VERSION HISTORY

V1	04/01/2023	Initial protocol draft
V2	13/02/2023	VW added proposed patient characteristic tables

AUTHORS

Venexia Walker, Rachel Denholm, Ben Goldacre, Angela Wood, Jonathan Sterne

TITLE

Comparison of risk factors for hospitalizations and death from winter infections

LAY SUMMARY

Flu, COVID-19 and other infections are more common in winter. Some risk factors, for example asthma or high blood pressure, may make you more likely to be infected or to become severely ill. When you visit your doctor or go to a hospital, information on your medical conditions is recorded. Approved researchers can access this information. We will use this information to learn more about which factors may make you more likely to have an infection in winter, and to become severely ill because of this infection. We also want to learn about which risk factors are linked with just one infection and which are linked with multiple infections. Knowing this will enable doctors and other health care professionals to take extra care of people at risk. This should mean that they are less likely to need to see a doctor or go to hospital and may reduce pressures on the national health service.

RESEARCH AIMS

The aims of this proposal are two-fold. Firstly, to determine risk factors for hospitalization and death from four winter infections: influenza, respiratory syncytial virus (RSV), pneumonia, and COVID-19. Secondly, to compare risk factors for hospitalization and death across these four winter infections.

DATA SOURCE

This study will be conducted using OpenSAFELY-TPP with the following linked data:

- Secondary Uses Services (SUS)
- Office of National Statistics (ONS)
- Death Registry
- Second Generation Surveillance System (SGSS)
- Index of Multiple Deprivation (IMD)

STUDY DESIGN

We will conduct two cohort studies using data from previous winters. Specifically, we will consider a Winter 2019 cohort that is followed up from 1st December 2019 to 28th February 2020 and a Winter 2021 cohort that is followed up from 1st December 2021 to 28th February 2022 inclusive. We have existing code to develop time-based cohorts (https://github.com/opensafely/post-covid-vaccinated) that can be repurposed to deliver the project within the given timeframe.

STUDY POPULATION

The study population for each cohort will include all individuals in OpenSAFELY-TPP who meet the following criteria:

- Alive on study start date
- Known age between 18 and 110 inclusive on the study start date
- Known sex
- Known deprivation
- Known region
- No record of hospitalization in the 30 days prior to the study start date
- Registered at the same TPP practice from 365 days prior to the study start date to the end of follow-up for that cohort (i.e., 28th February 2020 for the Winter 2019 cohort and 28th February 2022 for the Winter 2021 cohort)

RISK FACTORS

We will consider the following risk factors in this study defined from codes primary care: asthma, other chronic respiratory disease, chronic heart disease, diabetes mellitus, chronic liver disease, chronic neurological diseases, common autoimmune diseases (rheumatoid arthritis, systemic lupus erythematosus or psoriasis), solid organ transplant, asplenia, other immunosuppressive conditions, cancer, evidence of reduced kidney function, and raised blood pressure or a diagnosis of hypertension. These risk factors will be defined as in this repository (DOI: 10.1038/s41586-020-2521-4): https://github.com/opensafely/risk-factors-research/tree/main/codelists.

OUTCOMES

We will consider (1) hospital admission, (2) length of hospital admission (in days), (3) readmission to hospital within 30 days of discharge, and (4) death from four winter infections: influenza, RSV, pneumonia, and COVID-19. COVID-19 will not be studied for the Winter 2019 cohort. Outcomes will be defined using secondary care and mortality data only as testing in primary care differs by winter infection, limiting comparisons between infections. We will use the following definitions:

Outcome	Infection	Variable name
Hospital admission	influenza	out_date_flu_adm
	RSV	out_date_rsv_adm
	pneumonia_strep	out_date_pneustrep_adm
	pneumonia	out_date_pneu_adm
	COVID-19	out_date_covid_adm
Length of hospital admission	influenza	out_num_flu_stay
	RSV	out_num_rsv_stay
	pneumonia_strep	out_num_pneustrep_stay
	pneumonia	out_num_pneu_stay
	COVID-19	out_num_covid_stay

Readmission to hospital within 30 days of discharge	influenza	out_date_flu_readm
	RSV	out_date_rsv_readm
	pneumonia_strep	out_date_pneustrep_readm
	pneumonia	out_date_pneu_readm
	COVID-19	out_date_covid_readm
Death	influenza	out_date_flu_death
	RSV	out_date_rsv_death
	pneumonia_strep	out_date_pneustrep_death
	pneumonia	out_date_pneu_death
	COVID-19	out_date_covid_death

These definitions will be based on the following ICD-10 code lists:

Outcome	**Codelist**
influenza	https://www.opencodelists.org/codelist/bristol/influenza_icd10/71a06879/
RSV	https://www.opencodelists.org/codelist/bristol/rsv_icd10/5890f544/
Pneumonia due to Streptococcus pneumoniae	https://www.opencodelists.org/codelist/bristol/pneumonia_icd10/7453286f
Pneumonia	https://www.opencodelists.org/codelist/opensafely/pneumonia-secondary-care/2020-10-05/
COVID-19	https://codelists.opensafely.org/codelist/opensafely/covid-identification/2020-06-03/

COVARIATES

We will include the following covariates, defined prior to the study start date: age, sex (male and female), deprivation (Index of Multiple Deprivation quintiles), ethnicity (White, Black, South Asian, Mixed, Other, and missing), body mass index (World Health Organization definitions for no evidence of obesity, obese class I, obese class II, and obese class III), and smoking status (current, former and never). We will use the following definitions:

Covariate	**Variable name**	**Definition**
age*	cov_num_age	patients.age_as_of()
sex	cov_bin_male	patients.sex() with 0 indicating female and 1 indicating male
deprivation	cov_cat_deprivation	patients.address_as_of(returning="index_of_multiple_deprivation”)

ethnicity	cov_cat_ethnicity	https://github.com/opensafely/risk-factors-research/tree/main/codelists
body mass index	cov_cat_obese	https://github.com/opensafely/risk-factors-research/tree/main/codelists
smoking status	cov_cat_smoking	https://github.com/opensafely/risk-factors-research/tree/main/codelists

STATISTICAL ANALYSIS

We will fit Cox regression models for each risk factor with a calendar time scale using the cohort start date as the origin. This will ensure that all analyses account for changes with calendar time in rates of the outcome event. All analyses will be stratified by region, and we will consider two forms of adjustment: minimally adjusted, where age and sex are included as covariates in the model, and maximally adjusted, where all covariates listed above are included as covariates in the model. We will also fit an overall model, including all risk factors and covariates. We will repeat our analyses in subgroups defined by age (18-39, 40-5, 60-79, and 80-110), sex (male and female), ethnicity (White, Black, South Asian, Mixed, and Other), and care home residence (true and false) to interrogate whether there are particular risk factors for hospitalization and death from winter infections in these groups. We will report hazard ratios with 95% confidence intervals from all models, subject to disclosure controls.

PLANS FOR ADDRESSING MISSING DATA

Individuals with missing age, sex, or deprivation are excluded from our study population. We will include a missing category for ethnicity. The risk factors are all defined using the presence versus absence of specific codes, so have no identifiable missing values. We will therefore not use multiple imputation for this study.

QUALITY ASSURANCE

We will ensure data quality by applying the following quality assurance rules:

- Remove individuals whose year of birth is after their year of death
- Remove individuals whose year of birth is after date of data extract
- Remove individuals whose date of death is after date of data extract
- Remove men whose records contain pregnancy and/or birth codes
- Remove men whose records contain HRT or COCP medication codes
- Remove women whose records contain prostate cancer codes

PROPOSED MANUSCRIPT TABLES

Table 1: Patient characteristics for each cohort

Characteristic	Category	Number of individuals in Winter 2019 cohort (column %)	Number of individuals in Winter 2021 cohort (column %)
Total			
Age	18–39		
	40–49		
	50–59		
	60–69		
	70–79		
	80+		
Sex	Female		
	Male		
Obesity	Obese class I		
	Obese class II		
	Obese class III		
	Missing		
Smoking	Never		
	Former		
	Current		
	Missing		
Ethnicity	White		
	Mixed		
	South Asian		
	Black		
	Other		
	Missing		
IMD quintile	1 (least deprived)		
	2		
	3		
	4		
	5 (most deprived)		
Care home resident			
Hypertension			
Chronic respiratory disease excluding asthma			
Asthma	With no recent OCS use		
	With recent OCS use		
Chronic heart disease			
Diabetes	Controlled		
	Uncontrolled		
Cancer (non-haematological)	Diagnosed <1 year ago		
	Diagnosed 1–4.9 years ago		
	Diagnosed ≥5 years ago		

Haematological malignancy	Diagnosed <1 year ago		
	Diagnosed 1–4.9 years ago		
	Diagnosed ≥5 years ago		
Reduced kidney function	eGFR 30–60		
	eGFR < 30		
Chronic liver disease			
Stroke or dementia			
Other neurological disease			
Organ transplant			
Asplenia			
Common autoimmune disease (Rheumatoid arthritis, lupus, or psoriasis)			
Other immunosuppressive condition			

Supplementary Table 1a: Patient characteristics, including by infection, for the Winter 2019 cohort

Characteristic	Category	Number of individuals (column %)	Number of individuals with infection (% within stratum)				
			Influenza	RSV	Pneumonia due to Streptococcus pneumoniae	Pneumonia	COVID-19
Total							
Age	18–39						
	40–49						
	50–59						
	60–69						
	70–79						
	80+						
Sex	Female						
	Male						
Obesity	Obese class I						
	Obese class II						
	Obese class III						
	Missing						
Smoking	Never						
	Former						
	Current						
	Missing						
Ethnicity	White						
	Mixed						
	South Asian						
	Black						
	Other						
	Missing						
IMD quintile	1 (least deprived)						
	2						
	3						
	4						
	5 (most deprived)						
Care home resident							
Hypertension							
Chronic respiratory disease excluding asthma							
Asthma	With no recent OCS use						
	With recent OCS use						

Chronic heart disease							
Diabetes	Controlled						
	Uncontrolled						
Cancer (non-haematological)	Diagnosed <1 year ago						
	Diagnosed 1–4.9 years ago						
	Diagnosed ≥5 years ago						
Haematological malignancy	Diagnosed <1 year ago						
	Diagnosed 1–4.9 years ago						
	Diagnosed ≥5 years ago						
Reduced kidney function	eGFR 30–60						
	eGFR < 30						
Chronic liver disease							
Stroke or dementia							
Other neurological disease							
Organ transplant							
Asplenia							
Common autoimmune disease (Rheumatoid arthritis, lupus, or psoriasis)							
Other immunosuppressive condition							

Supplementary Table 1b: Patient characteristics, including by infection, for the Winter 2021 cohort

Characteristic	Category	Number of individuals (column %)	Number of individuals with infection (% within stratum)				
			Influenza	RSV	Pneumonia due to Streptococcus pneumoniae	Pneumonia	COVID-19
Total							
Age	18–39						
	40–49						
	50–59						
	60–69						
	70–79						
	80+						
Sex	Female						
	Male						
Obesity	Obese class I						
	Obese class II						
	Obese class III						
	Missing						
Smoking	Never						
	Former						
	Current						
	Missing						
Ethnicity	White						
	Mixed						
	South Asian						
	Black						
	Other						
	Missing						
IMD quintile	1 (least deprived)						
	2						
	3						
	4						
	5 (most deprived)						
Care home resident							
Hypertension							
Chronic respiratory disease excluding asthma							
Asthma	With no recent OCS use						
	With recent OCS use						

Chronic heart disease							
Diabetes	Controlled						
	Uncontrolled						
Cancer (non-haematological)	Diagnosed <1 year ago						
	Diagnosed 1–4.9 years ago						
	Diagnosed ≥5 years ago						
Haematological malignancy	Diagnosed <1 year ago						
	Diagnosed 1–4.9 years ago						
	Diagnosed ≥5 years ago						
Reduced kidney function	eGFR 30–60						
	eGFR < 30						
Chronic liver disease							
Stroke or dementia							
Other neurological disease							
Organ transplant							
Asplenia							
Common autoimmune disease (Rheumatoid arthritis, lupus, or psoriasis)							
Other immunosuppressive condition							