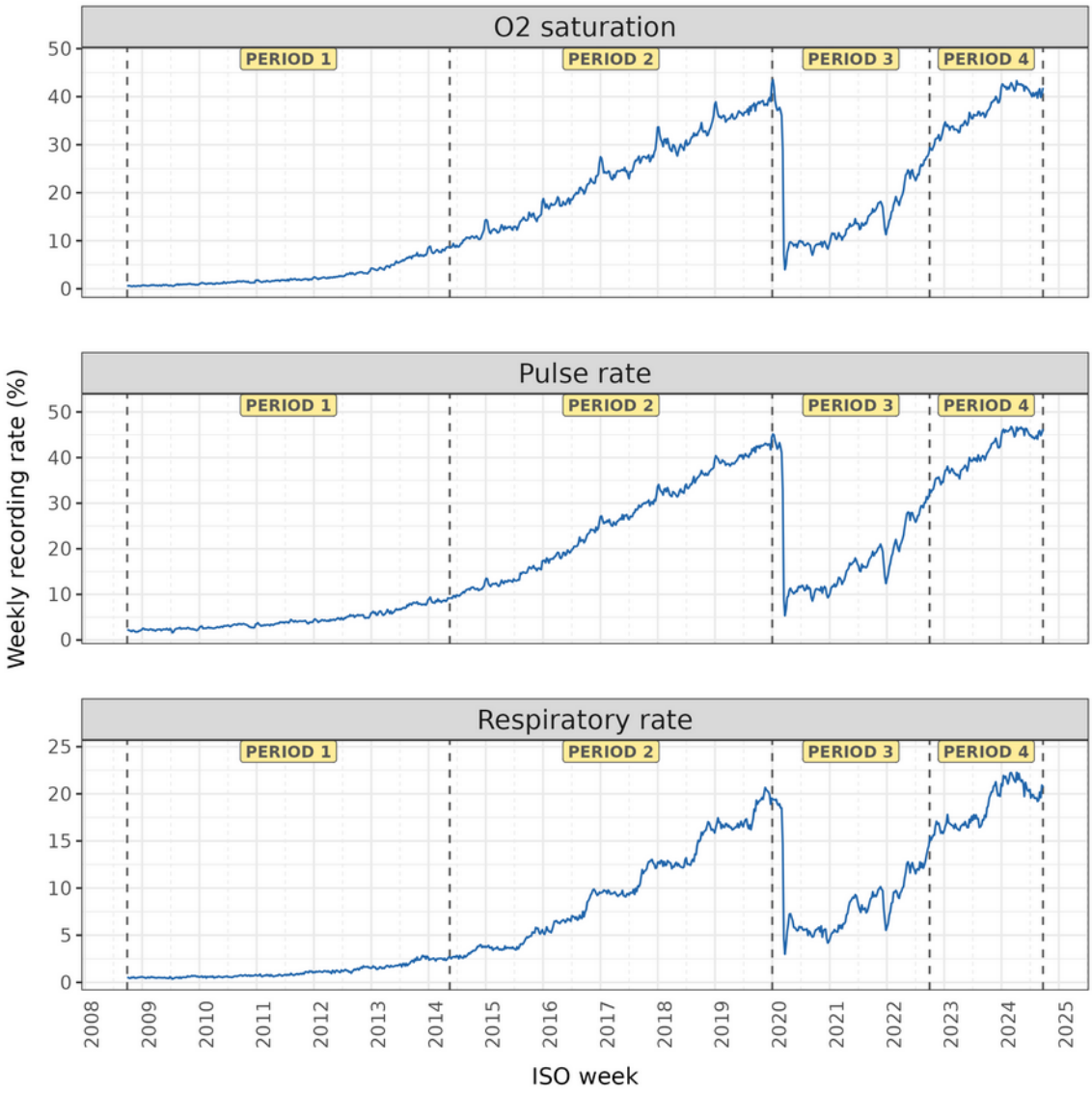


Figure A3.7: Time series of weekly sign (Oxygen saturation- top panel, pulse rate- middle panel, respiratory rate- bottom panel) recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO week: International Organization for Stadardardization.

A3.4.6 Health seeking behaviour

The overall median weekly recording rate for health seeking behaviour (any health seeking) was 1.15% (interquartile range 0.09–4.63). Within this group, NHS 111 (1.01%, IQR: 0.01–3.72) had the highest median weekly rate, followed by ambulance encounter (0.14%, IQR: 0.07–1.01). Very low weekly recording rates were observed for both ambulance

encounters and NHS 111 contacts in Period 1, with values of 0.07% (0.05–0.08) and 0.00% (0.00–0.01), respectively. A clear upward trend is evident by Period 2, and due to the very low baseline in Period 1 (particularly for NHS 111), a large 17.44-fold increase was seen in overall health seeking behaviour between Period 1 and Period 2. A further 4.87-fold increase was observed from Period 2 to Period 3 (COVID-19 pandemic), with spikes in ambulance encounters seen during the first and second waves of the pandemic. Recording of ambulance encounters in Period 4 has been higher than the pre-pandemic period (Period 2).

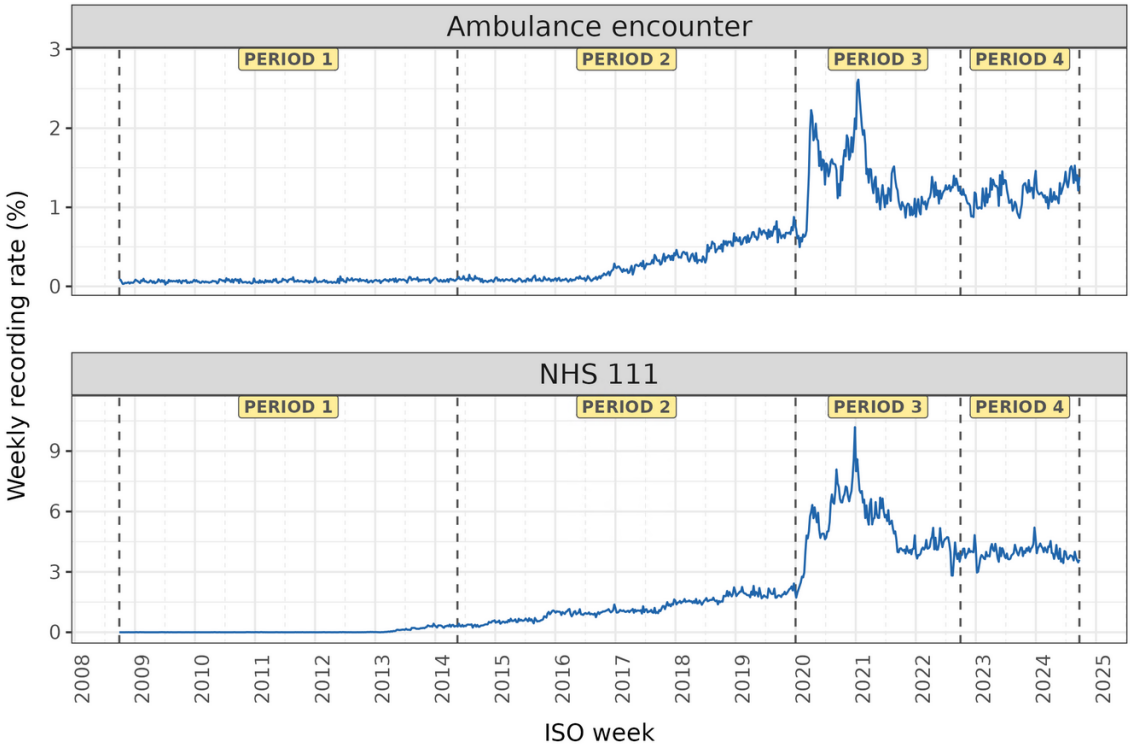


Figure A3.8: Time series of weekly health seeking (Ambulance encounter- top panel, NHS 111 encounter- bottom panel) recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO week: International Organization for Stadardardization.

A3.4.7 Clinical scores

The median weekly recording rate for all clinical scores was 2.05% (IQR: 0.19–3.24). Median recording rates for individual scores were generally much lower than those for clinical signs (31.95%, IQR: 18.75-51.81). AVPU had the highest median weekly rate (0.96%, IQR: 0.18–1.96), followed by Centor (0.74%, IQR: 0.00–1.12) and CRB65 (0.02%, IQR:

0.00–0.07). Centor scores were not recorded prior to April 2015, but use increased substantially during Period 2, displaying a seasonal pattern with peak recording rates occurring mid-year. The National Early Warning Score 2 (NEWS2) was not recorded to any meaningful extent before 2021 and remains infrequently used, with a median recording rate in Period 4 of just 0.11% (IQR: 0.09–0.14).

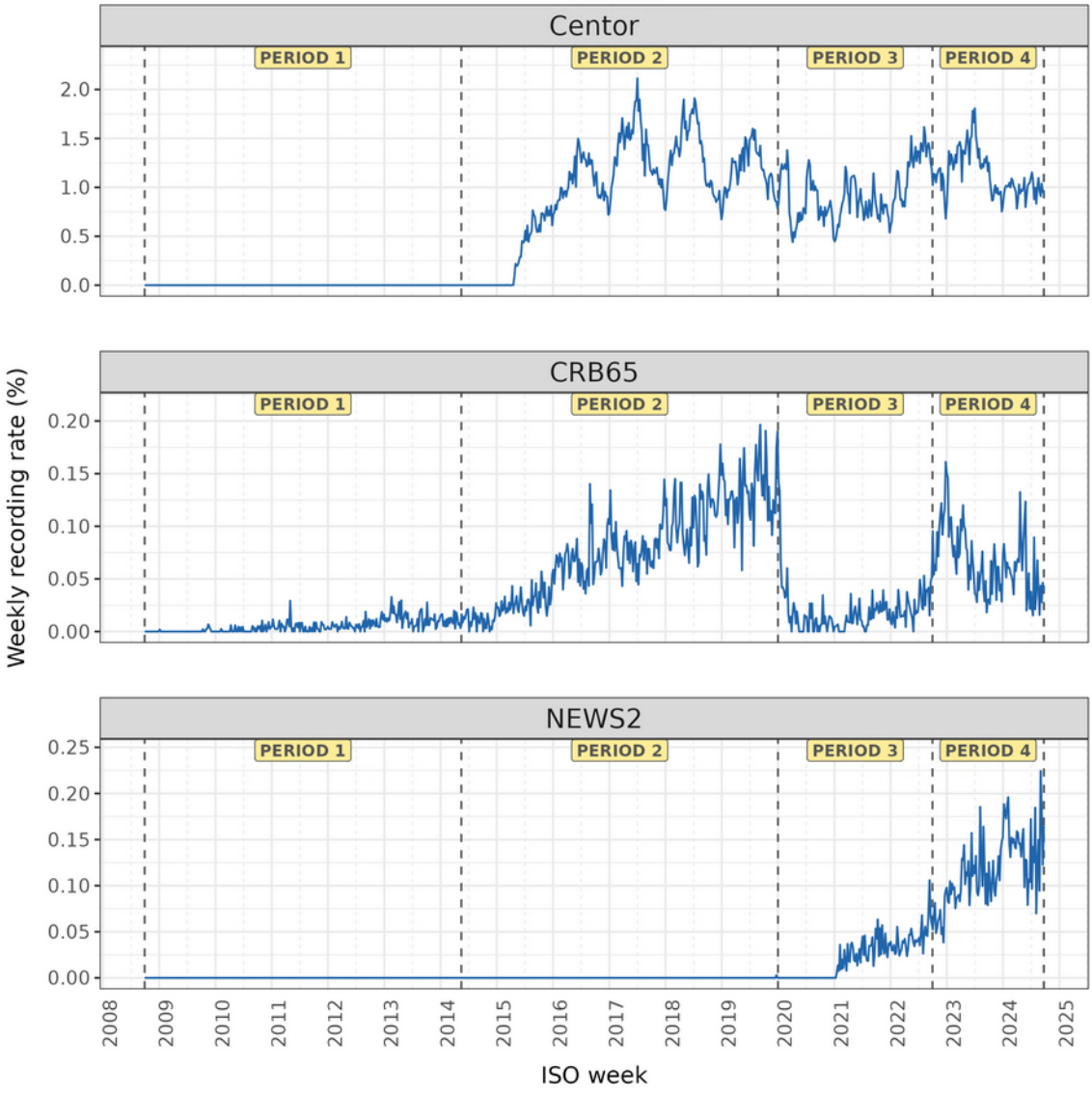


Figure A3.9: Time series of weekly clinical score (Centor- top panel, CRB65-middle panel and NEWS2-bottom panel) recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO week: International Organization for Stadardardization. Not all score time series are presented here due to very low recording rates.

A3.4.8 Investigations

The median weekly recording rate for investigations (any investigation) was 7.36% (IQR 6.36–8.41). Within this group, white blood cell count had the highest median weekly rate (5.61%, IQR: 4.91–6.33), followed by CRP (2.21%, IQR: 1.46–2.76) and chest X-ray (2.14%, IQR: 1.69–2.52). There was a modest 1.27-fold increase in overall investigation recording from Period 1 to Period 2. During the COVID-19 pandemic (Period 3), a slight decline was observed (fold-change 0.92). Peaks in the recording of all three investigations were seen during the first and second COVID-19 waves. Some seasonal variation is also evident, with investigation recording rates typically dipping in late December, around the Christmas holiday period.

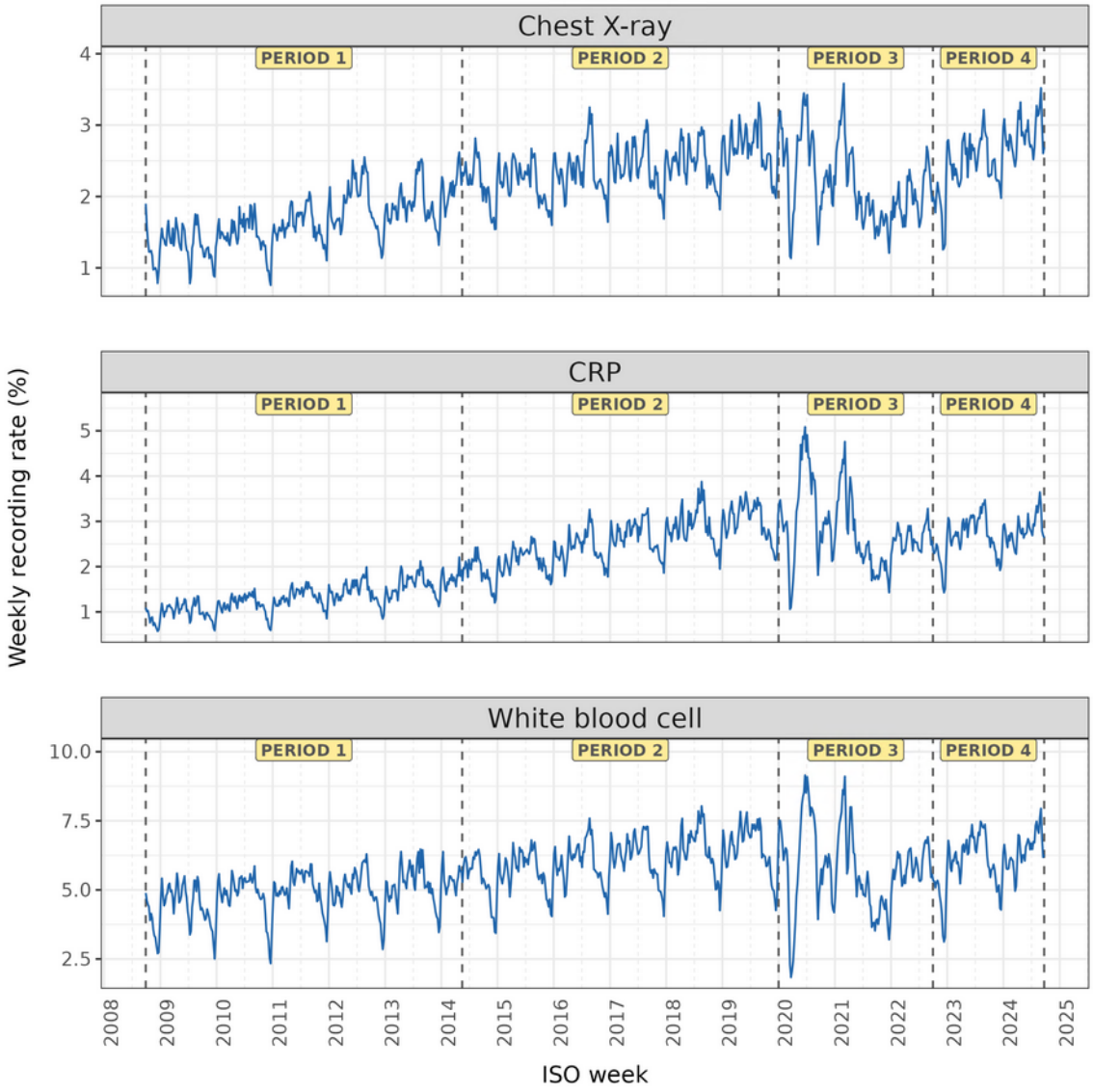


Figure A3.10: Time series of weekly investigation (Chest X-ray- top panel, C-reactive protein (CRP)-middle panel and white blood cell count-bottom panel) recording rates within 56 days of an ARI episode across the study period. Period 1: Early pre-pandemic; Period 2: Late pre-pandemic; Period 3: COVID-19 pandemic; Period 4: Post-pandemic. ISO week: International Organization for Stadardardization. Not all score time series are presented here due to very low recording rates.

Appendix A4

Chapter 5 Severity marker evaluation

A4.1 MATHEMATICAL DEFINITION OF CROSS CORRELATION

Table A4.1: Box 6.2: Mathematical definition of cross-correlation

Cross-correlation

Definition

$$r_{XY}(k) = \frac{\sum_{t=1}^{n-k} \left(X_t - \bar{X}\right)\left(Y_{t+k} - \bar{Y}\right)}{\sqrt{\sum_{t=1}^{n-k} \left(X_t - \bar{X}\right)^2} \sqrt{\sum_{t=1}^{n-k} \left(Y_{t+k} - \bar{Y}\right)^2}}$$

Symbols

$r_{XY}(k)$: cross-correlation at lag k.

$\bar{X}, \bar{Y}$: sample means of X and Y.

k: lag (may be positive or negative).

Interpretation

Positive k correlates X_t with future values Y_{t+k}; negative k correlates X_t with past values of Y. At $k = 0$, $r_{XY}(0)$ equals the Pearson correlation between X and Y.

Notes: The summation limits are written for $k \geq 0$; for negative k, choose limits so indices remain valid (equivalently, swap the roles of X and Y or write $r_{XY}(k) = r_{YX}(-k)$).

A4.2 FOREST PLOTS BY INFECTION TYPE AND AGE STRATA

Non-pandemic URTI (≤15 years)

Individual-level analysis: See Figure A4.1.

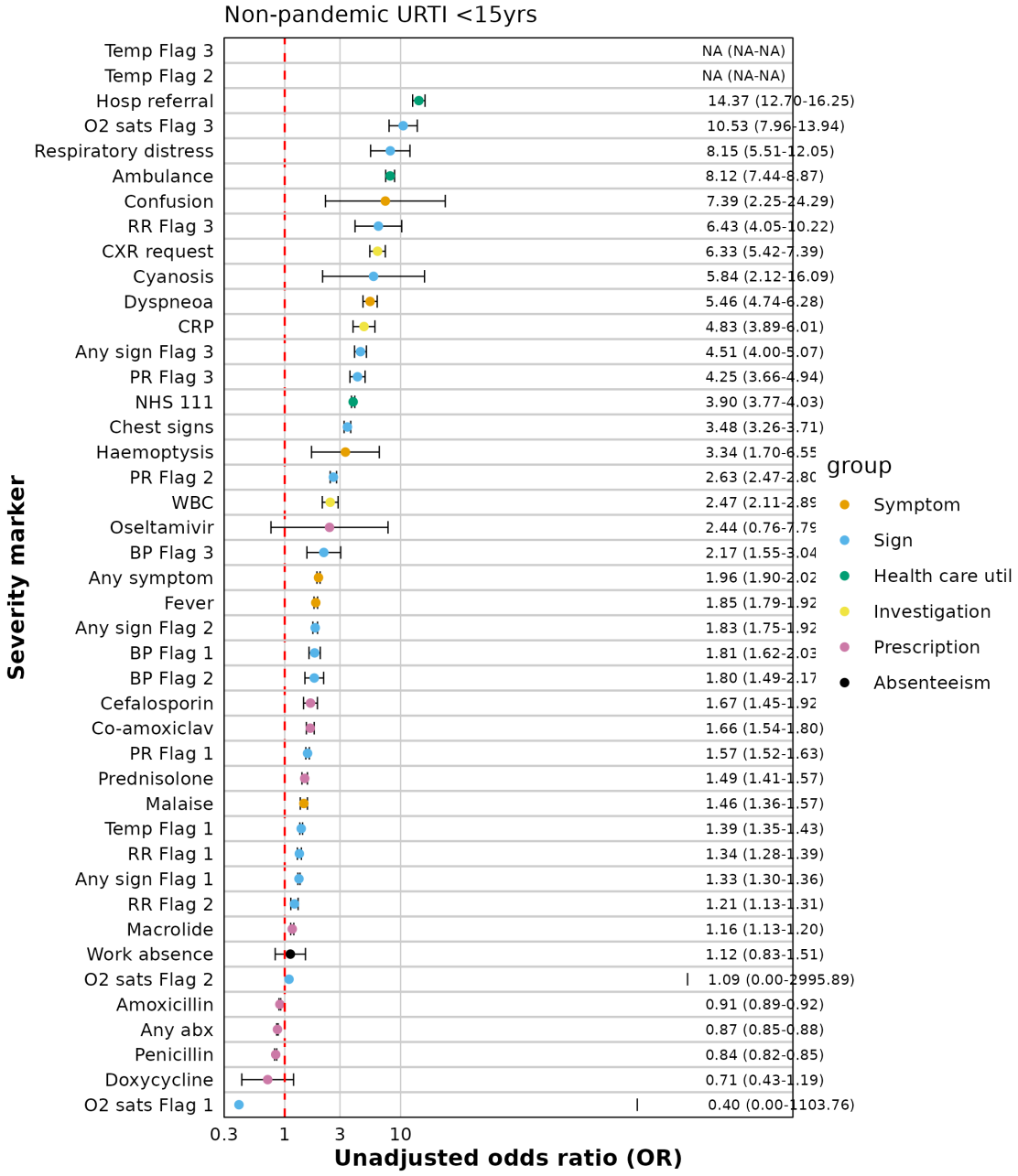


Figure A4.1: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic URTI (≤15 years)

Aggregate-level analysis: See Figure A4.2.

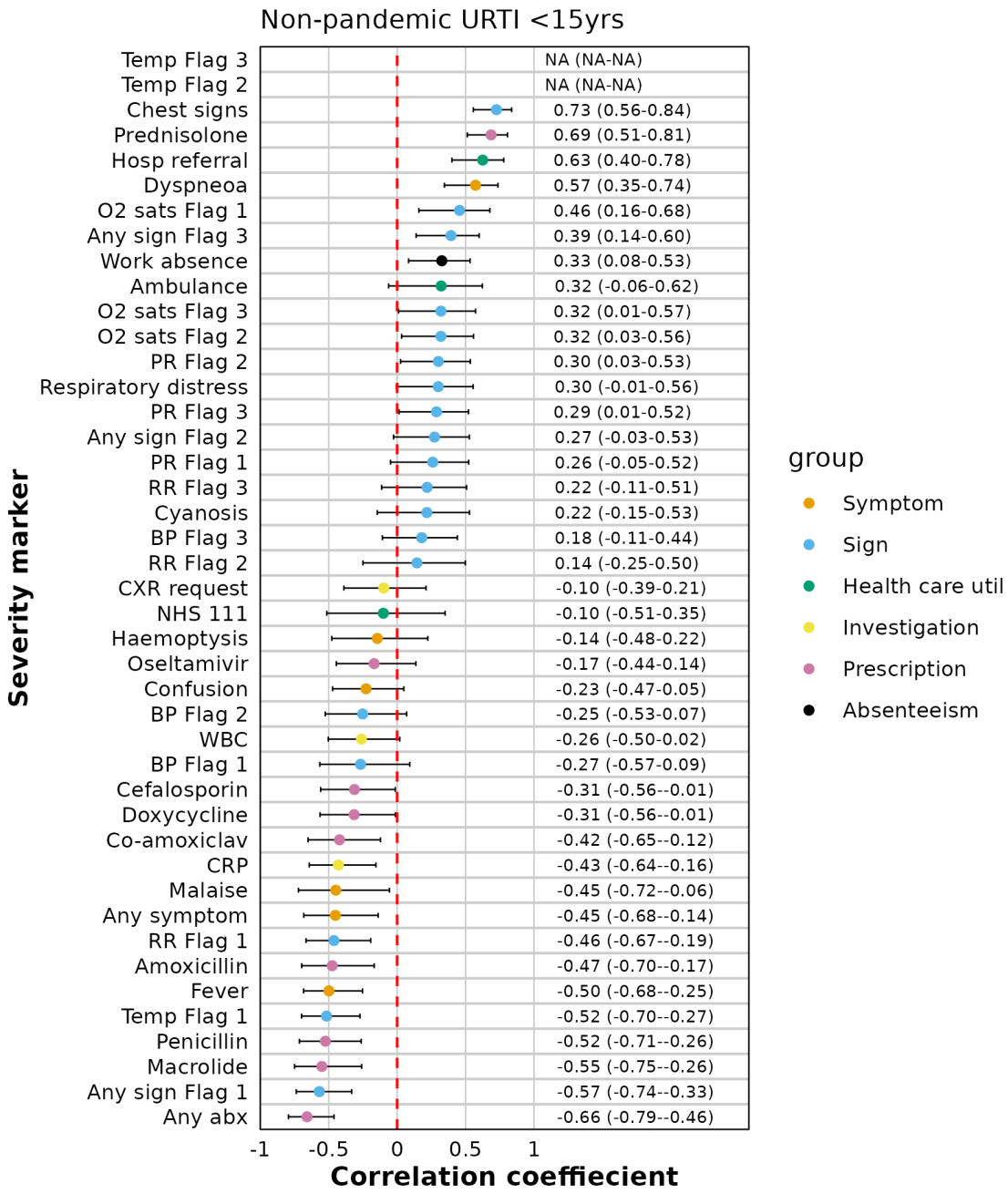


Figure A4.2: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic URTI (≤15 years)

Non-pandemic URTI (15–64 years)

Individual-level analysis: See Figure A4.3.

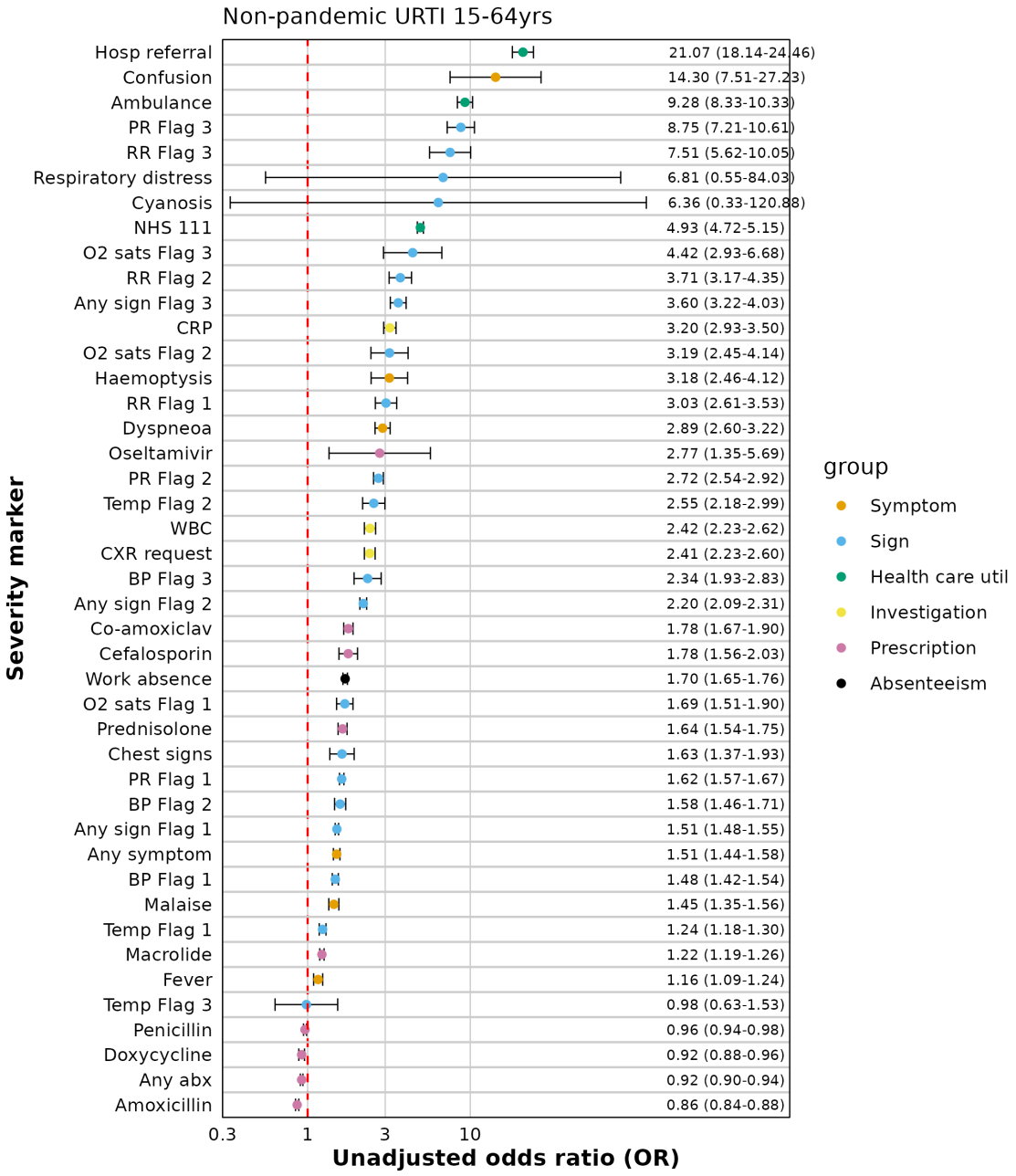


Figure A4.3: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic URTI (15–64 years)

Aggregate-level analysis: See Figure A4.4.

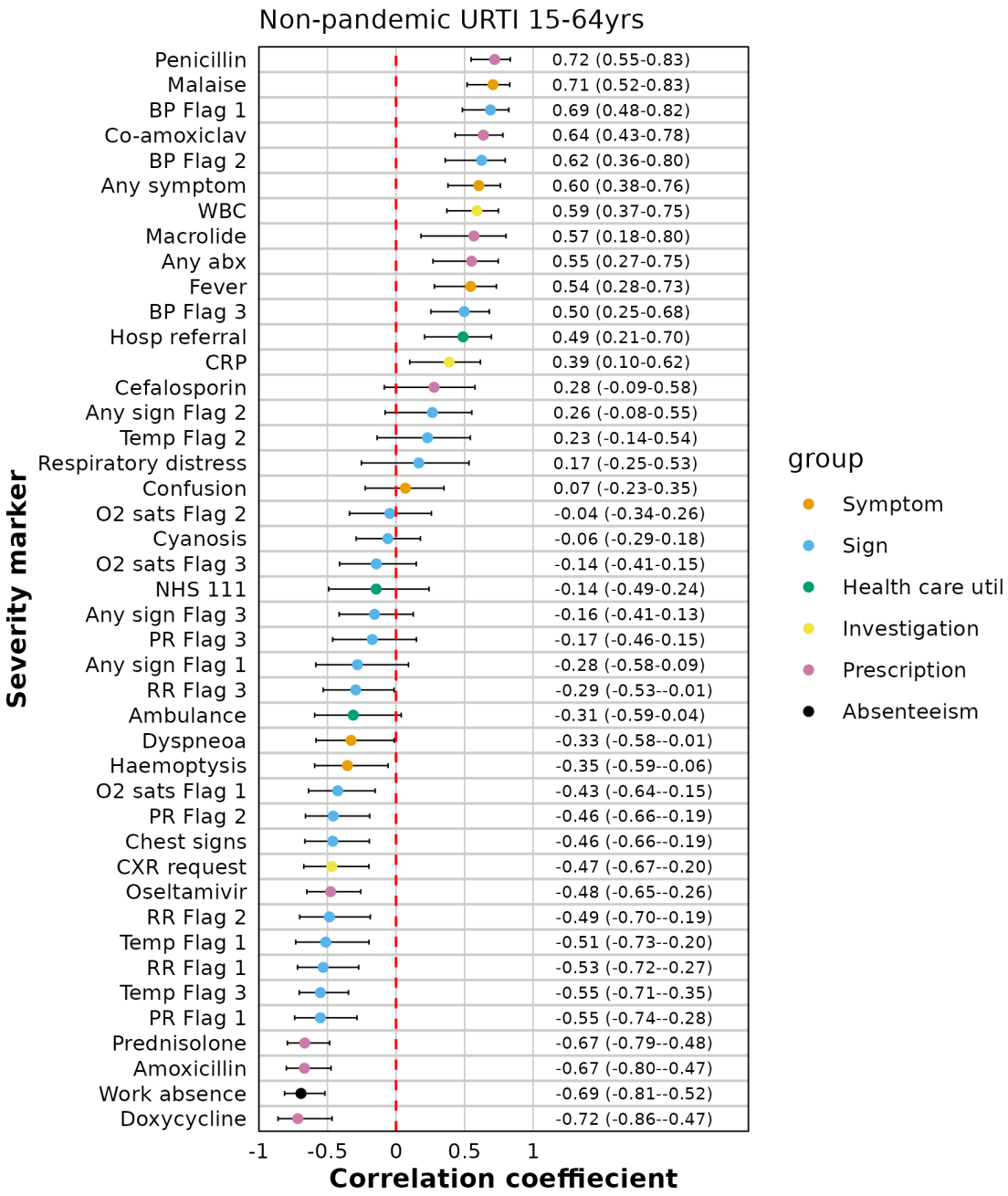


Figure A4.4: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic URTI (15–64 years)

Non-pandemic URTI (≥65 years)

Individual-level analysis: See Figure A4.5.

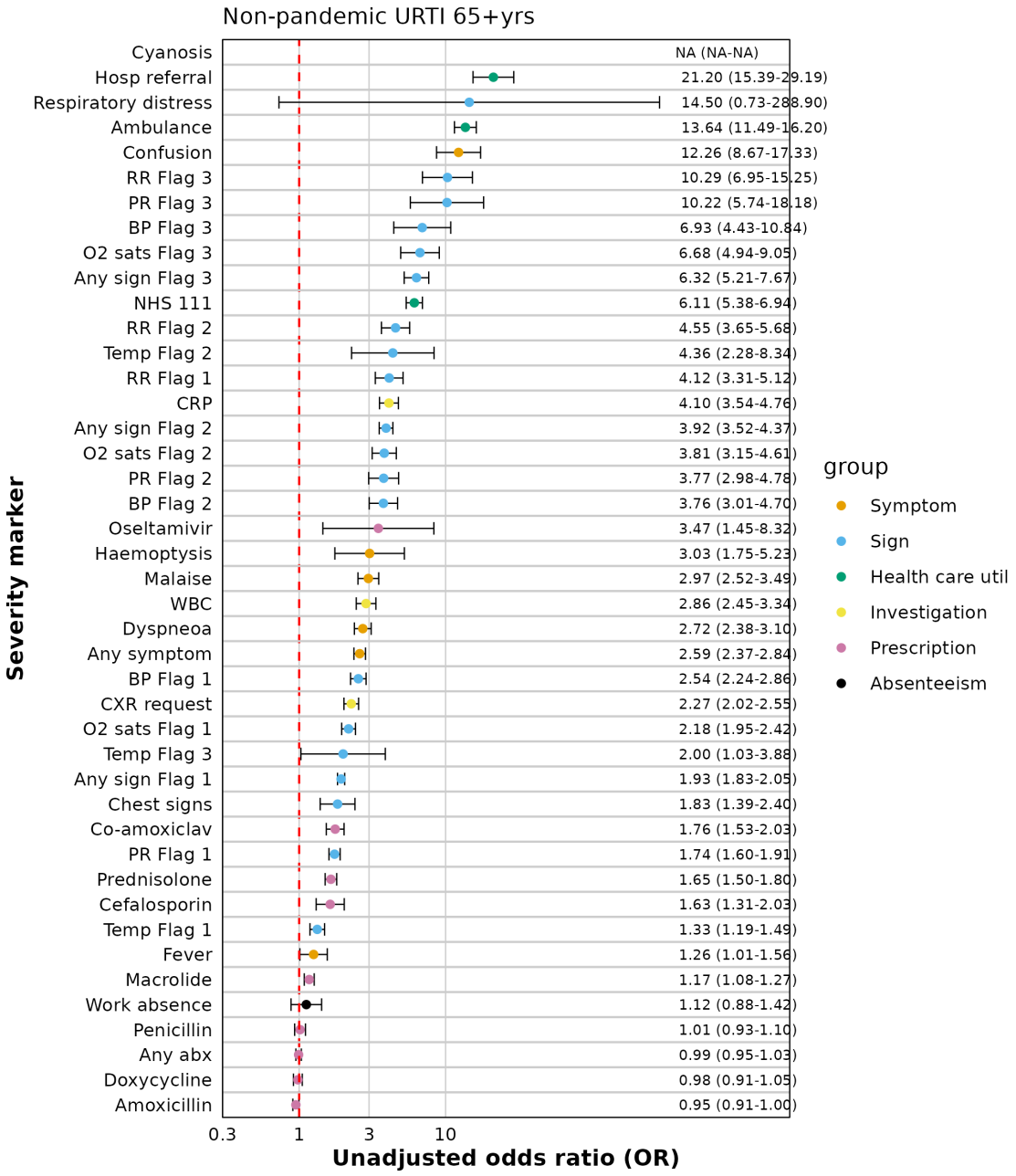


Figure A4.5: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic URTI (≥65 years)

Aggregate-level analysis: See Figure A4.6.

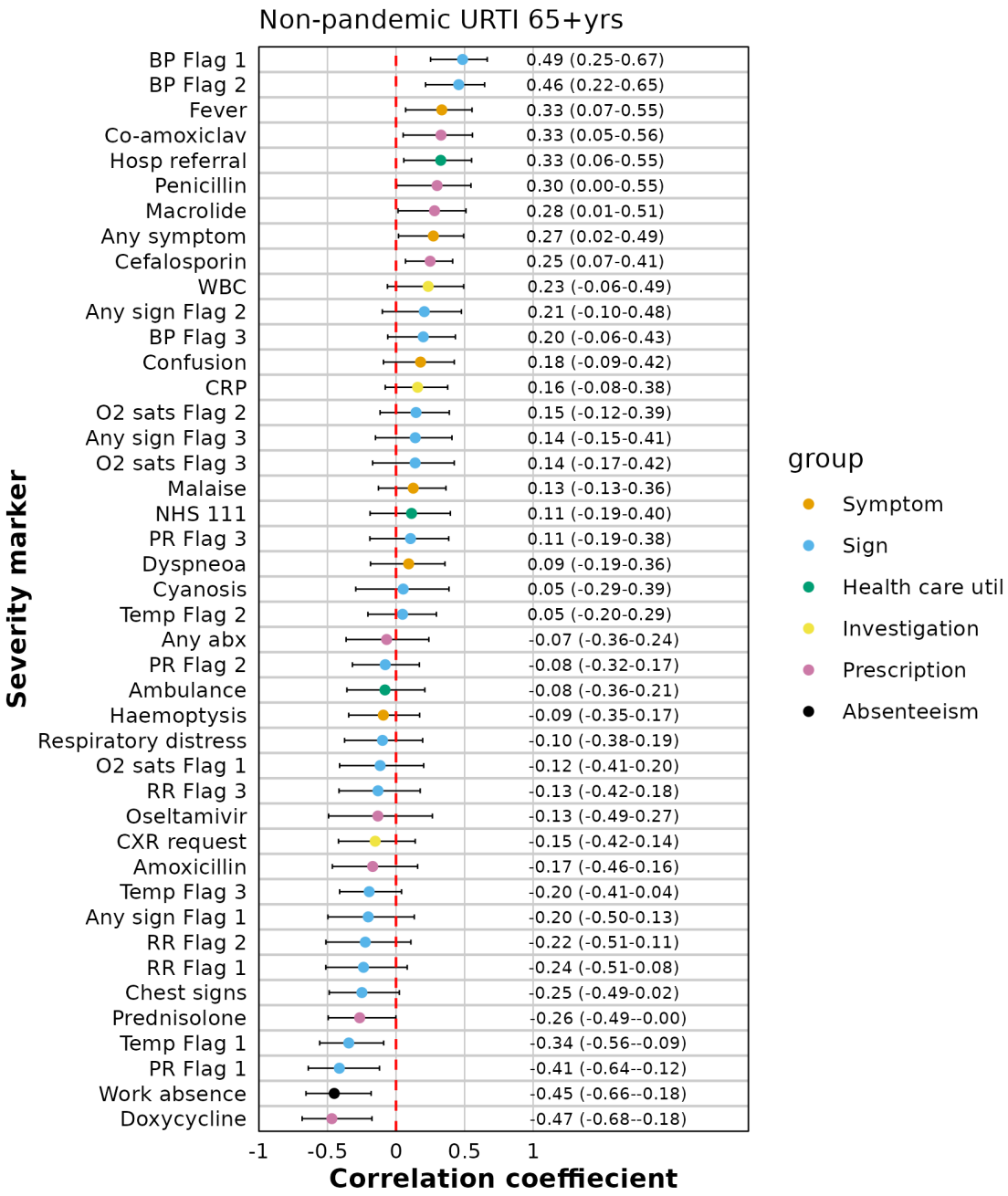


Figure A4.6: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic URTI (≥65 years)

Non-pandemic LRTI with ECLD (≤15 years)

Individual-level analysis: See Figure A4.7.

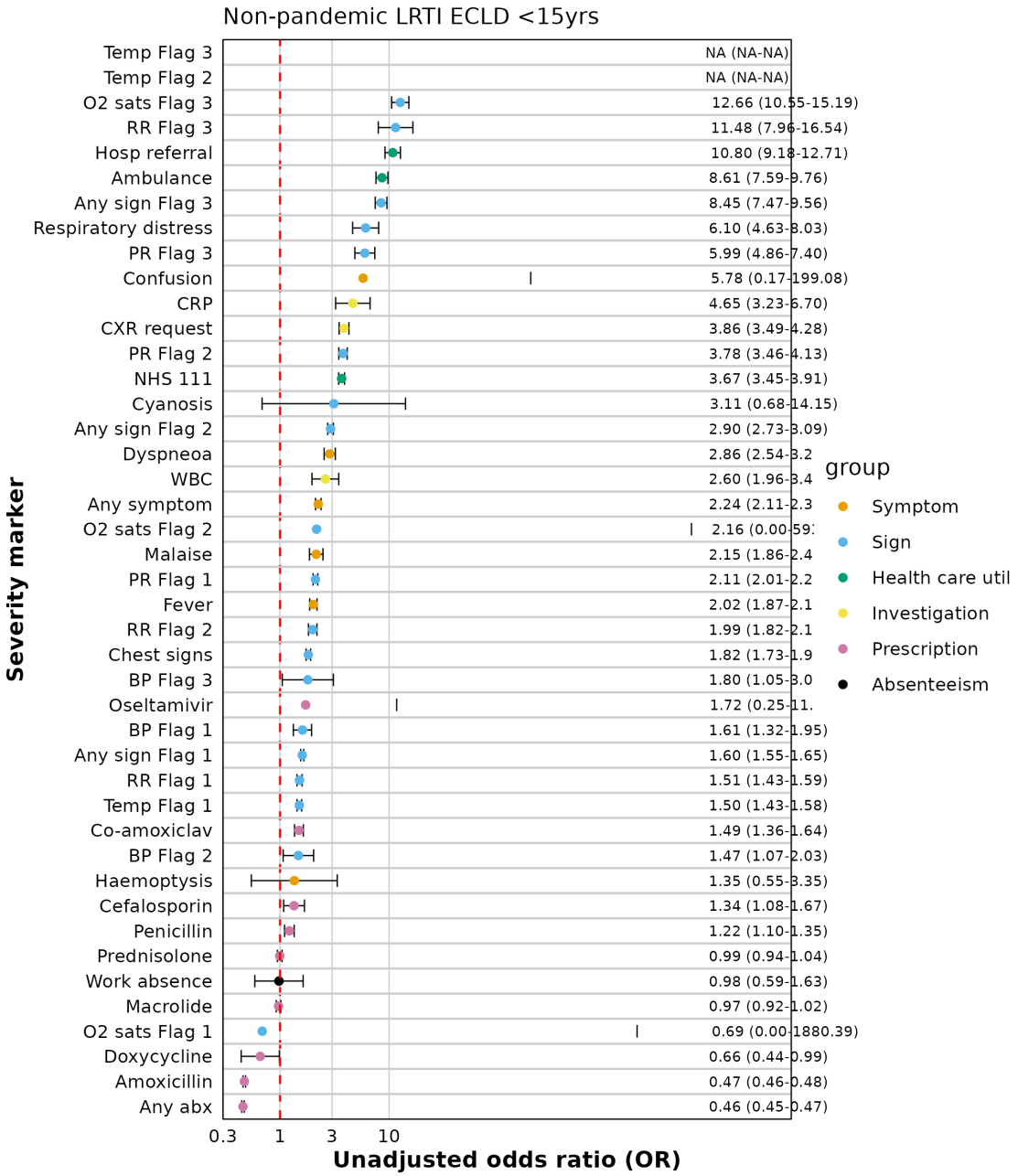


Figure A4.7: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic LRTI ECLD (≤15 years)

Aggregate-level analysis: See Figure A4.8.

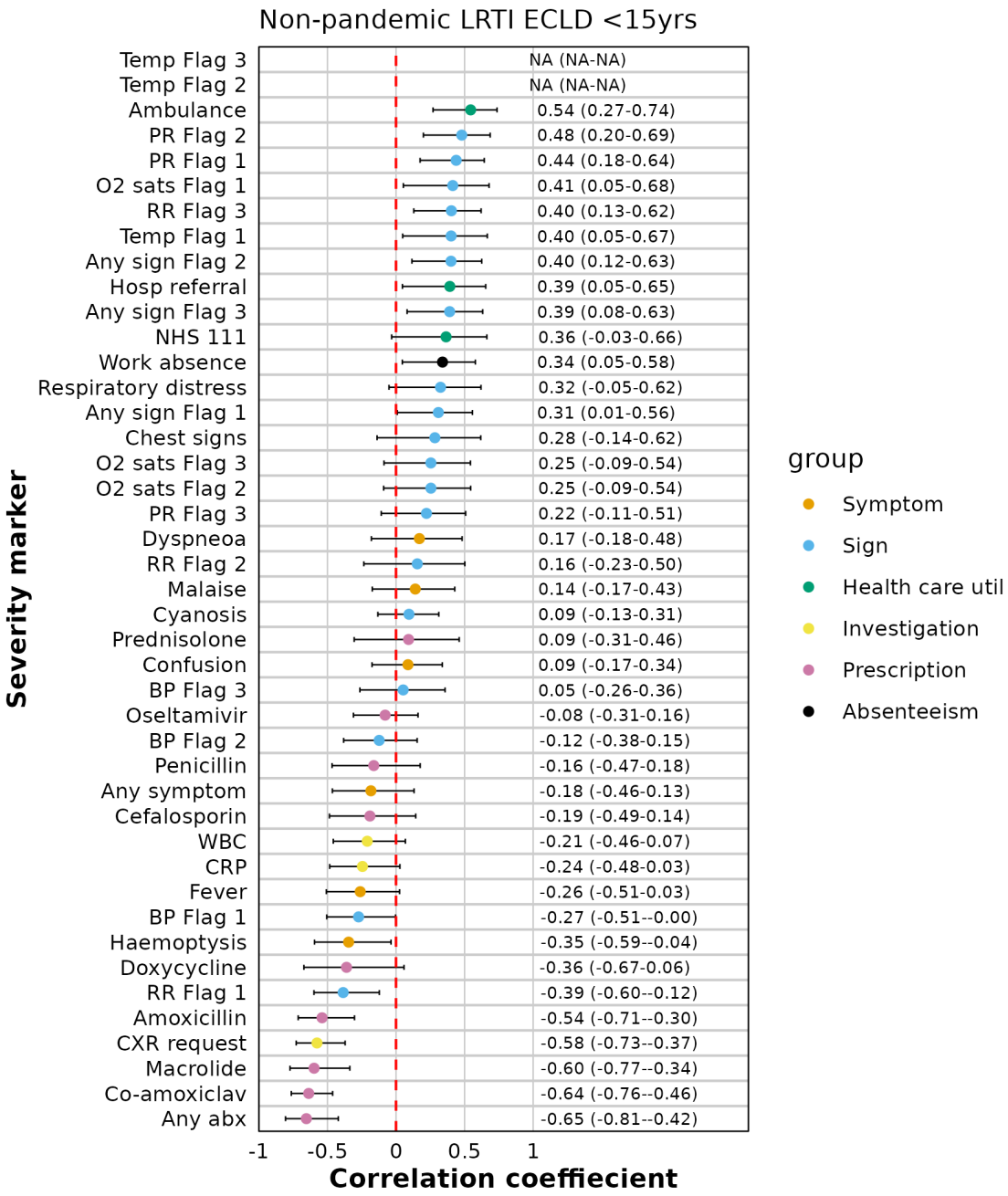


Figure A4.8: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic LRTI ECLD (≤15 years)

Non-pandemic LRTI with ECLD (15–64 years)

Individual-level analysis: See Figure A4.9.

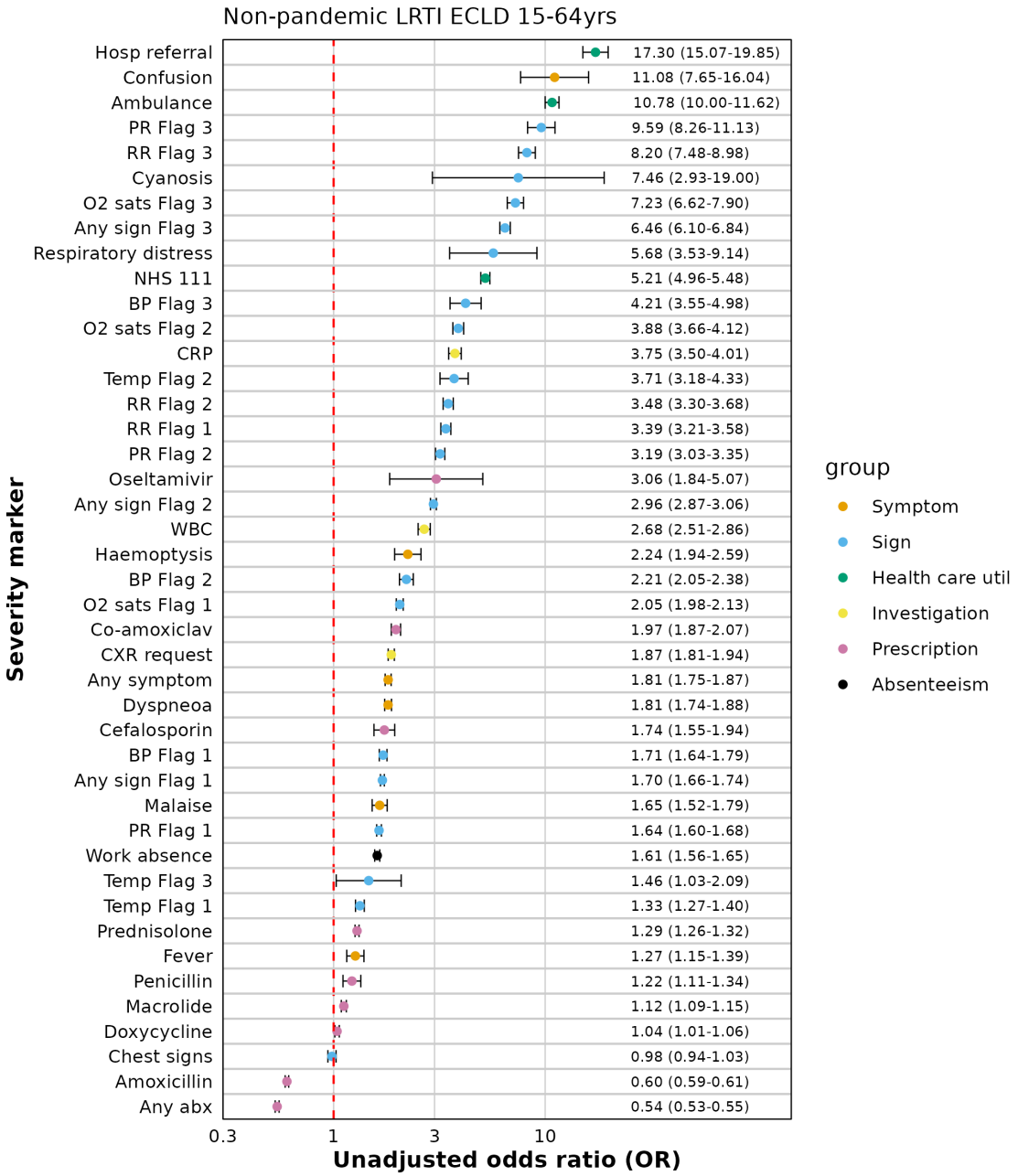


Figure A4.9: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic LRTI ECLD (15–64 years)

Aggregate-level analysis: See Figure A4.10.

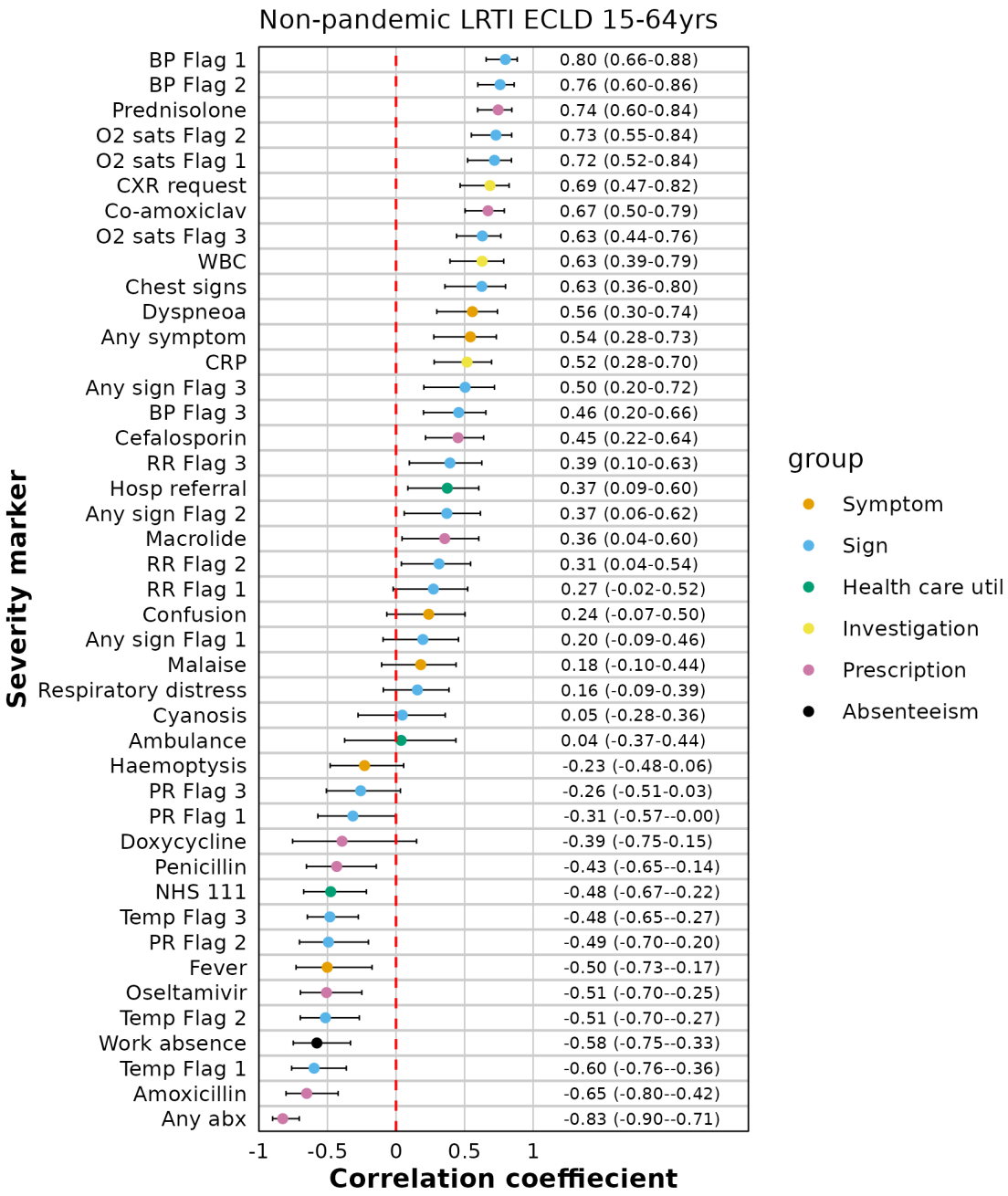


Figure A4.10: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic LRTI ECLD (15–64 years)

Non-pandemic ILI (≤15 years)

Individual-level analysis: See Figure A4.11.

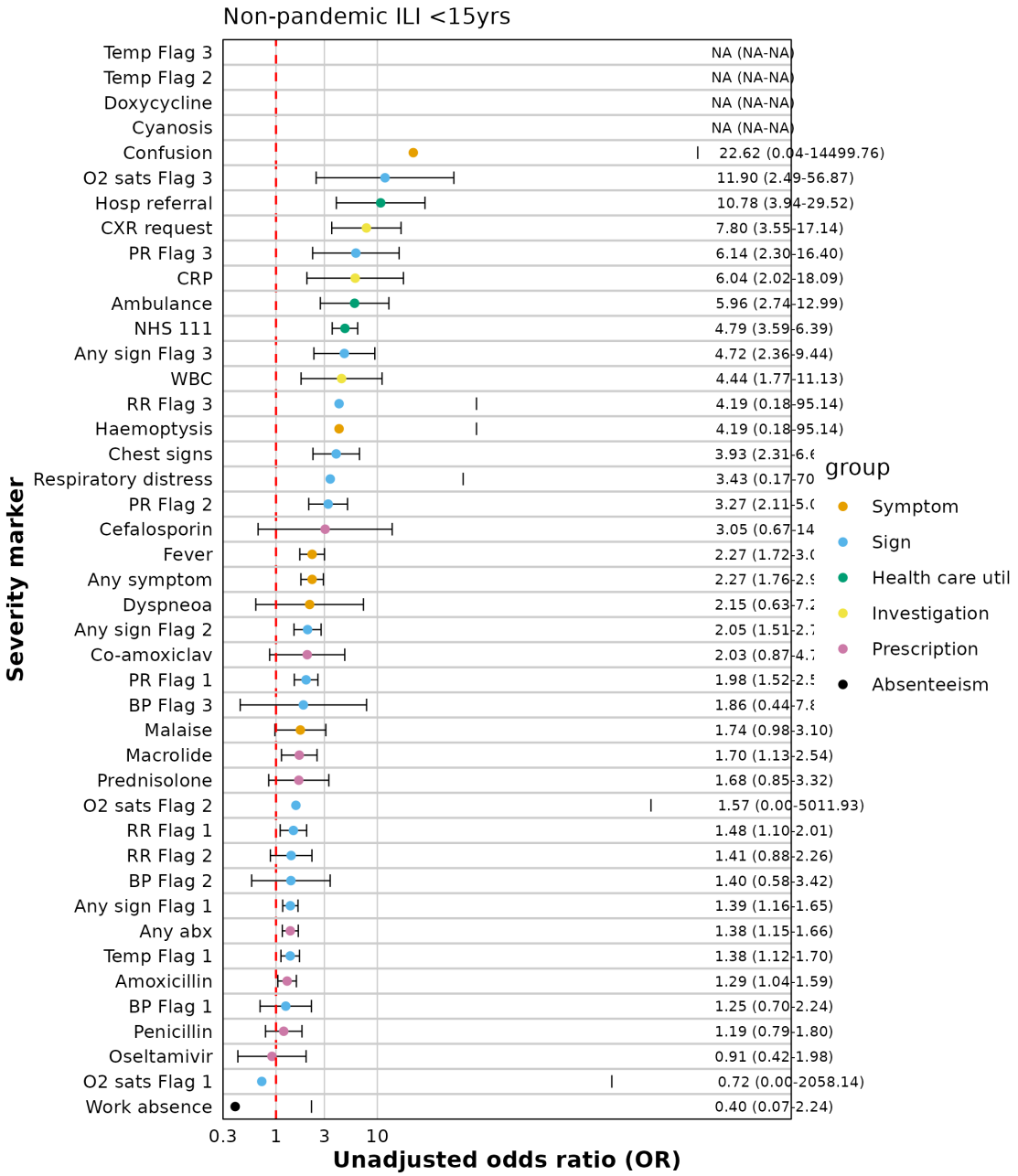


Figure A4.11: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic ILI (≤15 years)

Aggregate-level analysis: See Figure A4.12.

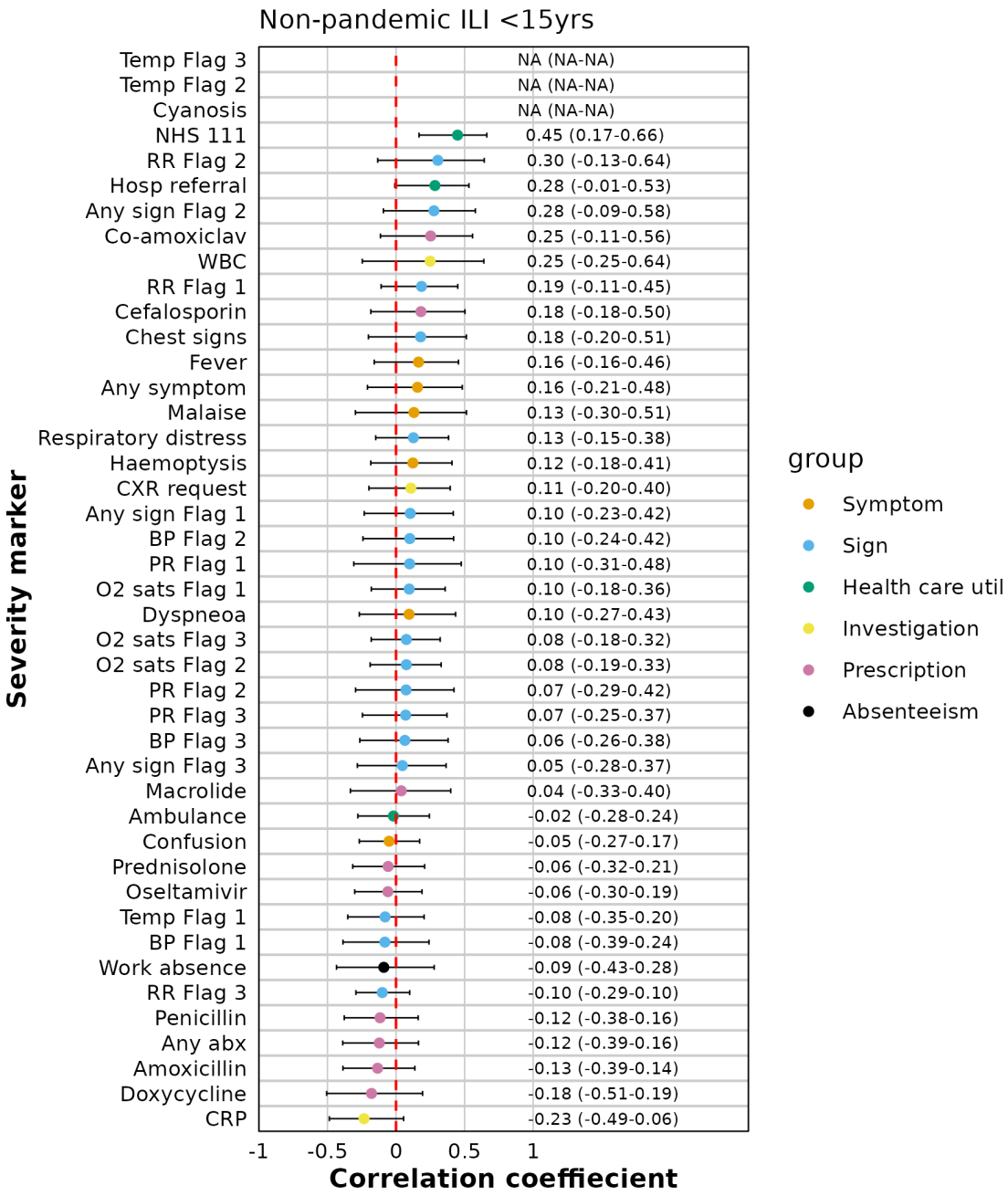


Figure A4.12: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic ILI (≤15 years)

Non-pandemic ILI (15–64 years)

Individual-level analysis: See Figure A4.13.

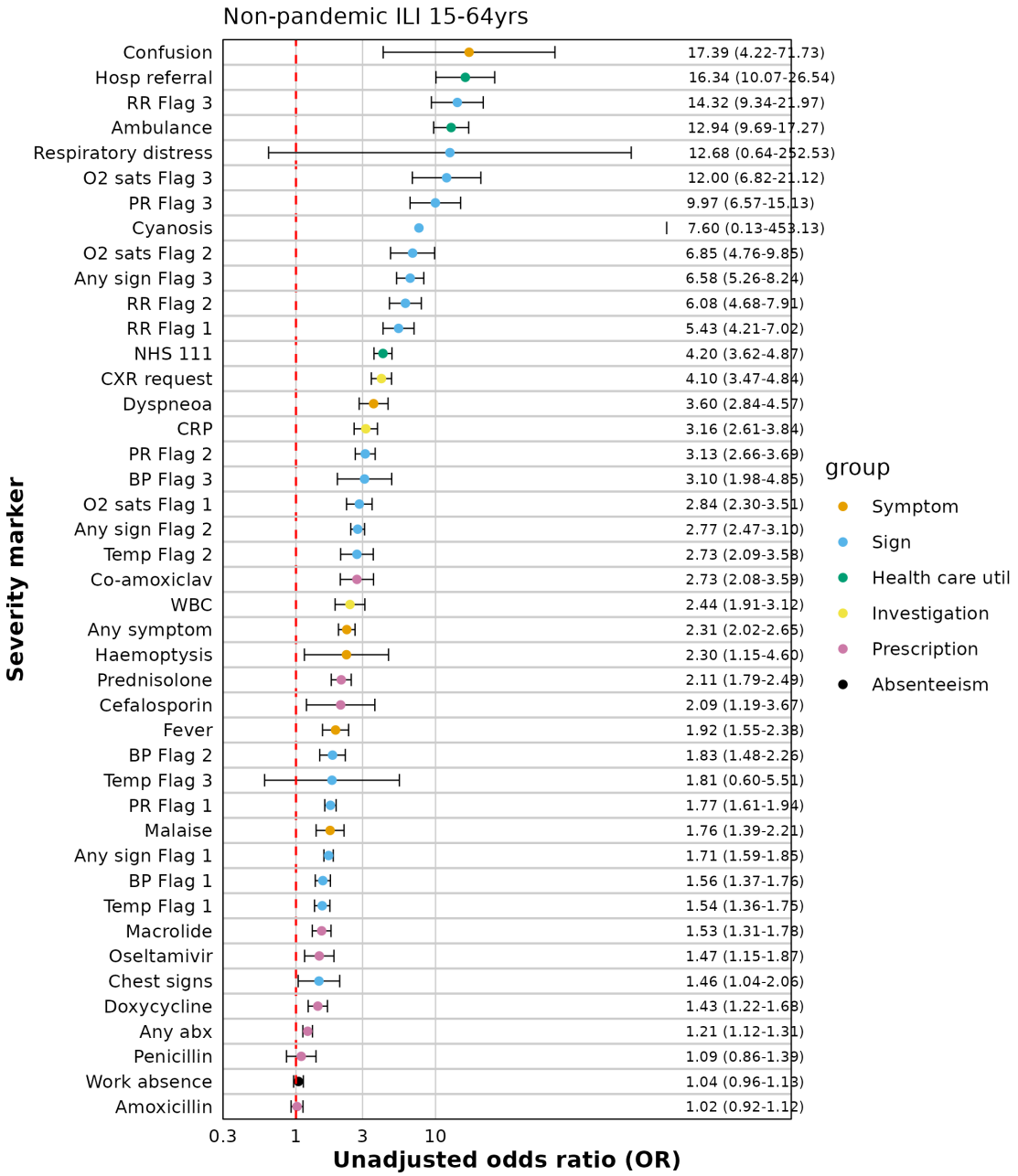


Figure A4.13: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic ILI (15–64 years)

Aggregate-level analysis: See Figure A4.14.

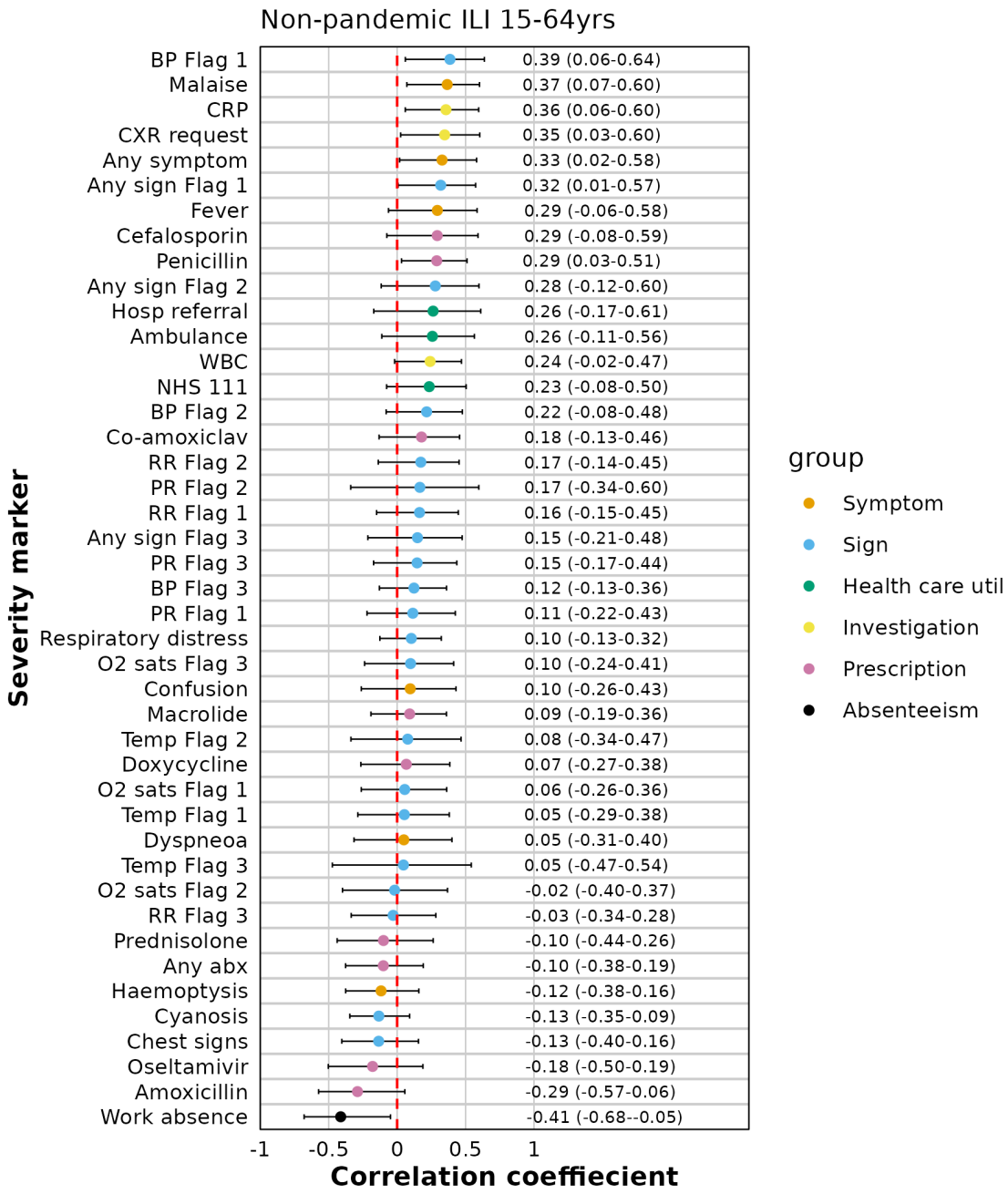


Figure A4.14: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic ILI (15–64 years)

Non-pandemic ILI (≥65 years)

Individual-level analysis: See Figure A4.15.

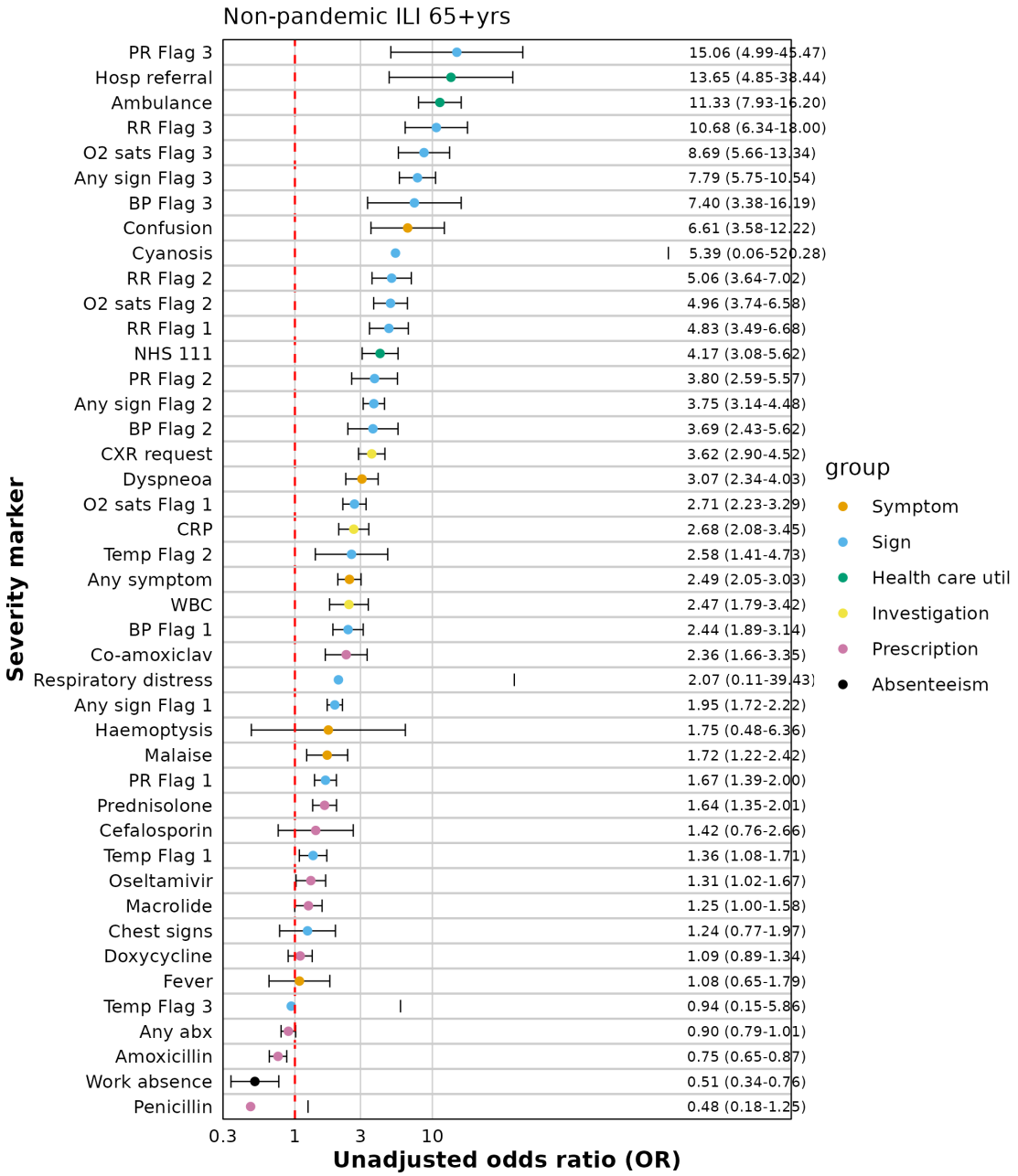


Figure A4.15: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic ILI (≥65 years)

Aggregate-level analysis: See Figure A4.16.

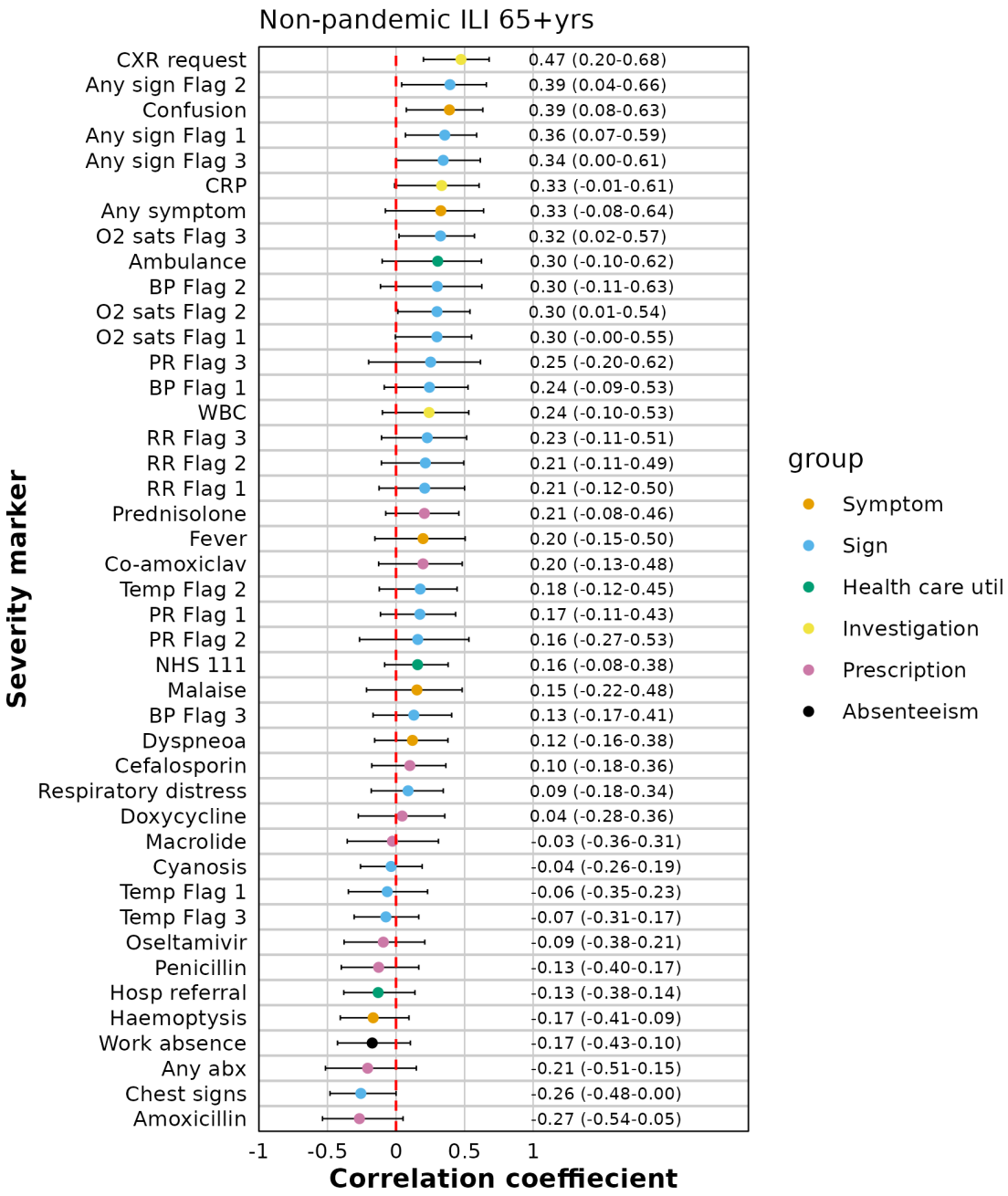


Figure A4.16: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death for non-pandemic ILI (≥65 years)

Pandemic suspected COVID-19 (15–64 years)

Individual-level analysis: See Figure A4.17.

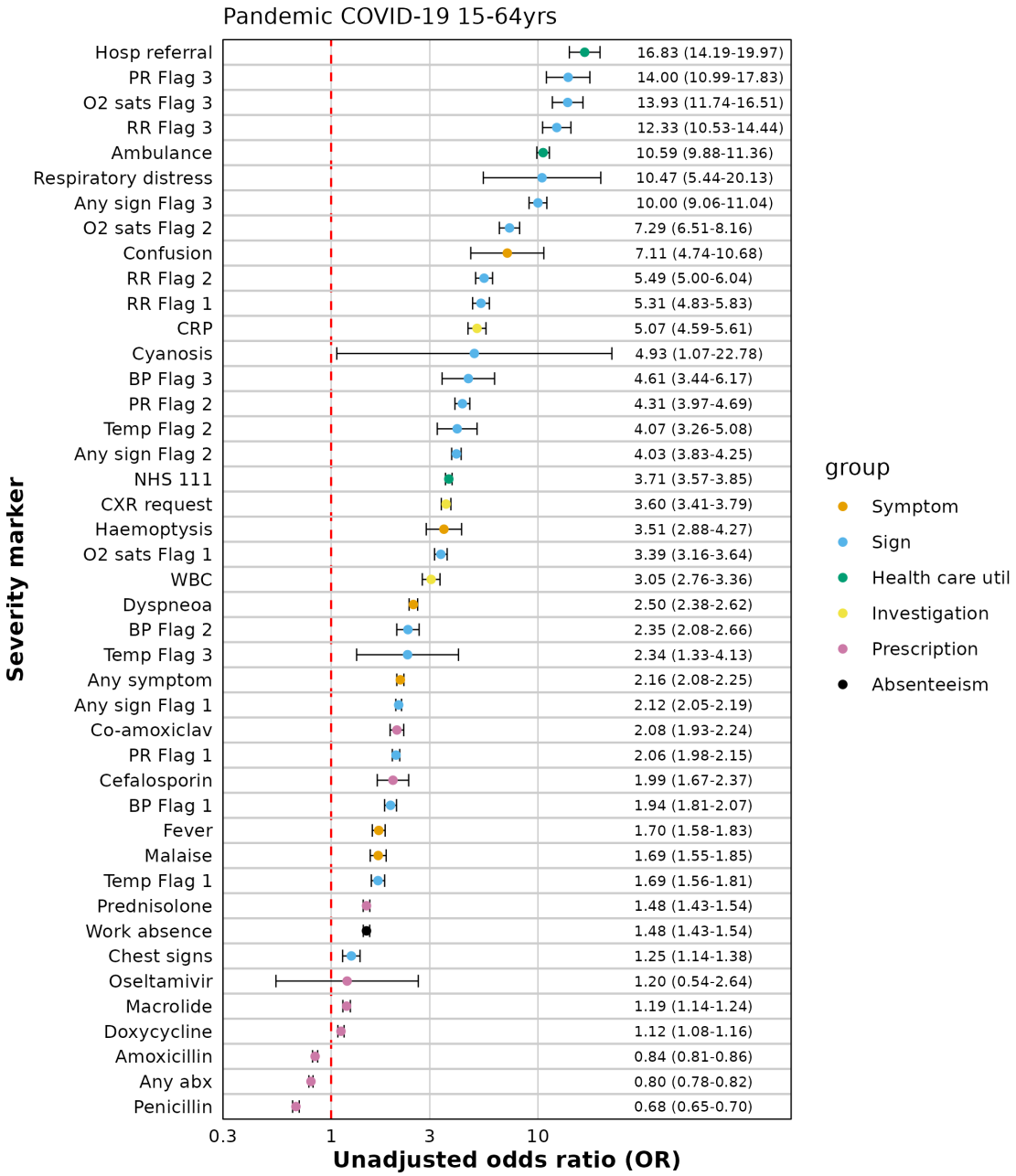


Figure A4.17: Forest plot showing relationship between severity markers and composite outcome of hospitalisation, complications, ICU admission or death for pandemic suspected COVID-19 (15–64 years)

Aggregate-level analysis: See Figure A4.18.

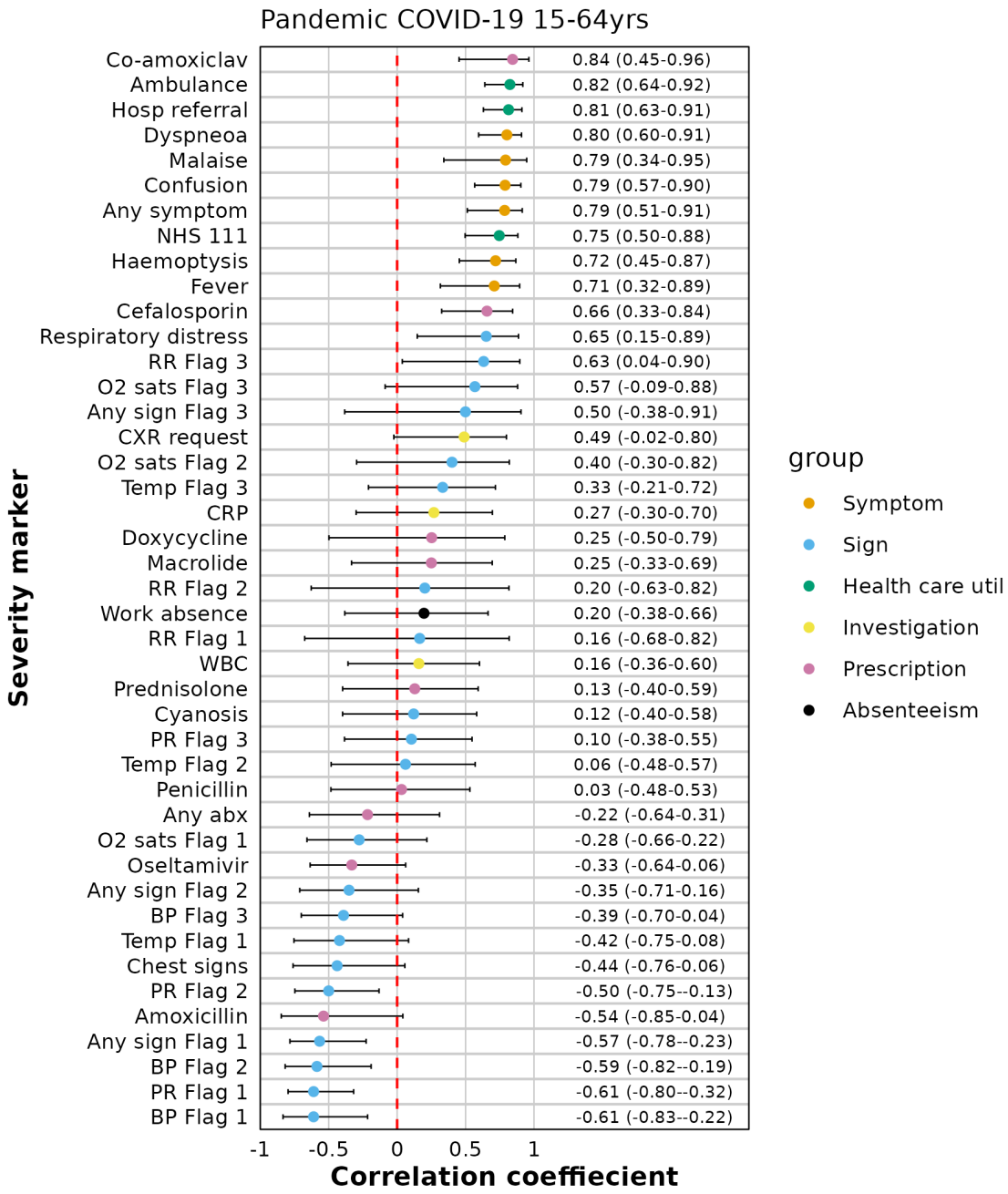


Figure A4.18: Forest plot showing relationship between severity indicators and composite outcome of hospitalisation, complications, ICU admission or death for pandemic suspected COVID-19 (15–64 years)

Appendix A5

Patient and Public Involvement (PPI)

A5.1 INTRODUCTION

The public are at the heart of this DPhil. The motivation to define operational severity indicators is ultimately to improve public health by reducing the harm that occurs as a consequence of ARIs. Not only this, but each individual member of the public in a registered RSC practice (who has not opted out of data sharing) has contributed their own data for the good of the population. It is only right, therefore, that they should have a role supporting this work. In this short chapter I present the outcome of a PPI workshop that aimed to gather public opinion on the future direction of this work including how the outputs should be communicated and shared.

A5.2 AIMS AND OBJECTIVES

A5.2.1 Aim

To gather public perspective on the future direction, communication, and accessibility of a public health research project, in order to enhance its relevance and public value.

A5.2.2 Objectives

1. To explore public views on how the research could be further developed.
2. To identify the types of outputs or resources that members of the public would find useful or engaging.

3. To gain feedback on how a lay summary of this work best communicate the findings to the public.

A5.3 METHODS

A PPI workshop was convened with contributors selected from the RSC PPI network. This network is primarily drawn from primary care patient participation groups at RSC member practices. This network has been regularly used to undertake PPI activities at the RSC. Individuals were invited via email, and adverts were also included in the weekly RSC communications to practices inviting them to share the opportunity with patients in their practice (Figure A5.1).

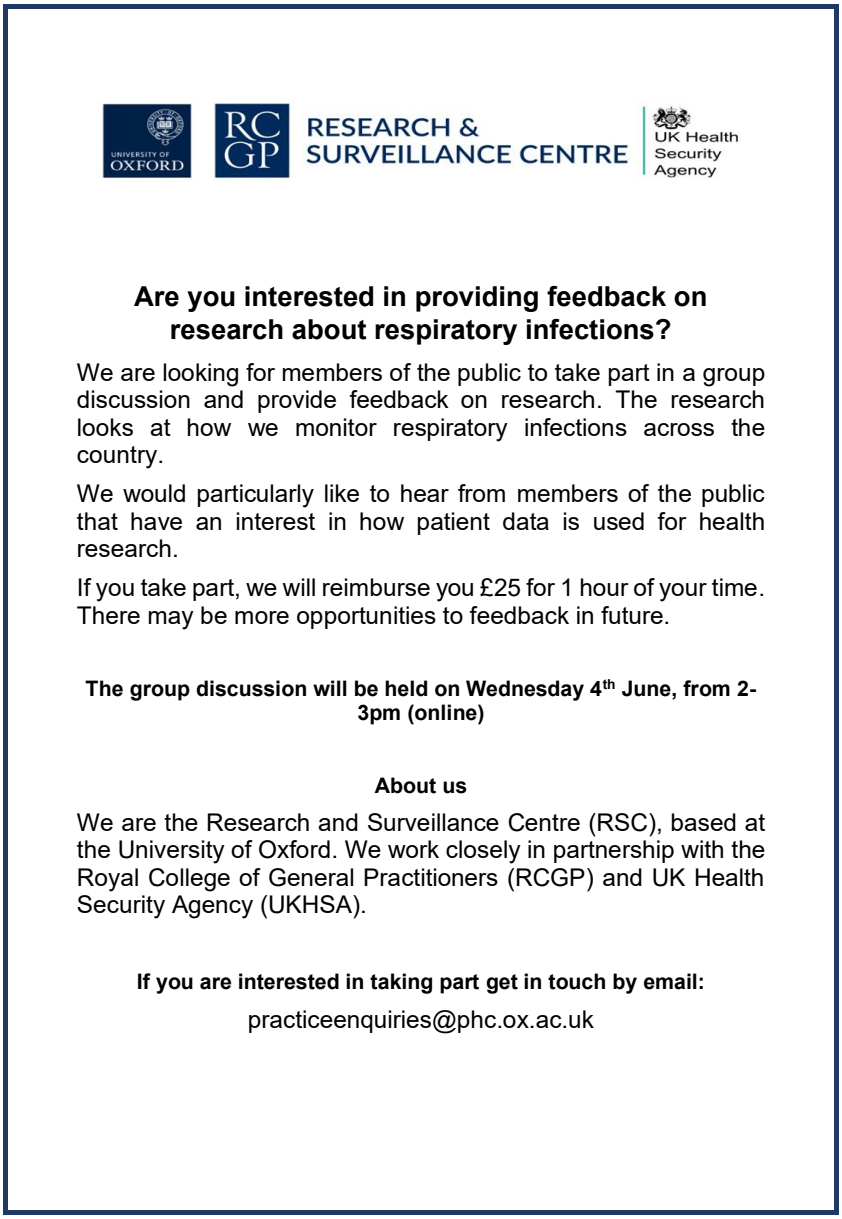
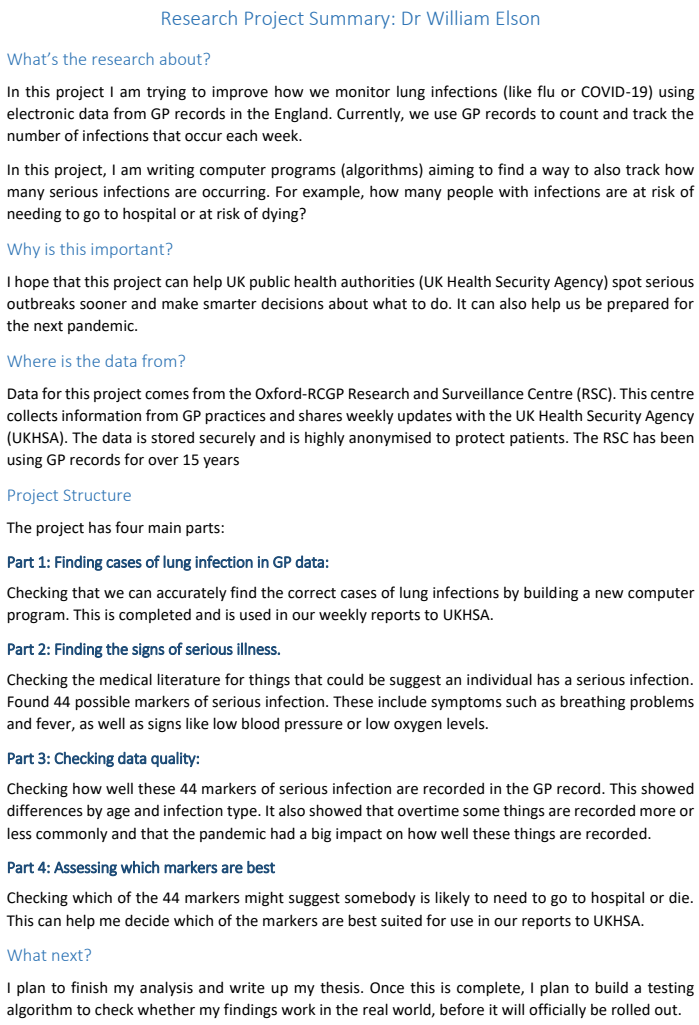


Figure A5.1: Advertisement sent to the Research and Surveillance Centre (RSC) member practices inviting public contributions to the Patient and Public Involvement (PPI) session

The session was conducted as an online workshop using Microsoft Teams. It was facilitated by a colleague at the RSC, VS, who is a member of the RSC practice liaison team and has significant experience in chairing and working in PPI over the past three years. Participants who agreed to contribute were provided with a one-page summary outlining the work completed in this DPhil and explaining the objectives of the workshop (Figure A5.2).

Four contributors (C1, C2, C3 and C4) took part in the session, representing a mixture of genders and ethnicities, and including some participants with chronic respiratory disease. No further personal details can be shared in order to protect confidentiality.



Research Project Summary: Dr William Elson

What's the research about?

In this project I am trying to improve how we monitor lung infections (like flu or COVID-19) using electronic data from GP records in the England. Currently, we use GP records to count and track the number of infections that occur each week.

In this project, I am writing computer programs (algorithms) aiming to find a way to also track how many serious infections are occurring. For example, how many people with infections are at risk of needing to go to hospital or at risk of dying?

Why is this important?

I hope that this project can help UK public health authorities (UK Health Security Agency) spot serious outbreaks sooner and make smarter decisions about what to do. It can also help us be prepared for the next pandemic.

Where is the data from?

Data for this project comes from the Oxford-RCGP Research and Surveillance Centre (RSC). This centre collects information from GP practices and shares weekly updates with the UK Health Security Agency (UKHSA). The data is stored securely and is highly anonymised to protect patients. The RSC has been using GP records for over 15 years

Project Structure

The project has four main parts:

Part 1: Finding cases of lung infection in GP data:

Checking that we can accurately find the correct cases of lung infections by building a new computer program. This is completed and is used in our weekly reports to UKHSA.

Part 2: Finding the signs of serious illness.

Checking the medical literature for things that could be suggest an individual has a serious infection. Found 44 possible markers of serious infection. These include symptoms such as breathing problems and fever, as well as signs like low blood pressure or low oxygen levels.

Part 3: Checking data quality:

Checking how well these 44 markers of serious infection are recorded in the GP record. This showed differences by age and infection type. It also showed that overtime some things are recorded more or less commonly and that the pandemic had a big impact on how well these things are recorded.

Part 4: Assessing which markers are best

Checking which of the 44 markers might suggest somebody is likely to need to go to hospital or die. This can help me decide which of the markers are best suited for use in our reports to UKHSA.

What next?

I plan to finish my analysis and write up my thesis. Once this is complete, I plan to build a testing algorithm to check whether my findings work in the real world, before it will officially be rolled out.

Figure A5.2: Project summary sent to Patient and Public Involvement (PPI) session contributors

The session was scheduled for one hour and began with a two-minute introduction from the chair (VS), which set the tone for the meeting. The purpose of this introduction was to explain that everyone would have the opportunity to express their views, and that

disagreement was acceptable provided it was expressed respectfully. Verbal consent was then obtained from all participants to record the session, and it was explained that their contributions would remain confidential. Their views would be summarised in this DPhil, but no personal details would be shared or communicated. Following the introduction, I presented a five-minute overview of the work undertaken in this DPhil using a single PowerPoint slide, which introduced the concepts already set out in Figure A5.2. Finally, the chair then walked through four questions one by one, offering individuals the chance to give their views. The questions were as follows:

1. What do you find interesting or important about this work (or otherwise)?
2. How do you think the research could potentially be developed further?
3. What kinds of ways do you think we can disseminate this information to the wider public?
4. What are your thoughts on the style and phrasing that might be used to communicate this research to the wider public?

The Guidance for Reporting Involvement of Patients and the Public, Short Form (GRIPP2-SF) was used for reporting, providing a standardised approach to documenting PPI when it is integrated into larger projects rather than a large stand-alone PPI project.

A5.4 RESULTS

A5.4.1 Summary of key questions and responses

In the following section I present the key quotes from the contributors and the responses from myself to illustrate the breadth of ideas and views expressed during the PPI workshop. Quotes are taken verbatim from the workshop transcript, with only minor edits (for example removal of duplication) for clarity without changing the meaning. Four cross-cutting themes were also identified including: the scope of the included data, data quality, targeted

dissemination for the public, and use of plain English in lay summaries of the information provided.

Q1: What is interesting or important about this work?

Individuals highlighted several areas of interest and importance. This included location, rurality, vaccination status, prevention, and antibiotic rescue packs. **Quote 1, Theme: Scope of included data:**

Contributor: *"To what extent are you looking at where people live, because clearly there may be some significant impacts in large urban areas versus rural areas,?" (C1)*

Response: *"This is a valuable observation. While geography and rurality are not directly addressed in this DPhil, surveillance reports already incorporate NHS region, and future developments may allow regional severity reporting."*

Quote 2, Theme: Scope of included data:

Contributor: *"Are you looking at prevention, like exercise and diet, so it prevents infections occurring because people are better informed?" (C2)*

Response: *"This is an important reminder of the broader context of ARI surveillance. Prevention factors are outside the scope of this analysis, which focuses on surveillance indicators, but they remain highly relevant for public health strategy."*

Quote 3, Theme: Scope of included data:

Contributor: *"A lot of people now, particularly elderly people and people with chronic conditions, are having RSV vaccinations. To what extent are you intending to follow through on that?" (C1)*

Response: *"This is a valuable suggestion. Vaccination status is not included in this DPhil, but future analyses could explore how peaks in severity vary by vaccination status."*

Quote 4, Theme: Data quality:

Contributor: *"I'm just wondering whether you're monitoring how many of the patients, have got emergency antibiotic kits?" (C3)*

Quote 5, Theme: Data quality:

Contributor: *". . . on my repeat prescription, they actually state it is in conjunction with*

a rescue pack." (C2)

Response: *"This is a very useful point. Rescue packs are a valid concern, as antibiotic prescribing may not always reflect severity when issued as standby treatment."*

Q2: How could the research be developed further?

Contributors suggested comorbidities, lifestyle factors, and improving data quality.

Quote 6, Theme: Scope of included data:

Contributor: *"... will those kind of underlying conditions, like a transplant, be looked at?" (C4)*

Response: *"This is a valuable consideration. Comorbidities such as transplants and kidney disease are not part of this analysis but are important for future research."*

Quote 7, Theme: Scope of included data:

Contributor: *"What about vaping, what impact that has on people's future health?" (C1)*

Response: *"This is an important point. Lifestyle factors like vaping are outside the scope of this work, but could provide valuable context in future surveillance studies."*

Quote 8, Theme: Data quality:

Contributor: *"How consistent is GP data across the country, I can well imagine some GPs are much more consistent than others." (C1)*

Quote 9, Theme: Data quality:

Contributor: *"... hospital told me one thing but it was recorded differently on my record, that's going to impact your information." (C2)*

Response: *"These are very important insights. Data quality and consistency are recognised limitations of CMR-based surveillance and remain a recurring challenge."*

Q3: How should findings be disseminated to the public?

Participants recommended multiple channels for dissemination.

Quote 10, Theme: Targeted dissemination:

Contributor: *"The one place I spend an awful lot of time is at outpatient appointments, they have lots of big displays up about research." (C3)*

Quote 11, Theme: Targeted dissemination:

Contributor: *"Any kind of disturbance in breathing or coughing, they should immediately contact their GP. That kind of measure would help the NHS." (C4)*

Quote 12, Theme: Targeted dissemination:

Contributor: *". . . they all bring different scenarios in, that might be a way of feeding it out to the public." (C2)*

Response: *"These are excellent suggestions. Dissemination via outpatient settings, rehab groups, and proactive education is valuable and should be considered in future work, though it is outside the direct scope of this DPhil."*

Q4: What style and phrasing should be used in communication?

There was strong emphasis on clarity, plain language, and personal relevance.

Quote 13, Theme: Plain English and accessibility:

Contributor: *"Respiratory infection just seems a big mass of ideas, if you can identify and say it includes COVID, bronchitis, then people will understand." (C1)*

Quote 14, Theme: Plain English and accessibility:

Contributor: *". . . be almost more simplistic as well and say chest infection or breathing problems, not everyone even realizes COVID was a respiratory infection." (C3)*

Quote 15, Theme: Plain English and accessibility:

Contributor: *"Sometimes you add too much about the research you've done, Oxford, Cambridge, that then detracts from the actual information." (C2)*

Quote 16, Theme: Plain English and accessibility:

Contributor: *"Particularly, what's the impact on me as a person? If I'm reading it, how is it going to affect me?" (C1)*

Quote 17, Theme: Plain English and accessibility:

Contributor: *". . . it should be really clear and simple for a layman. Even the term*

'renal', I didn't know that means kidney." (C4)

Response: *"These comments are extremely valuable. They reinforce the need for plain English, minimal jargon, and a focus on personal relevance rather than institutional detail."*

A5.5 DISCUSSION

The rich and varied discussion that took place during this PPI workshop identified four key themes that can help shape the future development of ARI severity indicators, including how these are communicated to the public. These themes highlight important additional variables for consideration, challenges relating to data quality, and contributors' ideas for dissemination and the structuring of public summaries.

Scope of included data

Contributors raised important points regarding the use of additional variables in this analysis. These included location, rurality, vaccination status, co-morbidities, and smoking status. Each of these factors can influence an individual's risk of developing severe disease and are therefore highly relevant to this work.

At present, RSC weekly surveillance reports present ARI incidence only by location and age band. If severity indicators are incorporated into these reports in the future, they will also aim to be presented by location and age band. Other variables raised by contributors are not currently included in weekly reporting, but they could play a valuable role in the annual report, which takes a deeper look at surveillance data across the year.

These comments also draw attention to health disparities, as additional variables can provide important insights into how infections are affecting specific populations. For example, are certain ethnic groups disproportionately impacted by ARIs at a given time, and what factors might explain this? Can culturally sensitive measures be introduced to sup-

port communities experiencing rises in ARI severity or incidence? An ideal surveillance system would therefore track severity indicators across a broader range of demographic characteristics, enabling a more equitable and targeted public health response.

The main challenge in presenting weekly indicators across too many variables is that it significantly increases operational complexity. Moreover, appropriate downstream outcomes must be considered, for example, how will additional layers of information translate into improvements in patient care or public health response? Ultimately, ensuring that intelligence is actionable is key. Nonetheless, all of these data are available and could be incorporated in the future if appropriate. Data quality assessments would also be required before operationalising their use.

Data quality

The theme of data quality was raised by three individuals. This issue is not only relevant to the PPI findings, but also emerges throughout this DPhil and is the focus of a dedicated chapter. Contributors highlighted several important aspects of data quality that affect the interpretation and utility of severity indicators.

First, context is lacking in certain situations where it could help interpretation of available data. Two participants discussed the challenge of distinguishing whether antibiotics issued to patients with chronic respiratory disease, such as chronic obstructive pulmonary disease (COPD), were linked to an acute event or simply provided as rescue packs. This reflects the fact that surveillance datasets can lack important clinical detail. In this study, all predictors of severe outcomes were required to be recorded within two weeks of the index ARI case. This approach substantially increases the likelihood that antibiotics were prescribed for a given acute episode, although it cannot eliminate the possibility entirely. A similar issue arises with work absenteeism, where the absence of employment status data introduces bias. Since employment rates vary across demographic groups, for example by ethnicity and deprivation, fit notes may be over-represented in some groups and underestimated in others. This is hard to adjust for.

Second, contributors raised concerns about variation in data recording across the country. Primary care IT systems are geographically distributed, with EMIS dominant in the West of England, London, and much of the South, and The Phoenix Partnership (SystmOne) (TPP) more widely used elsewhere. Factors specific to each system may influence how data are captured, which in turn affects analyses such as those undertaken in this DPhil. As this study included only EMIS practices, further work using combined datasets should be undertaken. Although the RSC holds such linked data, it was not accessible for this project due to governance restrictions.

Additionally, there will be heterogeneity in how data are coded across NHS regions, between practices, and even by individual primary care practitioners. However, the expectation is that the size of CMR datasets will allow meaningful signals to be detected, even if they are weak. Improvements in data quality would not only strengthen these signals but also reduce the risk of bias being introduced into the analysis.

Looking ahead, the public may have a role in supporting improvements in data quality, for example by directly updating demographic details or indicating when they have been hospitalised. Future PPI work could help identify practical and acceptable ways of involving the public in this process, and innovations that enable engagement with health records are likely to be especially valuable.

Accessible summaries and targetted dissemination

For Question 4, I asked contributors to reflect on the summary I had distributed prior to the meeting and to suggest how the style and communication could be improved. In general, feedback focused on the importance of using plain language and ensuring clarity. There was also an emphasis on reducing unnecessary detail so that the key messages stand out. One example highlighted that the term *respiratory infections* is not commonly understood, and that more familiar terms such as *COVID*, *bronchitis*, *chest infection*, or *breathing problem* may be clearer. Contributors also suggested removing references that add little value, such as where the research was conducted, to allow the basic and important

information to stand out.

In addition to feedback on content, contributors also suggested a more targeted approach to dissemination. One individual noted that when attending outpatient clinics for their specific health condition, they often read leaflets and posters in that setting because the information feels directly relevant. In contrast, they rarely engage with materials displayed in general practice, as these are often perceived as too generic. For this work, a targeted dissemination strategy could include feeding information back through our top surveillance practices and publishing tailored content on RSC and practice social media channels. Those who are already most engaged with the surveillance effort are likely to be the most receptive to such information.

Using reflections from this PPI workshop I have refined and modified my lay summary.

A5.6 LAY SUMMARY

What is this research about?

This work aims to help health authorities identify when and where outbreaks of serious lung infections (such as flu, COVID, or pneumonia) are happening across England. This is particularly important during the winter, when many different bugs are circulating, and during pandemics such as COVID.

Why is this important?

By knowing where and when these increases are happening, support can be given to GP practices and hospitals where it is needed. For example, additional staff or hospital beds could be provided, vaccination programmes expanded, or, in extreme situations, public health measures such as lockdowns considered. This will also help people get the treatment they need and can stop the bugs from spreading to other people.

How are we doing this?

We use information written by GPs in patients' medical notes to count how many infections are occurring around the country. We can then look at which infections are serious, and monitor how many serious infections are happening in different parts of the country.

What this means for the public?

By improving the way we track serious lung infections, health services can respond more quickly when problems arise. This means better planning and more support where it is most needed. In the long run, it also helps protect people from the worst effects of winter infections and prepares the country to deal with future pandemics.

A5.7 LIMITATIONS

This PPI workshop was incredibly valuable, though a number of limitations should be considered. Firstly, the group size was small and the time dedicated to the session was limited. This reduced the quantity and variety of feedback that could be obtained. Secondly, the PPI session took place towards the end of my DPhil, which limited the extent to which it could directly influence the overall direction of the work. Nevertheless, the session was instrumental in helping me to develop a lay summary of my research, shaped by the feedback received. In addition, I adapted the presentation of some of the data to include additional variables, such as ethnicity, in response to the points raised. Finally, the discussion surrounding data quality was incredibly valuable as it reinforced the prominence of this theme in my DPhil.

A5.8 FUTURE PPI WORK

Future work should seek to engage a broader and more diverse group of participants. One approach might be to identify individuals who are already more heavily engaged in surveillance through their priary care practices. For example, people who have contributed

virological samples may be especially interested in learning how the data they provide can be used to benefit the public. It would also be valuable to hear public perspectives on how data quality might be improved and the extent to which they feel they should have a role in this process. For instance, should patients be able to directly update their demographic details or indicate when they have been hospitalised or received treatment at another location? Involving patients in efforts to improve data quality could represent a novel and meaningful way of addressing one of the biggest challenges in research that uses CMRs for secondary purposes.

A5.9 SUMMARY

The public are integral to all surveillance practices at the RSC. Without their data, CMR-based surveillance systems could not exist. Engaging with the public helps build and maintain trust in how their highly sensitive health data are used. This trust can be strengthened by disseminating research findings in the most accessible way possible and by clearly communicating the value of the work. In addition, patient feedback and experiences can influence the nature and direction of the research itself. While the scope of this PPI activity was necessarily limited, it nonetheless generated valuable insights that directly informed the development of a lay summary and the presentation of key outputs in this DPhil.