Statistical analysis plan

Avon Community Acquired Pneumonia Study (Avon CAP):

A Pan-pandemic Acute Lower Respiratory Tract Disease Surveillance Study

Sponsor Project Reference Number	2020 - 2991
ISRCTN registry number	
IRAS ID Number	283899
REC Number	20/EE/0157
Pfizer reference number	WI255886-1

Contents

List of Abbreviations

Abbreviation	Definition
A&E	Accident and emergency
ARI	Acute respiratory illness
BNT162b2	Pfizer/BioNtech COVID-19 vaccine
BRI	Bristol Royal Infirmary
CAP	Community-acquired pneumonia
CHF	Congestive heart failure
COPD	Chronic obstructive pulmonary disease
COVID	Coronavirus disease
CRF	Case report form (eCRF = electronic CRF)
CSA	Clinical study agreement
CT	Computed tomography
CVA	Cerebrovascular accidents
CXR	Chest x-ray
GCP	Good Clinical Practices
ICH	International Council for Harmonisation
ICU	Intensive care unit
IRB/EC	Independent Review Board/ Ethics Committee
LRTD	Lower respiratory tract disease
MRI	Magnetic resonance imaging
NP	Nasopharyngeal
NSTEMI	Non-ST-elevation myocardial infarction
OP	Oropharyngeal
PCV7	7-valent pneumococcal conjugate vaccine
PCV13	13-valent pneumococcal conjugate vaccine
PCV20	20-valent pneumococcal conjugate vaccine
PPV23	23-valent pneumococcal polysaccharide vaccine
RSV	Respiratory syncytial virus
SAP	Statistical analysis plan
STEMI	ST-elevation myocardial infarction
TND	Test Negative Design
UAD	Urinary antigen detection assay
VE	Vaccine effectiveness

CHAPTER 1 – Study introduction and conduct

1. Rationale for study

Accurate incidence rates of acute LRTD and its disease subsets, such as pneumonia and LRTI, remain elusive and the impact of COVID-19 on respiratory disease burden is unclear. Accurate incidence rates of vaccine-preventable infection are required to assess the potential population-level impact of vaccination recommendations.

This population-based multi-hospital, active prospective surveillance is designed to determine population-based incidence rates of hospitalized adults ≥18 of age with community-acquired LRTI (including CAP) in Bristol, England. The involved Bristol hospitals nearly completely capture hospital admissions among residents of a well delineated geographic region allowing for calculation of population-based incidence rates of LRTI. Study data derived from surveillance activities will fully enumerate the number of acute LRTD cases in this region.

LRTD cases will be offered participation in the consented portion of this study involving enhanced testing for pneumococcal, RSV, and SARS-CoV-2 infection. This will allow for more complete characterization the incidence of these infections than standard of care (SOC) testing alone. Additionally, real world vaccine effectiveness (VE) estimates for COVID-19 vaccines are needed to demonstrate their effect in general populations as well as in risk groups outside of the clinical setting. These can be achieved using a test negative design (TND) case control study nested within the ongoing Bristol LRTD surveillance network.

2. Summary

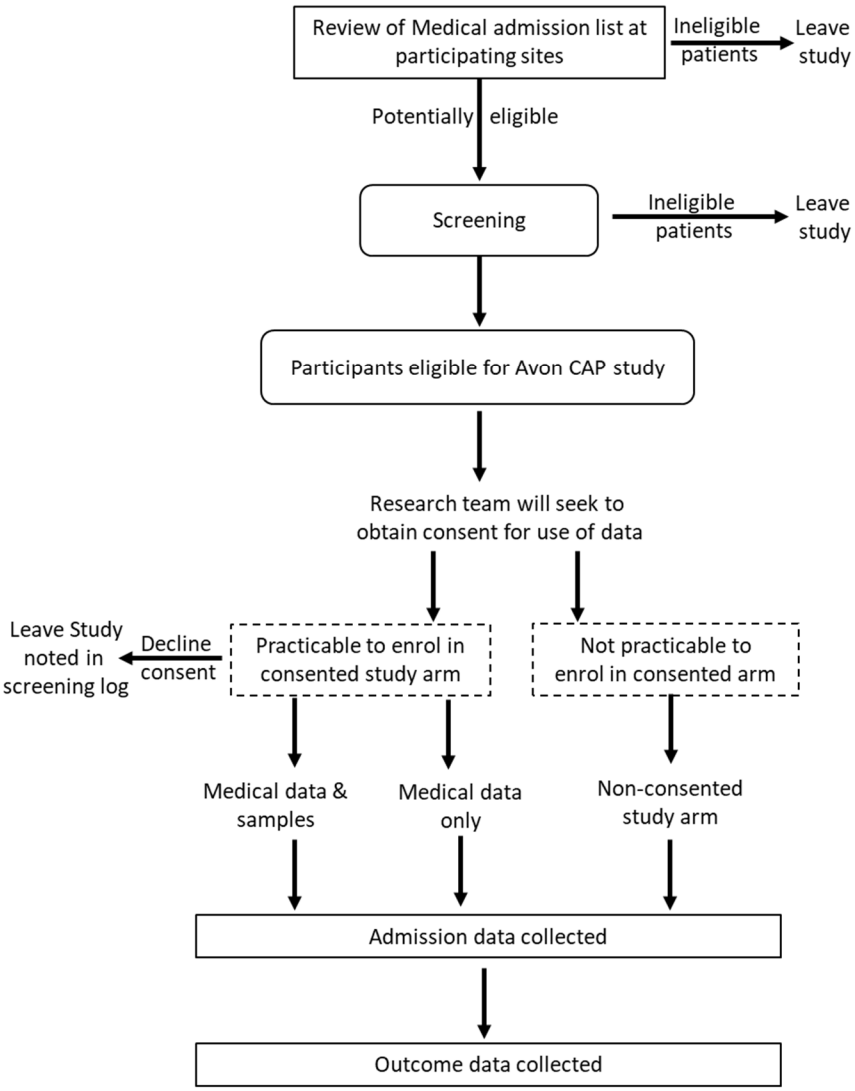
This study will undertake surveillance of lower respiratory tract disease (LRTD) in adults in a defined geographical area and will recruit a subset of the patients identified into the embedded study to provide extra samples and data, as described in the following chapters.

To calculate disease burden accurately and describe acute LRTD and its subsets, this study will record all cases of adults hospitalised with acute LRTD and its subgroups within the defined geographical area who meet study criteria. This will provide data for calculation of disease incidence, burden and outcome analysis in addition to other epidemiological analyses, including identifying risk factors for disease and poor outcome.

Individuals who decline consent for access to medical data will have data fields from medical records where information is recorded as part of SOC collected to determine features of clinical presentation and disease, including co-morbidities, routine healthcare tests and outcomes (e.g. requirement for intensive care and/or organ support, mortality).

The data collected during the non-consented surveillance activity will be combined with the data obtained from the consented arm to provide a comprehensive dataset of patients with acute LRTD. In this way, all patients and disease will be captured and described, allowing for a complete and accurate estimation of disease incidence and burden.

Additionally, VE analyses will be undertaken using data collected in the study. Analyses are described in full in each of the chapters following this introductory chapter.



3. Schedule of activities

3.1 Consented Arm

Sample collection and interview activities shown in bold below will be undertaken for consented participants only.

Table 1. Comprehensive List of Procedures by Visit for Consented-Arm, Including Visit Windows.

Procedure / Assessment [a]	**Screening/ Enrolment Visit [b]**	**Final Assessment/ Vital Status (Data Collection only)**	**Convalescent Visit ("Serology Subset" Only)[e]**
	Visit 1	**Visit 2**	**Visit 3**
	Day 1	**Day 30**	**Day 45**
Visit Window	**Within 48 hours of admission**	**Day 30 to 45**	**Day 22 to 60[c]**
Screening (surveillance)	X		
Clinical symptoms	X		
Medical History[c]	X	X	X
Vaccination History	X		X
CRB-65 and pneumonia severity (PSI) score	X		
Informed Consent	X		
Eligibility confirmation for embedded study	X		
Patient Interview, including approved questionnaires	X		X
Collect urine specimen	X		
Collect upper respiratory tract sample for SARS-CoV-2, RSV and other pathogens[d]	X		X[d]
Collect blood specimen[e]	X		X
Obtain residual (scavenged) standard care specimens from clinical laboratory	X	X	X
Record SOC respiratory specimen testing results	X		
Record SOC blood culture results	X		
Record SOC chest imaging results	X		
Final acute LRTD Illness Diagnosis		X	
Cardiac complications		X	
Vital Status/Mortality		X	
Hospitalization duration, readmission, & level of care		X	

Procedure / Assessment [a]	Screening/ Enrolment Visit [b]	Final Assessment/ Vital Status (Data Collection only)	Convalescent Visit ("Serology Subset" Only)[e]
	Visit 1	Visit 2	Visit 3
	Day 1	Day 30	Day 45
Collect Research-related Injuries (RRIs)	X	X	X

Abbreviations: SOC = standard of care; UAD = urine antigen detection assay; RSV= respiratory syncytial virus; RRI = research-related injury; CRB-65 = Confusion, Respiratory rate, Blood pressure, 65 years of age and older.

a. Sample collection and interview activities shown in bold above will be undertaken for consented participants only.
b. All participants in the consented enhanced diagnostic testing will be have confirmation of their eligibility confirmed and be subsequently enrolled at Visit 1 / Day 1.
c. At Visit 2 and 3, relevant changes to medical history since the last visit will be documented.
d. Respiratory samples will be collected for RSV and other respiratory pathogen testing if such testing has not already been ordered or completed as part of standard-of-care testing. Details on the type and collection process for samples will be included in the laboratory manual. Another respiratory swab will be collected at the convalescent visit, only if the participant experienced a new ARI after hospital discharge.
e. For those that consent to participate in the Serology subset, a blood sample will be collected for serologic testing. A remnant of an appropriate blood sample from standard of care testing can be used if appropriate for this use per laboratory manual specifications. For this subset of patients, an additional convalescent visit will take place approximately 22–60 days after enrolment for collection of a convalescent blood sample. However, effort should be made to schedule this visit as close to day 42 as possible. If subject has another enrolment qualifying acute LRTD event(s) prior to their convalescent visit, they will have acute serology specimen taken at each enrolment and only one convalescent serology visit will be completed 42 days after last acute specimen was taken.

3.2 Non-consented arm

Collection of SoC data will be done for all identified LRTD events in study hospitals and entered into the surveillance databases hosted by the Trusts.

Table 2. Comprehensive List of Information to be collected for non-consented arm.

Procedure / Assessment	First Data Collection	Second Data Collection
	Admission	Outcomes and Results
	Day 1	Day 30
Visit Window	**Within 48 hours of admission**	**Day 30 to 45**
Surveillance	X	
Clinical symptoms	X	
Medical History[a]	X	X
Vaccination History	X	
CRB-65 and pneumonia severity (PSI) score	X	
Record SOC respiratory specimen testing results	X	

Procedure / Assessment	First Data Collection Admission Day 1	Second Data Collection Outcomes and Results Day 30
Record SOC blood culture results	X	
Record SOC chest imaging results	X	
Final acute LRTD Illness Diagnosis		X
Cardiac complications		X
Vital Status/Mortality		X
Hospitalization duration, readmission, & level of care		X

Abbreviations: SOC = standard of care; UAD = urine antigen detection assay; RSV= respiratory syncytial virus; RRI = research-related injury; CRB-65 = Confusion, Respiratory rate, Blood pressure, 65 years of age and older.

a. At Visit 2, relevant changes to medical history since the last visit will be documented.

4. Research methods

4.1 Study design

Adults with LRTD will be screened using population-level surveillance at study hospitals, and collection of SOC data will be performed on all LRTD events, including from SOC laboratory tests. Patients presenting during the recruitment period of the study, with documented or suspected COVID-19 will fulfil study eligibility criteria, and therefore all references to LRTD also encompass documented or suspected COVID-19 cases who may not otherwise qualify as LRTD. LRTD patients will be offered participation in the enhanced diagnostic testing portion of this study with informed consent, which will involve collection of urine, respiratory and, in some cases, blood samples. These samples will be used for study-specific testing and, if necessary, for COVID-19, pneumococcus, and RSV tests if not available from SOC records for any reason. A short patient questionnaire on COVID-related risk behaviours will also be administered. The pneumococcal testing will include serotype to allow estimation of the proportion of the burden that is potentially vaccine preventable – either by the currently available PCV13 or the anticipated PCV20, which is currently in the final phases of clinical development. Information about the additional pneumococcal, SARS-CoV-2 and RSV infection testing will be integrated with the population-level surveillance data to allow for more accurate population-based estimates of vaccine-preventable pneumococcal and COVID-19 and RSV-related LRTD incidence. The epidemiologic data generated from the study will serve as the baseline for future vaccine effectiveness studies – either for current and possibly future SARS-CoV-2 vaccines as well as expected PCV20 and investigational RSV vaccines that are also currently under development at Pfizer.

An additional group of approximately 400 contemporaneous UAD and viral testing control participants without acute LRTD, comparable by age group and season to subjects in the enhanced testing group, will be enrolled. These participants will be identified as adults who are:

1. Patients attending outpatient clinics in participating hospitals in specialties other than respiratory or cardiology OR
2. Patients currently admitted at participating hospitals on non-medical wards with no current acute respiratory or cardiovascular conditions.

Control participants will provide urine for UAD and BinaxNOW testing and an upper respiratory swab sample for RSV/respiratory pathogen testing. The control population is used in every UAD study to determine serotype and population specific cut-points for UAD positivity. For most adult populations these cut-points have remained stable, but for study rigor, Pfizer policy is to obtain controls in each population.

4.2 Patient selection

4.2.1 Screening inclusion criteria

Patients must meet all the following inclusion criteria to be eligible for enrolment:

1. Aged ≥18 years of age
2. Patients with illness with following 2 characteristics:
   a. Acute illness (i.e., present for 28 days or less); **AND**
   b. Evidence of acute LRTD:
      i. Patients with current or suspected COVID-19 or previous proven COVID-19 within last 28 days **OR**
      ii. Clinical or radiologic diagnosis of pneumonia or an acute LRTI **OR**
      iii. New onset or worsening of ≥2 of following 8 LRTD symptoms or clinical findings:
         1. fever (>38.0°C) or hypothermia (<35.5°C) before or within 24 hours of enrolment;
         2. pleuritic chest pain;
         3. cough (including nocturnal only);
         4. sputum production or purulence;
         5. dyspnea (shortness of breath) including orthopnea or on exertion only;
         6. tachypnea (respiratory rate ≥20/min) documented by healthcare professional;

7. abnormal auscultatory findings suggestive of LRTD (e.g., crepitations/rales or evidence of pulmonary consolidation including dullness on percussion, bronchial breath sounds, wheezing, or egophony);
8. radiologic finding that is consistent with LRTD, including pneumonia, and/or acute congestive heart failure (e.g., pleural effusion, increased pulmonary density due to infection, the presence of alveolar infiltrates (multilobar, lobar or segmental) containing air bronchograms, or interstitial oedema).

4.2.2 Screening exclusion criteria

Patients meeting any of the following criteria will not be included in the study:

1. Any patient who develops signs and symptoms of LRTD after being hospitalized for ≥48 hours (either at current hospital, another transferring hospital, or a combination of these), unless admitted with current, previous proven, or suspected COVID-19 infection.
2. Previously enrolled participants readmitted ≤7 days after discharge for their study qualifying admission, unless admitted with current, previous proven, or suspected COVID-19 infection
3. At the time of enrolment, an LRTD-related diagnosis has been excluded or another diagnosis confirmed (for example, patient was found to have fever and tachypnoea due to an intraabdominal process such as cholecystitis)

4.2.3 Study inclusion/exclusion criteria

To participate in the enhanced diagnostic testing, individuals must meet all of the following:

1. Meet all screening inclusion criteria in section 5.2.2
2. Meet none of screening exclusion in section 5.2.3
3. Informed consent document signed and dated by patient, or the requirements for patients unable to provide consent have been fulfilled (as detailed below)

4.2.4 UAD and respiratory pathogen control group

4.2.4.1 Control group inclusion criteria

Individuals must meet all of the following inclusion criteria:

1. Age 18 years and older.
2. Informed consent document signed and dated by patient
3. Individuals who are willing and able to provide urine and respiratory swab

4.2.4.2 Control group exclusion criteria

Individuals presenting with any of the following will not be included in the control group:

1. Individuals who are investigational site staff members or relatives of those site staff member or participants who are Pfizer employees directly involved in the conduct of the trial.
2. Individuals with suspicion of pneumonia or other respiratory infectious diseases or documented, concomitant infectious disease.
3. Individuals residing in any long-term care facilities (for example, nursing homes or respite care facilities).
4. Individuals with known bronchial obstruction or a history of post-obstructive pneumonia. (Chronic obstructive pulmonary disease (COPD) is permissible, provided there has not been an exacerbation within the 3 months prior to enrolment.)
5. Individuals with primary lung cancer or another malignancy metastatic to the lungs.
6. Individuals with fever (measured temperature of ≥38.0° C measured by a healthcare provider).
7. Individuals with significant immunosuppressive disease such as leukaemia.
8. Individuals with either pneumococcal conjugate vaccine (PCV) and/or pneumococcal polysaccharide vaccine (PPV) administration within the past 30 days

4.2.5 COVID-19 VE TND analysis group

For the purposes of the primary COVID VE analysis, the WHO definition of ARI will be used and is defined as an acute respiratory infection that occurred within the last 10 days and required hospitalization with history of fever or measured fever of ≥ 38 C° AND cough. Because many COVID-19 patients do not necessarily have fever (~15%) or cough (~15%) (Garg et al 2020), we will vary this definition in sensitivity analyses.

5. Data management

5.1 Data sources

Data will be recorded on an eCRF for the study, relating to outcomes of:

- Radiology tests
- Microbiology tests
- Clinical parameters regarding hospital admission (including co-morbidities, symptom type and duration, admission observations, etc)
- Vaccination status
- Patient treatment (including requirement for ventilation, renal support, etc)

- Patient outcomes (including hospital length of stay, requirement for intensive care, haemofiltration and complications arising from their LRTD).

Consent will be obtained to record which GP practice the participant is registered with, in order to use this to generate a denominator for incidence calculation as well as for administration of the COVID-risk behaviour questionnaire which the patient will complete with the research team member.

5.2 NHS Databases

A site-specific study database will be built within the NHS IT domain at each participating NHS site, using the REDCap database management software. The REDCap database will be programmed to mark identifiable fields (and therefore restrict their access or download), and to calculate fields such as length of hospital admission, survival (up to 30 days following admission), age at admission. These fields will be exported in the pseudonymised dataset as opposed to the specific date, thereby removing identifiers and aggregating data.

5.3 University of Bristol Data

Data from each participating NHS Trust will be imported and held in two separate databases:

1. A REDCap database within a secure University IT domain- containing only pseudonymised data
2. A password protected database containing identifiable data, with restricted user access on a bespoke server

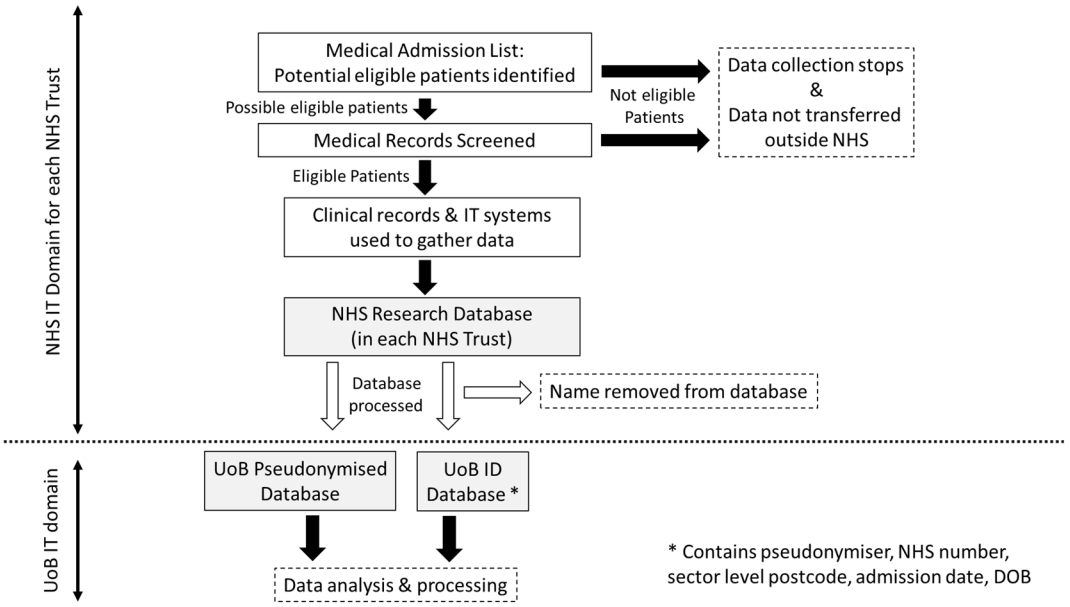
This will create a single unifying research dataset, allowing for analysis of data to meet the study objectives, whilst enabling data security measures to protect data and participants.

5.4 Pseudonymisation

Each patient with an admission that may meet eligibility will be assigned a surveillance number within the database, assigned by the research team. The database will be cleaned using the NHS number to identify individuals who have more than one qualifying admission, both within one NHS Trust and across participating study sites. These individuals will be identified in the pseudonymised database using the surveillance number of their first study eligible admission. The NHS number will subsequently be deleted following processing at the end of the study. This will provide a method of

ensuring that there is an accurate calculation of disease incidence within a defined geographical area with multiple NHS hospitals providing acute care, whilst maintaining a pseudonymised database. A data flow diagram is included below.

5.5 Data Flow Diagram



5.6 Data collection

5.6.1 Consented participants

V1 data collection

Patients meeting screening criteria will have basic demographic and clinical data collected on an e-CRF after consent. This will involve collection of the following data at visit 1:

- Patient details
- Eligibility checks – inclusion/ exclusion criteria components
- Date of hospitalization, length of stay in hospital
- Demographics (age, gender, race/ethnicity), socioeconomic status estimated by postcode, vaccination (influenza/ PCV13/PPV23 and date of last administration), smoking status, alcohol/drug use
- Details of present illness (symptoms, date of onset, vital signs on admission to hospital, used antibiotics in the 14 days prior to admission)

- Standard-of-care test results
    - Routine biochemistry and haematology test results on admission, including C-reactive protein and NT-proBNP
    - Blood cultures
    - Respiratory microbiology testing, including bacterial (e.g., PCR and culture) and viral testing (e.g., COVID19, RSV and influenza)
    - Pneumococcal testing (e.g., BinaxNOW urine test),
    - Antibiotic resistance results for pneumococcal isolates
    - Sputum Gram stain results
    - COVID-19 testing data collection will include tests from up to 14 days prior to hospitalisation based on patient self-report and medical records
- Hospitalization data:
    - Admission hospital
    - Dates and time of admission/discharge
    - ICU stay (yes/no)
    - Number of days in ICU
    - Mechanical ventilation and days of ventilator use
    - Requirement for new/increased haemofiltration
    - Non-invasive ventilation requirement and days of usage
- New York Heart Association (NYHA) Heart Failure Classification, Pneumonia Severity (CRB65 and PSI scores)
- Relevant Medical History and Major Comorbidities
- Rockwood Frailty Score and Charlson Comorbidity Index
- COVID risk behaviours questionnaire including household structure, inclusion in social bubble, mask wearing (see appendix 1)

V2 data collection

- Final standard of care/ clinical diagnosis of acute LRTD illness
- Vital status at Day 30
- Occurrence of cardiovascular events within 30 days of illness onset (pneumonia only):
    - non-ST elevation myocardial infarction (NSTEMI)

- ST-elevation myocardial infarction (STEMI)
- cerebrovascular accident
- new episode of atrial fibrillation or other clinically significant arrhythmia
- deep venous thrombosis or pulmonary embolism
- new or worsening congestive heart failure
- cardiovascular-related death

5.6.2 Patient outcome

The follow-up information collected 30 days after enrolment or when the patient is discharged from hospital (whichever occurs later):

- Record final clinical/standard-of-care diagnosis for qualifying acute LRTD illness:
  - CAP – radiologically or clinically confirmed, acute bronchitis/LRTI, exacerbation of underlying chronic respiratory disease, LRTI not otherwise specified, congestive cardiac failure, empyema/lung abscess, non-infective process, and non-respiratory infection-related diagnosis.
- Vital status at Day 30 after enrolment:
  - deceased, not recovered, recovered, recovered with sequelae, recovery ongoing, unknown
- Cardiac complications through Day 30 after enrolment:
  - ST- and non-ST elevation myocardial infarction, cerebrovascular accident, new episode of atrial fibrillation or other arrythmia, venous thromboembolism, cardiovascular-related death

5.7 Surveillance participants (non-consented)

5.7.1 Admission data collection

Patients meeting screening criteria will have basic demographic and clinical data collected on an e-CRF. This will involve collection of the following data at admission:

- Patient details
- Eligibility checks – inclusion/ exclusion criteria components
- Date of hospitalization

- Demographics (age, gender, race/ethnicity, socioeconomic status estimated by postcode, vaccination (influenza/ PCV13/PPV23/COVID-19 and date of last administration), smoking status, alcohol/drug use)
- Details of present illness (symptoms, date of onset, vital signs on admission to hospital)
    - Used antibiotics in the 14 days prior to admission
- Standard-of-care test results
    - Routine biochemistry and haematology test results on admission, including C-reactive protein and NT-proBNP, and blood group
- Hospitalization data:
    - Admission hospital
    - Dates and time of admission
- New York Heart Association (NYHA) Heart Failure Classification, Pneumonia Severity (CRB65 and PSI scores)
- Relevant Medical History and Major Comorbidities
- Rockwell Frailty Score and Charlson Comorbidity Index for each patient based on history of chronic medical and immunocompromising conditions (Table 2).

5.7.1.1 Vaccination history

Vaccination history, including date(s) of vaccination and product(s) given, will be obtained from medical records collected from relevant healthcare providers (e.g., primary care, public health department), pharmacies, and any local, or national immunization registries for each enrolled patient as follows:

Pneumococcal vaccine (e.g., 23-valent pneumococcal polysaccharide vaccine or 13-valent pneumococcal conjugate vaccine) receipt in the last 5 years. History of vaccination with any other newly-licensed or investigational pneumococcal vaccine (e.g., 20-valent or 15-valent pneumococcal conjugate vaccine which are under development by Pfizer, Inc. and Merck & Co., respectively) will also be collected (if applicable) and categorized as ever received *vs* never received in the last 5 years.

Influenza vaccine receipt in the year prior to enrollment.

Any COVID-19 vaccine receipt at any time.

5.8 Patient Outcome measures

The follow-up information should be collected 1 month after enrolment or when the patient is discharged from hospital (whichever occurs later). At day 30 after admission, the following will be assessed and recorded in the CRF:

- Microbiological investigation results
    - Blood cultures.
    - Respiratory microbiology testing, including bacterial (e.g., PCR and culture) and viral testing (e.g., COVID-19, RSV and influenza),
    - Pneumococcal testing (e.g., BinaxNOW urine test),
    - Antibiotic resistance results for pneumococcal isolates
    - Sputum Gram stain results

- Hospitalization data:
    - Dates of discharge
    - ICU stay (yes/no)
    - Number of days in ICU
    - Mechanical ventilation and days of ventilator use
    - Requirement for new/increased haemofiltration
    - Non-invasive ventilation requirement and days of usage

- Vital status at Day 30 after enrolment:
    - Deceased, not recovered, recovered, recovered with sequelae, recovery ongoing, unknown
    - Date of death if under 30 days from admission

- Record final clinical/standard-of-care diagnosis for qualifying acute LRTD illness:
    - CAP – radiologically or clinically confirmed, acute bronchitis/LRTI, exacerbation of underlying chronic respiratory disease, LRTI not otherwise specified, congestive cardiac failure, empyema/lung abscess, non-infective process, and non-respiratory infection-related diagnosis.

- Cardiac complications through Day 30 after enrolment:

- ST- and non-ST elevation myocardial infarction, cerebrovascular accident, new episode of atrial fibrillation or other arrythmia, venous thromboembolism, cardiovascular-related death

- Hospital related adverse events (e.g., falls in hospital and hospital-acquired infections)

References (chapter 1)

Garg S, Kim L, Whitaker M, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. Mmwr 2020;69:458-64.

Chapter 2 – COVID-19 Test Negative Design Analysis for Pfizer BNT162b2

1. Introduction

The UK began vaccinating its adult population in December 2020, using a Department of Health defined risk-based strategy that prioritized vaccination based on age, comorbidity and key worker status. Initially, only the Pfizer/Biontech mRNA COVID-19 vaccine BNT162b2 was used. BNT162b2 is a nucleoside modified messenger ribonucleic acid vaccine that encodes the full-length, membrane-anchored spike (S) glycoprotein of SARS-CoV-2 with two introduced proline mutations to lock it in the prefusion conformation (Kariko et al. 2008; Pardi et al. 2015, Wrapp et al. 2020). BNT162b2 showed an acceptable safety profile in a Phase 1/2 study (Walsh et al 2020) and, in a Phase 3 trial, was tolerable and demonstrated 95% efficacy against COVID-19 (Polack et al. 2020). The vaccine currently has temporary authorisation for supply under MHRA regulation 174, and data confirming the effectiveness of the vaccine outside of the clinical trial setting are needed. Additionally, real world vaccine effectiveness (VE) estimates for BNT162b2 are needed to demonstrate its effect in general populations as well as in risk groups. This can be achieved using test negative design (TND) case control analysis. For this analysis, cases are individuals hospitalized with acute respiratory tract infection (ARI) and tested positive for SARS-CoV-2 up to 14 days prior to admission or on admission to hospital. Controls are those in whom SARS-CoV-2 was not detected in the same timeframe. Almost all data needed to conduct this analysis are already being collected in this study, including COVID-19 disease and vaccination status from standard of care records alongside other medical history and current illness details. To allow for more complete multivariable adjustment for potential confounding differences between the cases and controls, additional information on COVID-19-related behavioural risk factors will be collected from participants using a standardised questionnaire, such as mask use and inclusion in a social bubble.

Residual nasopharyngeal or other respiratory specimens tested for COVID-19 as part of standard of care will be saved prior to being discarded for destruction. Swabs will be retained frozen until such time as they are selected and sent for variant determination which may include whole genome sequencing (WGS). WGS will occur to identify variants of interest, to identify prevalent circulating strains, and to assess VE against specific variants. WGS will not be used for patient care.

2. Objectives

2.1 Primary

To estimate the effectiveness of 2 doses of BNT162b2 vaccine (i.e. fully vaccinated) against hospitalization for ARI due to SARS-CoV-2 infection.

2.2 Secondary

1. To describe the effectiveness of 1 dose of BNT162b2 vaccine (i.e., partially vaccinated) against hospitalization for ARI due to SARS-CoV-2 infection.
2. To describe the effectiveness of ≥1 dose of BNT162b2 vaccine (i.e., ever vaccinated) against hospitalization for ARI due to SARS-CoV-2 infection.
3. To evaluate the effectiveness of BNT162b2 against ARI hospitalization stratified by prevalent or important viral strains.
4. To evaluate the effectiveness of BNT162b2 against severe hospitalization-related outcomes (e.g., ICU admission, mechanical ventilation, and death).

2.3 Exploratory

1. To describe the effectiveness of BNT162b2 against hospitalization for any ARI stratified by various patient characteristics (e.g., age group, sex, race/ethnicity, chronic medical conditions, history of SARS-CoV-2 infection, long term care facility residence, pregnancy status, receipt of influenza vaccine, time since vaccination (i.e., durability), and time between doses among those who received 2 doses).
2. To calculate the incidence rate reduction of BNT162b2 against ARI hospitalization due to SARS-CoV-2 infection and all ARI hospitalization calculated as VE * background incidence.
3. To describe the proportion of patients with ARI where SARS-CoV-2 was identified.
4. To summarize the proportion and characteristics of hospitalized ARI and SARS-CoV-2 patients who receive 0, 1, or 2 doses of BNT162b2.
5. To summarize the time between administration of the first and second dose of BNT162b2 among patients who received 2 doses.
6. To summarize time since vaccination with BNT162b2 (most-recent dose) from illness onset.
7. To describe demographic, clinical, and laboratory characteristics (i.e., viral strain) and disease severity of any BNT162b2 vaccine failures.
8. To describe COVID-19 disease severity for vaccinated and unvaccinated cases.

3. Selection criteria for inclusion in the analysis dataset

This analysis can fulfil its objectives only if appropriate participants are included. The following criteria will be used to select individuals for analysis dataset:

3.1 Inclusion Criteria

Participants must meet all the following inclusion criteria to be included in the analysis dataset:

1. Age 18 years or older.
2. Admitted to hospital for ARI* at a participating site.
3. Have a result from NP or nasal swab and all relevant data available in the study database.

* For the purposes of the primary COVID VE analysis, the WHO definition of ARI will be used and is defined as an acute respiratory infection that occurred within the last 10 days and required hospitalization with history of fever or measured fever of $\geq 38\ C^{\circ}$ AND cough. Because many COVID-19 patients do not necessarily have fever (~15%) or cough (~15%) (Garg et al 2020), we will vary this definition in sensitivity analyses.

3.2 Exclusion Criteria

Participants will be excluded from the analysis dataset if any of the following criteria apply:

1. Received SARS-CoV-2-directed antiviral treatment within the past 30 days, or COVID-19 monoclonal antibody therapy or COVID-19 convalescent serum therapy within the past 90 days prior to collection of required study-related procedures for the detection of SARS-CoV-2 (i.e., NP or nasal swab).
2. Previous enrolment in this study within the past 30 days. Patients can contribute >1 ARI event to the study if a subsequent ARI event for the same patient occurred >30 days after the previous event.
3. Receipt of any COVID-19 vaccine apart from BNT162b2

4. Test Negative Design

A test negative design (TND) will be used, with patients identified on the basis of a clinical case definition (e.g. acute respiratory infection). Patients are then tested for a vaccine aetiologic agent, and VE is estimated by comparing the odds of vaccination among patients testing positive for vaccine-type aetiology compared to those testing negative. Adjusting can be undertaken for potential confounding factors. TND studies are considered a robust type of observational study for evaluating VE against infectious respiratory diseases (De Serres et al 2013, Jackson et al 2013, Lipsitch et al 2016,

Sullivan et al 2014, Foppa et al 2016, Orenstein et al 2007). The main advantages of the TND are that it helps avoid bias due to (unmeasured) healthcare-seeking behaviour and its ease of access to a series of controls that are representative of the source population. Previous research has shown that this simplicity does not necessarily come at the cost of validity (De Serres et al 2013, Jackson et al 2013, Lipsitch et al 2016, Sullivan et al 2014, Foppa et al 2016, Orenstein et al 2007, Schwartz et al 2017). Specifically, in observational settings, the TND is less susceptible to bias caused by differences in healthcare-seeking behavior among cases and controls (De Serres et al 2013, Jackson et al 2013, Lipsitch et al 2016, Haber et al 2015). This key premise of the TND is based on the idea that individuals with a propensity to seek care when ill may be more likely to receive recommended vaccines and to exhibit other behaviour(s) that reduces the risk of a given vaccine-preventable illness. Such healthcare-seeking behaviour could confound the association between vaccination and the development of infection in traditional cohort or case-control studies. Moreover, healthcare-seeking behavior is difficult, if not impossible, to measure directly. Thus, healthcare-seeking behaviour will, in general, be an unmeasured confounder in traditional cohort or case-control studies that could threaten the validity of VE estimates. By conditioning on patients presenting to healthcare providers with the same clinical syndrome, the TND helps avoid bias due to (unmeasured) healthcare-seeking behaviour.

Further, to prevent selection bias and obtain valid results from a case-control study, controls must be representative of the source population (Miettinen et al 1976, Rothman et al 2017). As such, controls should be individuals who, theoretically, would have been identified as cases had they acquired the outcome, condition, or etiology of interest. If a control had had the disease, *would* they have been likely to be enrolled as a case? (Aschengrau et al 2014). Test-negative controls are derived from patients presenting with similar clinical syndromes as the cases, for example acute respiratory infection, and differ only on the specific pathogen, serotype or strain identified as aetiologic agent. Given appropriate sensitivity and specificity of the test used to distinguish cases from controls, (Lipsitch et al 2016, Orenstein et al 2007) the TND provides some inherent reassurance that controls emerge from the same source population as cases, and that the primary difference between cases and controls is which particular pathogen, serotype, or strain was identified as the causative agent of the disease syndrome.

More than thirty years ago, Broome et al. first applied the TND approach and used patients infected by non-vaccine-type invasive pneumococcal disease to serve as controls in their analysis of pneumococcal vaccine effectiveness (Broome et al 1980). The TND has since been extensively applied to evaluate the effectiveness of influenza vaccine and is an extension of the same logic (De Serres et al 2013, Jackson et al 2013, Lipsitch et al 2016, Sullivan et al 2014, Foppa et al 2016, Orenstein et al 2007, Schwartz et al 2017). Another recent study extended this design to an evaluation of 13-valent pneumococcal conjugate vaccine effectiveness against hospitalized community-acquired pneumonia caused by vaccine serotypes (McLaughlin et al 2018). The current proposed evaluation of BNT162b2 will closely resemble these previous study designs.

Like all observational designs, the TND still requires assessment for bias and confounding that may exist in the absence of randomized participation and blinded follow-up. As such, these studies will collect a data for each participant to assess and control for potential confounding.

4.1 Detection of cases

SARS-CoV-2 will be detected by molecular techniques (i.e. nucleic acid amplification tests, NAAT) from biological specimens collected from nasopharyngeal (NP) or nasal swabs. This will be undertaken via collecting of results of SoC tests or from research specimen (if not undertaken as part of SoC testing).

Cases will be defined as patients who meet selection criteria and:

1. Test positive for SARS-CoV-2 via NAAT performed at hospital admission or study enrollment, OR
2. Tested positive by NAAT from samples collection ≤14 days prior to hospital admission or study enrollment.

Patients who have positive results for SARS-CoV-2 from samples taken ≥72 hours after hospital admission/enrollment will be considered to have nosocomial COVID-19 and will be excluded from the primary analysis. The impact of excluding these patients on VE estimates will be examined in sensitivity analyses. Laboratory confirmation is defined by NAAT, however, results from viral culture, direct or indirect fluorescent antibody staining, or rapid antigen testing will be collected and analyzed in sensitivity analyses.

4.2 Definition of Test-Negative Controls

All other patients who met study inclusion criteria (e.g., at least one NP or nasal swab that was tested for SARS-CoV-2 using NAAT) but for whom SARS-CoV-2 is not identified from NAAT will serve as test-negative controls. This approach mimics the definition of test-negative controls that is commonly used in TND studies of influenza (De Serres et al 2013, Jackson et al 2013, Lipsitch et al 2016, Sullivan et al 2014, Foppa et al 2016, Orenstein et al 2007, Schwartz et al 2017) and pneumococcal vaccines (McLaughlin et al 2018).

4.3 Primary exposure of interest

The primary exposure of interest is history of vaccination with BNT162b2 vaccine. As an observational study design, vaccination with BNT162b2 is not part of the study procedures, rather it would be given as part of the national vaccination campaign. The dates of administration and confirmation that BNT162b2 vaccine was given, which will be recorded in the medical records, will be included in the study CRF data. Patients who received newly licensed or investigational SARS-CoV-2 vaccines other than Pfizer's BNT162b2 will be excluded from all analyses.

BNT162b2 vaccination status will be also captured as part of the patient interview and can be self-reported. Self-reported (only) vaccination status will be analyzed in sensitivity analyses.

Vaccination status will be defined as follows:

1. **Fully vaccinated (primary objective):** defined as 2 doses of BNT162b2 vaccine received with ≥7 days between receipt of the 2nd dose and ARI symptom onset. This group will serve as the 'exposed' group evaluated in the primary objective. Patients who received only 1 dose or 2 doses of BNT162b2 vaccine with <7 days between receipt of the 2nd dose and ARI symptom onset will be excluded from this analysis. In sensitivity analyses, VE will also be calculated for 2 doses of BNT162b2 received with ≥14 days between receipt of the 2nd dose and ARI symptom onset.
2. **Partially vaccinated (secondary objective):** defined as 1 dose (only) of BNT162b2 vaccine received with ≥14 days between receipt of the 1st dose and ARI symptom onset. This group will serve as the 'exposed' group as a secondary endpoint. Patients who received 2 doses or 1 dose of BNT162b2 vaccine with <14 days between ARI symptom onset and receipt of the 1st dose will be excluded from this analysis.

3. **Ever vaccinated (secondary objective):** defined as ≥1 dose of the same BNT162b2 vaccine received with ≥14 days between receipt of the 1st dose and ARI symptom onset. Patients who received 1 dose of BNT162b2 vaccine received with <14 days between receipt of the 1st dose and ARI symptom onset will be excluded from this analysis. This group will serve as the 'exposed' group as a secondary endpoint.
4. **Never vaccinated** defined as never received BNT162b2 vaccine. This group will serve as the reference exposure group (i.e., 'unexposed' group) in all VE analyses.

4.4 Patient Sociodemographic Characteristics, Clinical History, Health Behaviours and Lifestyle, and History of Vaccination

4.4.1 Patient Sociodemographic Characteristics, Clinical History, and Health Behaviours and Lifestyle

Sociodemographic and clinical characteristics for each participant will be collected and described. Importantly, these characteristics will be used to assess and control for potential confounding in adjusted VE models. Factors assessed for confounding will include individual-level variables related to participant sociodemographic characteristics (i.e., time and site of enrollment, age, sex, race, ethnicity, and socioeconomic status using index of multiple deprivation [IMD] which uses postcode of residence to assign a score of 1-10), clinical history and disease severity (i.e., presence of chronic medical or immunocompromising conditions, history of SARS-CoV-2 infection, body mass index (BMI), nursing home residency or other healthcare facility exposure in the past 3 months, and antibiotic use in previous 14 days), health and lifestyle behaviors (e.g., smoking status, weekly exposure to children <5 years of age, current drug abuse, and workplace exposure to COVID-19), and vaccination history (e.g., influenza vaccine in last year and pneumococcal vaccine(s) in last 5 years). These are factors that we have either found to be important covariates in previous work, have been identified in other risk factor literature, or are variables that may be associated with the exposure as well as outcome (i.e. prior positive SARS-CoV-2 PCR test, etc.) (Kumar et al 2020, Petrilli et al 2020, Popkin et al 2020, Singu et al 2020, Tartof et al 2020, Webb et al 2020, Wu et al 2020, Zhou et al 2020, CDC website, Lancet editorial 2020). Section 5.6 of chapter 1 of this statistical analysis plan describes the data collection in detail.

Clinical characteristics describing the presence or absence of underlying immunocompromising and chronic medical conditions will also be collected. History of prior SARS-CoV-2 infection (prior to the current testing period and documented through a positive test in their medical records) will be

assessed. History of immunocompromising conditions will include a history of human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS), leukemia, lymphoma, Hodgkin's Disease, generalized malignancy (excluding skin cancer), diseases requiring treatment with immunosuppressive drugs including long term corticosteroids or radiation therapy, nephrotic syndrome, chronic renal failure (including end-stage renal disease), organ transplantation, multiple myeloma, sickle cell disease, functional or anatomic asplenia, and immune deficiency or other conditions consistent with an immunocompromised state. Other chronic medical conditions and health behaviors to be evaluated will include: hypertension, obesity, chronic heart disease (including heart failure, coronary artery disease, cardiomyopathy, and pulmonary hypertension), diabetes (including type 1, type 2, or gestational), stroke, asthma (moderate or severe persistent), chronic lung disease (such as chronic obstructive pulmonary disease (COPD, including emphysema and chronic bronchitis), idiopathic pulmonary fibrosis, and cystic fibrosis), chronic kidney disease (with dialysis), chronic liver disease (including cirrhosis), alcoholism, or cigarette smoking. A classical and updated Charlson comorbidity index (cCCI (Charlson et al 1987) and uCCI (Quan etc at 2011)) will be calculated

TABLE 1. CATEGORIZATION OF VARIABLES DESCRIBING SOCIODEMOGRAPHIC CHARACTERISTICS, CLINICAL HISTORY, HEALTH BEHAVIORS AND LIFESTYLE, AND HISTORY OF VACCINATION

Variable	d.f.	Categories
Sociodemographic Characteristics		
Time of enrollment	TBD	in 1-week intervals
Recruitment site - Bristol	2	Site 1 – Southmeads Site 2 - BRI
Age group, years at enrollment	1	linear tail-restricted cubic splines
Sex	1	Female Male
Index of multiple deprivation (IMD)	Group level	1-10 (inc)
Race	6	White British White other Mixed Black or African American Asian Other Unknown
Clinical History and Disease Severity		
History of prior SARS-CoV-2 infection	1	Yes
		No
classical Charlson Comorbidity Index (cCCI) (Charlson et al 1987)	5	0 1 2 3 4 5 or higher
updated Charlson Comorbidity Index (uCCI) (Quan et al 2011)	5	0 1 2

Variable	d.f.	Categories
		3
		4
		5 or higher
Body mass index (BMI) [e]	5	Underweight (<18.5)
		Normal or healthy weight (18.5−24.9)
		Overweight (25.0−29.9)
		Obese, class 1 (30.0−34.9)
		Obese, class 2 (35.0−39.9)
		Obese, class 3 (≥40.0)
Immunocompromising condition	1	Yes
		No
Hypertension	1	Yes
		No
Chronic heart disease	1	Yes
		No
Diabetes	1	Yes
		No
Asthma	1	Yes
		No
Chronic lung disease	1	Yes
		No
Chronic kidney disease	1	Yes
		No
Chronic liver disease	1	Yes
		No
Alcoholism	1	Yes
		No
Antibiotic use in 14 days prior to enrollment	1	Yes
		No
Health Behavior and Lifestyle		
Healthcare facility exposure in past 3 months[d]	1	Yes

Variable	d.f.	Categories
		No
Living in long-term care, assisted living, or skilled nursing facility at time of enrollment	1	Yes No
Current drug abuse	1	None Alcohol excess, IVDU, Marijuana, other smoked drugs
Smoking status	3	Current smoker Ex-smoker Non-smoker Unknown
Weekly exposure to children <5 years of age	1	Yes No
School-age children currently living at home with you	1	Yes No
Employed in a healthcare setting	6	Hospital or emergency medical services (EMS) Ambulatory care clinic (e.g., urgent care, outpatient clinic, physician's office) Nursing home or long-term care setting Pharmacy Home health care Other health care setting No, I don't work in the health care setting
Required to leave home to go to work	3	Yes, I cannot work from home Most of the time, but I work from home 1–2 days per week Occasionally, but I work from home 3 or more days per week No, I work from home exclusively
Job resulted in being offered vaccination against COVID-19	1	Yes No
Travelled internationally in the last month	1	Yes No
Able to practice currently recommended social distancing measures (for your area) for combatting COVID-19	4	Always Usually About half the time Seldom

Variable	d.f.	Categories
		Never
Wear a mask or facial covering when you are in public?	4	Always Usually About half the time Seldom Never
Social interaction with other not sheltering (living) with you	4	A great deal A moderate amount Occasionally Seldom Never
Other Vaccination History		
Influenza vaccination within previous year	1	Yes No
PPSV23 vaccination in previous 5 years	1	Received ≥1 dose PPSV23 Never received a dose of PPSV23
PCV13 vaccination in previous 5 years	1	Received ≥1 dose PCV13 Never received a dose of PCV13

d.f.=degrees of freedom; PCV13=13-valent pneumococcal conjugate vaccine; PPSV23=23-valent pneumococcal polysaccharide vaccine; SDI=social deprivation index; TBD=to be determined.

a. *Other includes* American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, or other or unspecified racial categories.
b. BMI is calculated based on weight and height and is reported in kg/m^2.
c. *Includes dialysis centers or* long-term care, skilled-nursing, assisted-living, rehabilitation, or other healthcare facilities.

TABLE 2. CLASSICAL AND UPDATED CHARLSON COMORBIDITY INDEX (CCI) SCORING

Comorbid condition	Charlson Comorbidity Index	
	cCCI	uCCI
History of myocardial infarction	1	0
Congestive heart failure	1	2
Peripheral vascular disease	1	0
Cerebrovascular disease	1	0
Dementia	1	2
Chronic pulmonary disease	1	1
Rheumatic disease	1	1
Peptic ulcer disease	1	0
Mild liver disease	1	2
Diabetes without chronic complication	1	0
Diabetes with chronic complication	2	1
Hemiplegia or paraplegia	2	2
Renal disease	2	1
Any malignancy without metastasis	2	2
Leukemia	2	
Lymphoma	2	
Moderate or severe liver disease	3	4
Metastatic solid tumor	6	6
AIDS (excluding asymptomatic infection)	6	4
Maximum (summary) comorbidity score	33	24

AIDS= acquired immunodeficiency syndrome; cCCI= classical Charlson comorbidity index (Charlson et al 1987); uCCI= updated Charlson comorbidity index (Quan etc at 2011)

5. Statistical Analysis

5.1 Per protocol analysis population

The Per-Protocol Population will include all participants who:

1. Meet all inclusion and exclusion criteria,
2. Have a final diagnosis consistent with ARI*
3. Meet the definition of either case or test-negative control
4. Report SARS-Cov-2 vaccination or have a history available from government-issued COVID-19 vaccination cards, medical and billing records collected from relevant healthcare providers (e.g., primary care, public health department), health-insurance providers, pharmacies, and any local, state, or national immunization registries
5. Did not receive any other newly licensed or investigational SARS-CoV-2 vaccine or COVID-19 prophylactic agent other than Pfizer's BNT162b2.

**Note:* As defined by WHO as an acute respiratory infection that occurred within the last 10 days and required hospitalization with history of fever or measured fever of ≥ 38 C° AND cough.

5.2 Estimated Crude (Unadjusted) VE

Odds of having received BNT162b2 (fully, ever, and partially vaccinated) for cases and test-negative controls will be constructed and compared using ORs and 95% CIs. VE will be calculated as 1−OR multiplied by 100%. Corresponding 95% CIs will be calculated using the Wald method.

5.3 Estimating Adjusted VE

In addition to constructing crude OR and VE estimates, logistic regression modeling to assess BNT162b2 VE after adjustment for potentially confounding factors, including (but not limited to) age, gender, nursing home residence, comorbidities, history of COVID-19 infection, and study week will be performed. These are factors that either been found to be important covariates in previous work, have been identified in other risk factor literature, or are variables that may be associated with the exposure as well as outcome (e.g., prior positive COVID-19 test) (Kumar et al 2020, Petrilli et al 2020, Popkin et al 2020, Singu et al 2020, Tartof et al 2020, Webb et al 2020, Wu et al 2020, Zhou et al 2020, CDC website, Lancet editorial 2020). Covariates will be entered in the logistic regression model in backward stepwise manner. Only variable(s) that change the estimated OR for BNT162b2 by ≥10% (i.e., confounder) (Mickey et al 1989) will remain in the final VE model. Corresponding 95% CIs will be calculated using the Wald method. Mixed-effects models will be used to model group-level socioeconomic variables as a random intercept. In addition to results from the final model, univariate VE results will be presented for each independent variable that is assessed for potential confounding, as the results from a fully adjusted model.

5.4 Handling missing data

Crude estimates of VE will be based on the observed, determinate SARS-CoV-2 test results and BNT162b2 vaccination status. For adjusted VE evaluations, patients with all available covariates will be included in the logistic regression model. If there is a substantial amount of missing data (>10%) for any variables deemed necessary to include in our final analyses, sensitivity analysis will be performed using multiple imputation for missing covariates (under the assumption of missing at random) to understand the impact of excluding patients with missing information in adjusted models.

5.5 Analysis Timings

The COVID-19 pandemic continues to cause substantial morbidity and mortality with new data becoming rapidly available. As such, flexibility and a rapid response to the changing dynamics of disease is needed. We plan to conduct informal interim analyses to inform decision making which may be presented or published. Analyses for the primary, secondary, and exploratory outcomes will be analyzed when available, in the form of ad hoc interim analyses, following LPLV of the study.

5.6 Hypothesis Testing

The overall study type-I error is 5%. For the primary objective, hypothesis testing will be used to assess if 2 doses of BNT162b2 are effective in preventing ARI requiring admission to hospital where SARS-CoV-2 is identified. The primary null hypothesis (H0) vs the alternative hypothesis (H1) is H0: VE≤ 20% vs H1: VE> 20%, where VE is 1–OR multiplied by 100%, and the OR is equal to the odds of being fully vaccinated with BNT162b2 (i.e., 2 doses of BNT162b2 with ≥7 days between receipt of the 2nd dose and ARI symptom onset) for ARI cases requiring hospitalization where SARS-CoV-2 is identified relative to the odds of being fully vaccinated with BNT162b2 for ARI cases requiring hospitalization where SARS-CoV-2 is not identified. For the primary objective, BNT162b2 will be considered effective for preventing ARI requiring hospitalization where SARS-CoV-2 is identified if the lower bound of the 95% confidence interval for the estimated VE is >20%.

5.7 Sample Size Determination

This will be an event-driven study based on the number of cases identified. Study sample size is based on the primary endpoint (BNT162b2 VE against ARI requiring hospitalization where SARS-CoV-2 is identified). The required sample size will depend primarily on i) the proportion of all-cause ARI requiring hospitalization caused by SARS-CoV-2 (which determines the number of cases identified and the ratio of cases to controls in the primary analysis), ii) the average uptake of BNT162b2 in the study population over the duration of the study, and iii) the assumed VE of BNT162b2 against ARI requiring hospitalization where SARS-CoV-2 is identified. Sample size calculations were based on the following fixed assumptions:

- Two-sided, type-I error of 5%
- 90% power
- Log(OR) following approximated normal distribution
- Assumed true BNT162b2 VE varying from 70–90% to prevent ARI requiring hospitalization was modeled
- 5% of all-cause ARI episodes requiring hospitalization will test positive for SARS-CoV-2. A range of 5–30% was also modeled given the attack rate of COVID-19 may vary based on social distancing and shelter-in-place measures, underlying levels of population immunity, and other factors.

Average BNT162b2 vaccine uptake in controls over the study period was allowed to vary in sample size calculations (range: 10–90%) and will depend on potential future vaccination uptake scenarios and timing of the conduct of the study. Final study enrollment size will also depend on the proportion of enrolled patients excluded from the Per Protocol Population because i) vaccination records could

not be obtained, ii) they received a newly-licensed or investigational SARS-CoV-2 vaccine other than BNT162b2 vaccine, or iii) they received BNT162b2 vaccine, but did not receive the full 2-dose schedule.

Appendix 1 presents sample size calculations for various scenarios of BNT162b2 uptake and the proportion of all-cause ARI where SARS-CoV-2 is identified. Depending on the uptake of BNT162b2 and the proportion of ARI hospitalizations where SARS-CoV-2 is identified at the time of the study, approximately 3,000 to 12,000 persons ≥18 years of age will be enrolled. A Statistics Center will monitor BNT162b2 uptake among controls and the proportion of all-cause ARI where SARS-CoV-2 is identified to inform decisions on the sample size required to reach an effectiveness endpoint.

References (chapter 2)

Aschengrau A, Seage GR. Essentials of Epidemiology in Public Health. Third ed. Jones & Bartlett Learning; 2014.

Broome CV, Facklam RR, Fraser DW. Pneumococcal disease after pneumococcal vaccination: an alternative method to estimate the efficacy of pneumococcal vaccine. *N Engl J Med*. Sep 04 1980;303(10):549-52. doi:10.1056/NEJM198009043031003

Butler DC, Petterson S, Phillips RL, Bazemore AW. Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery. Health Serv Res. Apr 2013;48(2 Pt 1):539-59. doi:10.1111/j.1475-6773.2012.01449.x

Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83. doi:10.1016/0021-9681(87)90171-8

De Serres G, Skowronski DM, Wu XW, Ambrose CS. The test-negative design: validity, accuracy and precision of vaccine efficacy estimates compared to the gold standard of randomised placebo-controlled clinical trials. *Euro Surveill*. Sep 12 2013;18(37)

Foppa IM, Ferdinands JM, Chaves SS, et al. The case test-negative design for studies of the effectiveness of influenza vaccine in inpatient settings. *Int J Epidemiol*. Dec 1 2016;45(6):2052-2059. doi:10.1093/ije/dyw022

Garg S, Kim L, Whitaker M, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. Mmwr 2020;69:458-64.

Haber M, An Q, Foppa IM, Shay DK, Ferdinands JM, Orenstein WA. A probability model for evaluating the bias and precision of influenza vaccine effectiveness estimates from case-control studies. *Epidemiol Infect*. May 2015;143(7):1417-26. doi:10.1017/S0950268814002179

Jackson ML, Nelson JC. The test-negative design for estimating influenza vaccine effectiveness. *Vaccine*. Apr 19 2013;31(17):2165-8. doi:10.1016/j.vaccine.2013.02.053

Kumar A, Arora A, Sharma P, Anikhindi SA, Bansal N, Singla V, Khare S, Srivastava A. Is diabetes mellitus associated with mortality and severity of COVID-19? A meta-analysis. Diabetes Metab Syndr. 2020 Jul-Aug;14(4):535-545. doi: 10.1016/j.dsx.2020.04.044. Epub 2020 May 6.

Lipsitch M, Jha A, Simonsen L. Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies. Int J Epidemiol. Dec 01 2016;45(6):2060-2074. doi:10.1093/ije/dyw124

McLaughlin J, Jiang Q, Isturiz R, Sings, H, Swerdlow D, Gessner B, Carrico RM, Peyrani P, Wiemken TL, Mattingly WA, Ramirez JA, Jodar. Effectiveness of 13-Valent Pneumococcal Conjugate Vaccine Against Hospitalization for Community-Acquired Pneumonia in Older US Adults: A Test-Negative Design CID 2018 Oct 30;67(10):1498-1506. doi: 10.1093/cid/ciy312.

Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. Am J Epidemiol. Jan 1989;129(1):125-37.

Miettinen O. Estimability and estimation in case-referent studies. *Am J Epidemiol*. Feb 1976;103(2):226-35.

Orenstein EW, De Serres G, Haber MJ, et al. Methodologic issues regarding the use of three observational study designs to assess influenza vaccine effectiveness. *Int J Epidemiol*. Jun 2007;36(3):623-31. doi:10.1093/ije/dym021

Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ.* 2020;369:m1966.

Popkin BM, Du S, Green WD, Beck MA, Algaith T, Herbst CH, Alsukait RF, Alluhidan M, Alazemi N, Shekar M. Individuals with obesity and COVID-19: A global perspective on the epidemiology and biological relationships. Obes Rev. 2020 Nov;21(11):e13128. doi: 10.1111/obr.13128. Epub 2020 Aug 26.

Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. Mar 15 2011;173(6):676-82. doi:10.1093/aje/kwq433

Rothman KJ. Invited Commentary: When Case-Control Studies Came of Age. Am J Epidemiol. Jun 01 2017;185(11):1012-1014. doi:10.1093/aje/kwx074

Schwartz LM, Halloran ME, Rowhani-Rahbar A, Neuzil KM, Victor JC. Rotavirus vaccine effectiveness in low-income settings: An evaluation of the test-negative design. Vaccine. Jan 3 2017;35(1):184-190. doi:10.1016/j.vaccine.2016.10.077

Singu S, Acharya A, Challagundla K, Byrareddy SN. Impact of Social Determinants of Health on the Emerging COVID-19 Pandemic in the United States. Front Public Health. 2020 Jul 21;8:406. doi: 10.3389/fpubh.2020.00406. eCollection 2020.

Sullivan SG, Feng S, Cowling BJ. Potential of the test-negative design for measuring influenza vaccine effectiveness: a systematic review. Expert Rev Vaccines. Dec 2014;13(12):1571-91. doi:10.1586/14760584.2014.96669

Tartof SY, Qian L, Hong V, et al. Obesity and Mortality Among Patients Diagnosed With COVID-19: Results From an Integrated Health Care Organization. Ann Intern Med. 2020;173(10):773-781.

Webb Hooper M, Napoles AM, Perez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. *JAMA.* 2020;323(24):2466-2467.

Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA.* 2020;323(13):1239-1242.

Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020;395(10229):1054-1062.

The plight of essential workers during the COVID-19 pandemic. The Lancet. 2020 May 23;395(10237):1587. doi: 10.1016/S0140-6736(20)31200-9.

WHO website – ARI definition:
https://www.who.int/influenza/surveillance_monitoring/ili_sari_surveillance_case_definition/en/

Centers for Disease Control and Prevention. COVID-19 and People with Certain Medical Conditions. 2021; https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html. Accessed March 2, 2021.

Appendix 1- Sample size requirements for the final analysis population to detect BNT162b2 VE >20% assuming true VE=70–90% with 90% power and type-I error of 5% (2-sided) under various BNT162b2 uptake scenarios (5 to 30% of ARI hospitalizations due to SARS-CoV-2)

Assume true VE=70%																
	5% of ARI is SARS-CoV-2 positive				*15% of ARI is SARS-CoV-2 positive*				*25% of ARI is SARS-CoV-2 positive*				*30% of ARI is SARS-CoV-2 positive*			
BNT162 Uptake	Controls	Cases	Tot Eval	Tot Enroll*	Controls	Cases	Tot Eval	Tot Enroll*	Controls	Cases	Tot Eval	Tot Enroll*	Controls	Cases	Tot Eval	Tot Enroll*
10	6769	356	7125	11875	2104	371	2475	4125	1171	390	1561	2602	938	402	1340	2233
20	3266	172	3438	5730	1022	180	1202	2003	573	191	764	1273	461	198	659	1098
30	2108	111	2219	3698	665	117	782	1303	377	126	503	838	304	130	434	723
40	1540	81	1621	2702	491	87	578	963	281	94	375	625	229	98	327	545
50	1213	64	1277	2128	392	69	461	768	228	76	304	507	187	80	267	445
60	1015	53	1068	1780	335	59	394	657	199	66	265	442	165	71	236	393
70	909	48	957	1595	308	54	362	603	187	62	249	415	157	67	224	373
80	905	48	953	1588	318	56	374	623	200	67	267	445	171	73	244	407
90	1174	62	1236	2060	435	77	512	853	287	96	383	638	251	108	359	598

Assume true VE=80%																
	5% of ARI is SARS-CoV-2 positive				*15% of ARI is SARS-CoV-2 positive*				*25% of ARI is SARS-CoV-2 positive*				*30% of ARI is SARS-CoV-2 positive*			
BNT162 Uptake	Controls	Cases	Tot Eval	Tot Enroll*	Controls	Cases	Tot Eval	Tot Enroll*	Controls	Cases	Tot Eval	Tot Enroll*	Controls	Cases	Tot Eval	Tot Enroll*
10	4945	260	5205	8675	1518	268	1786	2977	832	277	1109	1848	661	283	944	1573
20	2325	122	2447	4078	717	127	844	1407	396	132	528	880	315	135	450	750
30	1455	77	1532	2553	452	80	532	887	252	84	336	560	201	86	287	478
40	1024	54	1078	1797	321	57	378	630	181	60	241	402	146	63	209	348

50	770	41	811	1352	245	43	288	480	140	47	187	312	114	49	163	272
60	608	32	640	1067	197	35	232	387	115	38	153	255	95	41	136	227
70	505	27	532	887	169	30	199	332	102	34	136	227	85	36	121	202
80	455	24	479	798	160	28	188	313	101	34	135	225	86	37	123	205
90	513	27	540	900	196	35	231	385	132	44	176	293	116	50	166	277
Assume true VE=90%																
	5% of ARI is SARS-CoV-2 positive				*15% of ARI is SARS-CoV-2 positive*				*25% of ARI is SARS-CoV-2 positive*				*30% of ARI is SARS-CoV-2 positive*			
BNT162 Uptake	Controls	Cases	Tot Eval	Tot Enroll*	Controls	Cases	Tot Eval	Tot Enroll*	Controls	Cases	Tot Eval	Tot Enroll*	Controls	Cases	Tot Eval	Tot Enroll*
10	4275	225	4500	7500	1294	228	1522	2537	698	233	931	1552	549	235	784	1307
20	1955	103	2058	3430	594	105	699	1165	322	107	429	715	253	108	361	602
30	1183	62	1245	2075	361	64	425	708	197	66	263	438	155	66	221	368
40	798	42	840	1400	245	43	288	480	135	45	180	300	107	46	153	255
50	568	30	598	997	176	31	207	345	98	33	131	218	78	33	111	185
60	417	22	439	732	132	23	155	258	74	25	99	165	60	26	86	143
70	313	16	329	548	101	18	119	198	59	20	79	132	49	21	70	117
80	241	13	254	423	83	15	98	163	51	17	68	113	43	18	61	102
90	212	11	223	372	82	14	96	160	56	19	75	125	50	21	71	118

Assume true VE=90%																
BNT162 Uptake	Cases	Controls	Tot Eval	Tot Enroll*	Cases	Controls	Tot Eval	Tot Enroll*	Cases	Controls	Tot Eval	Tot Enroll*	Cases	Controls	Tot Eval	Tot Enroll*
10	224	22162	22386	37310	225	4275	4500	7500	227	2039	2266	3777	228	1295	1523	2538
20	102	10125	10227	17045	103	1955	2058	3430	104	934	1038	1730	105	594	699	1165
30	62	6116	6178	10297	62	1183	1245	2075	63	567	630	1050	64	361	425	708
40	42	4116	4158	6930	42	798	840	1400	43	383	426	710	43	245	288	480
50	30	2921	2951	4918	30	568	598	997	30	274	304	507	31	176	207	345
60	22	2131	2153	3588	22	417	439	732	23	203	226	377	23	132	155	258
70	16	1580	1596	2660	16	313	329	548	17	154	171	285	18	101	119	198
80	12	1194	1206	2010	13	241	254	423	14	122	136	227	15	83	98	163
90	10	992	1002	1670	11	212	223	372	13	115	128	213	14	82	96	160

*Assumes 40% of enrolled participants will be unevaluable (i.e., excluded from the Per Protocol Population because i) vaccination records could not be obtained, ii) they received a newly-