Inequalities in Impact of Respiratory Viruses: Development of Respiratory Virus Phenotypes in EHRs from England Using OpenSAFELY

E Prestige[1], J Quint[2], C Warren-Gash[1], W Hulme[3], EPK Parker[1], E Williamson[1], and RM Eggo[1]

[1] London School of Hygiene & Tropical Medicine, London, UK.
[2] Imperial College London, London, UK.
[3] Bennett Institute for Applied Data Science, University of Oxford, Oxford, UK.

Introduction

The burden of respiratory viruses in England is both large, and unequally distributed. Those from socioeconomically deprived areas and from minority ethnic groups often shoulder this burden.

To explore these disparities computable phenotypes can be designed for use with electronic health records (EHRs). These phenotypes dictate how to identify patients with certain outcomes, using a collection of relevant clinical codes from their records. Phenotypes can be designed to varying degrees of specificity in order to identify patients with higher accuracy or at a larger scale - the specificity required depends on the research question at hand.

In this research we design a series of phenotypes to identify patients with respiratory virus outcomes, with a key focus on **respiratory syncytial virus (RSV), influenza, and COVID-19**. We then apply these phenotypes to our patient populations in order to explore disparities in health outcomes.

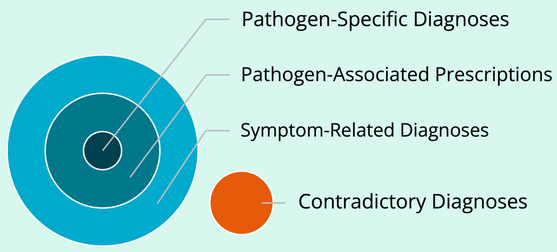
Data Source

With the approval of NHS England, we used pseudonymized **GP data** in OpenSAFELY (1), linked with **Hospital Episode Statistics (HES)** and **ONS mortality data**, to extract respiratory virus outcomes from approximately 45% of the English population. We extract outcomes from patients between **1st September 2016 and 31st August 2024**.

Phenotype Design

We developed a series of computable phenotypes to capture **mild** (primary/emergency care) and **severe** (secondary care) **respiratory outcomes**. For each virus (RSV, influenza, and COVID-19) a maximally specific and maximally sensitive phenotype were designed to capture cases with more accuracy or frequency respectively. We nest the specific phenotype within the sensitive phenotype.

We designed the phenotypes such that specific phenotypes use diagnosis codes which are highly pathogen-specific, and that sensitive phenotypes expand on this by additionally using pathogen-related diagnosis and medication codes. The sensitive phenotypes also contain exclusion criteria which consist of diagnosis codes which indicate other conditions, e.g. rhinovirus.



Validating Phenotypes

When looking to validate the phenotypes, we must find sources of **surveillance data** to compare to. We will also compare the outcomes identified for each patient, dependent on the phenotype used.

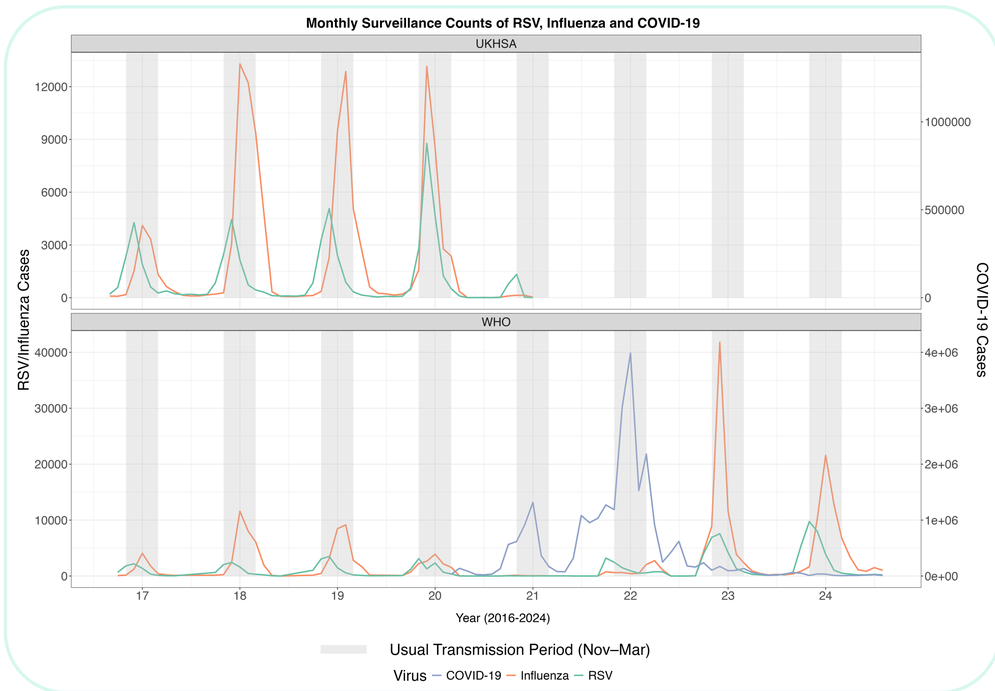
We are interested in the following:

- whether the patient is identified by both specific and sensitive phenotypes
- when a patient is identified by multiple sensitive phenotypes, e.g. RSV and influenza
- the seasonality of outcomes identified with respect to the phenotype used

Validation Sources

To determine the suitability of these phenotypes we will compare these to publicly available surveillance data from the **WHO** (2) and **UKHSA** (3).

However, sources of surveillance data can differ in the cases reported. Additionally, it is often difficult to unpick surveillance data when publicly available - for example which data provided to the WHO is hard to ascertain but it is provided by UKHSA.



Subsequent Analyses

Once phenotypes have been designed and implemented, the events identified can be modelled statistically. We will compare rates per 1000 person-years, adjusting for age group, sex, rurality, and where relevant, vaccination status, in different socioeconomic quintiles and ethnic groups.

Each season (September - September), cohort (infants, children and adolescents, adults, and older adults), and pathogen (RSV, influenza, and COVID-19) will be analysed separately. This will provide an understanding of **whether disparities are consistent across pathogen and age group** and how they have changed over time.

References

(1) Bennett Institute for Applied Data Science. OpenSAFELY Home
(2) World Health Organisation. RespiMart Data
(3) UK Health Security Agency. UKHSA Data

Available At:

`https://www.opensafely.org/.`
`https://www.who.int/tools/RespiMart.`
`https://www.gov.uk/government/publications/.`

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Contact Information:
em.prestige@lshtm.ac.uk

`https://github.com/opensafely/disparities-comparison`

FUNDED BY NIHR | National Institute for Health and Care Research

OpenSAFELY

learn more about me