BMJ Open

Respiratory syndromic disease study in Shanghai community population

Xiao Yu,[1] Zexuan Wen,[2,3] Wenyong Zhou,[2,3,3] Yaxu Zheng,[1] Jian Chen,[1] Wen-jia Xiao,[4] Sheng Lin,[1] Hongyu Liang,[2] Xiaolei Duan,[2] Weibing Wang ,[2,3] Huanyu Wu,[1] Xin Chen[1]

To cite: Yu X, Wen Z, Zhou W, *et al*. Respiratory syndromic disease study in Shanghai community population. *BMJ Open* 2026;**16**:e103001. doi:10.1136/bmjopen-2025-103001

► Prepublication history for this paper is available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2025-103001).

XY, ZW and WZ contributed equally.

Received 31 March 2025
Accepted 30 January 2026

[1]Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China
[2]Department of Epidemiology, Fudan University Shanghai, Shanghai, China
[3]Fudan University Shanghai, Shanghai, China
[4]Division of Infectious Disease Control and Prevention, Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China

Correspondence to
Professor Weibing Wang; wwb@fudan.edu.cn,
Xin Chen; chenxin_scdc@126.com and
Huanyu Wu; wuhuanyu_scdc@126.com

ABSTRACT

Purpose This prospective community-based cohort study (Acute Respiratory Infection Epidemiological Characteristics Assessment Study (ARI-ECAS)) aims to systematically monitor acute respiratory infection (ARI) incidence, characterise multiple pathogen coinfection patterns and explore microbial landscape dynamics in Shanghai's general population. By integrating syndromic surveillance, molecular diagnostics and metagenomic sequencing, the study seeks to enhance understanding of ARI epidemiology, seasonal variation and host–pathogen interactions to inform predictive modelling and optimise public health interventions in high-density urban environments.

Participants The study enrolled 15 199 permanent residents from all 16 districts of Shanghai, with baseline oropharyngeal swab samples across five representative districts (Xuhui, Jing'an, Jiading, Songjiang and Fengxian). Inclusion criteria required residency ≥6 months and consent for weekly follow-ups. Exclusion criteria addressed mobility limitations (planned relocation >6 months) and recent ARI history. Participants provided demographic, behavioural and clinical data via the Shanghai Health Cloud platform, with baseline and symptomatic-phase biological samples collected for analysis.

Findings to date During the initial 8-month surveillance period (May 2024–January 2025), the ARI-ECAS cohort demonstrated critical insights into the epidemiology of acute respiratory infections in Shanghai's urban communities. Among 15 199 participants, 10.96% reported symptomatic episodes, of whom 21.43% experienced recurrent infections. Pathogen detection using targeted next-generation sequencing (tNGS) identified microbial aetiologies in 53.52% of symptomatic cases, revealing a high prevalence of coinfections: 27.96% involved dual pathogens, while 33.01% showed polymicrobial interactions (≥3 pathogens). Notably, 85.09% of symptomatic episodes were self-managed, underscoring a low healthcare-seeking rate (14.91%) consistent with patterns observed in urban China during postpandemic transitions.

Future plans The current phase of data collection will conclude in June 2025; however, syndromic surveillance and tNGS protocols will be sustained to capture multiyear seasonal transmission patterns. To enhance comparative rigour, future protocols will aim to collect samples from participants during asymptomatic periods in the subsequent year to serve as seasonal baseline controls.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Prospective longitudinal design: first community-based cohort integrating weekly symptom monitoring and multi-omics profiling (targeted next-generation sequencing (tNGS)/metagenomics) for acute respiratory infection (ARI) in a megacity.
⇒ Digital epidemiology platform: real-time data collection via the Shanghai Health Cloud enhances surveillance accuracy and scalability.
⇒ Comprehensive pathogen detection: capability to identify 198 respiratory pathogens and resistance genes through standardised tNGS protocols.
⇒ Under-representation of sampling during asymptomatic periods: lacking ability to monitor longitudinal asymptomatic carriage and seasonal fluctuations of the respiratory microbiome in the absence of symptoms.
⇒ Self-reported symptom recall: potential misclassification of asymptomatic/mild cases and timing bias in symptom onset reporting.
⇒ Observational constraints: limited causal inference for microbial landscape–ARI relationships without experimental validation.

Building on this foundation, the study will integrate contact behaviour and mobility surveys to quantify parameters critical for understanding pathogen transmission dynamics (eg, household contacts and public transportation usage). Furthermore, pathogen detection and metagenomic data will be combined with transcriptomic and metabolomic profiling in selected cases to model multipathogen interaction networks and delineate host immune response pathways, thereby advancing mechanistic insights into polymicrobial cocirculation.

INTRODUCTION

Acute respiratory infections (ARIs) are widely prevalent globally with high disease burden.[1] Since the initial detection of SARS-CoV-2 at the end of 2019, the aetiology and epidemiological complexity of ARI have intensified due to the emergence of new respiratory pathogens and increased circulation of existing pathogens.[2] In China, the prolonged 'zero-COVID' policy temporarily reduced ARI incidence by affecting

population mobility and altering social behaviours.[3 4] However, following adjustments to epidemic prevention policies on 8 December 2022,[5] ARI incidence has fluctuated with the cocirculation of pathogens like SARS-CoV-2, influenza viruses, respiratory syncytial virus (RSV) and mycoplasma, presenting new challenges for prevention and control.[6] Understanding the interactions between these respiratory pathogens and their impact on population health is essential for developing effective epidemic prediction models and preventive strategies in the post-COVID-19 era.[7]

China's current ARI surveillance infrastructure relies heavily on hospital-based data sources, including routine influenza-like illness surveillance and a well-established fever clinic network.[8] While these resources offer substantial data on ARI incidence, they are limited in scope by their dependence on healthcare-seeking behaviours and passive reporting through healthcare providers and diagnostic laboratories. Mild or asymptomatic cases are less likely to visit health facilities, and critical information about the full spectrum of ARI is therefore missing.[8 9] Therefore, community-based, systematic follow-up and syndromic surveillance are necessary to capture a more realistic infection rate of ARI with multiple pathogens and identify risk factors associated with disease severity and spread.[10]

Although previous studies have provided valuable insights into the epidemiology of ARI and respiratory pathogen interactions, these studies are often restricted to short-term, cross-sectional designs that fail to capture seasonal fluctuations and evolving patterns over extended periods.[2 8 11] Seasonal changes in pathogen prevalence, environmental factors and social behaviours contribute to variations in ARI incidence, complicating understanding of disease transmission dynamics and the development of responsive public health interventions.[12] To address this gap, it is essential to establish longitudinal data collection systems that monitor ARI occurrence and pathogen characteristics over time, especially in large, diverse urban populations like Shanghai, where respiratory infections pose unique challenges due to high-density living conditions and significant internal and international migration.[13]

As one of China's largest metropolitan areas, Shanghai is home to nearly 25 million residents, with over 42.1% of the registered population consisting of non-local households and an annual entry–exit population of around 20 million.[14] The city's demographic diversity and high mobility contribute to the elevated risk of respiratory infections, with frequent introduction and reintroduction of various respiratory pathogens.[15] As COVID-19 transitions to an endemic state, it is increasingly important to assess how these pathogens interact, circulate and impact the health of urban populations.[16] Addressing these questions requires a dedicated, systematic surveillance approach that can capture the actual burden of ARI and the contribution of different pathogens in community settings.[17]

In response to this need, we initiated an Acute Respiratory Infection Epidemiological Characteristics Assessment Study (ARI-ECAS) in May 2024. This prospective, community-based cohort study aims to systematically monitor the incidence of ARI, identify cocirculating respiratory pathogens and analyse the dynamics of respiratory infections in the Shanghai population. The ARI-ECAS cohort will facilitate weekly syndromic surveillance across all 16 districts of Shanghai, supported by biological sample collection during infection periods for pathogen detection and metagenomic sequencing. ARI-ECAS will provide crucial data for refining predictive models and enhancing epidemic preparedness and control measures in urban settings by evaluating pathogen prevalence, coinfection patterns and host–pathogen interactions within this diverse community cohort.

This study outlines the design and implementation of ARI-ECAS. This syndromic surveillance cohort will provide comprehensive insights into the epidemiology of ARI and the varied pathogen landscapes in Shanghai. Through this research, we aim to reinforce public health strategies for managing respiratory infections and help develop more effective, evidence-based ARI prevention and control policies in large urban populations. The study seeks to monitor the incidence of ARI, describe its transmission characteristics within the general population of Shanghai, analyse the relationship between different respiratory pathogens and the microbial landscape in individuals with ARI and characterise the coinfection patterns of respiratory pathogens to gain a better understanding of interactions among pathogens and their effects on ARI incidence and severity.

COHORT DESCRIPTION

Study design and setting

The study is designed as a single prospective cohort based on a community syndromic surveillance approach, covering all 16 districts of Shanghai. This cohort tracks ARI incidence through weekly follow-ups and collects biological samples during infection periods for pathogen detection. The samples collected will support in-depth pathogen analysis and microbial landscape assessment in the natural population setting. Data are systematically collected through the 'Shanghai Health Cloud' platform, facilitating efficient tracking and real-time updates. The initial registration documents participants' basic demographic and health-related information, including name, ID number, height, weight, occupation, workplace and address, educational background, medical insurance coverage, chronic diseases, smoking and alcohol consumption habits, other health-related behaviours and details about their household members. Subsequently, vaccination and medical consultation records are retrieved through the 'Shanghai Health Cloud' platform, encompassing (a) vaccination records: vaccine name, manufacturer, vaccination date and dosage details; (b) outpatient information: date of visit, medical institution

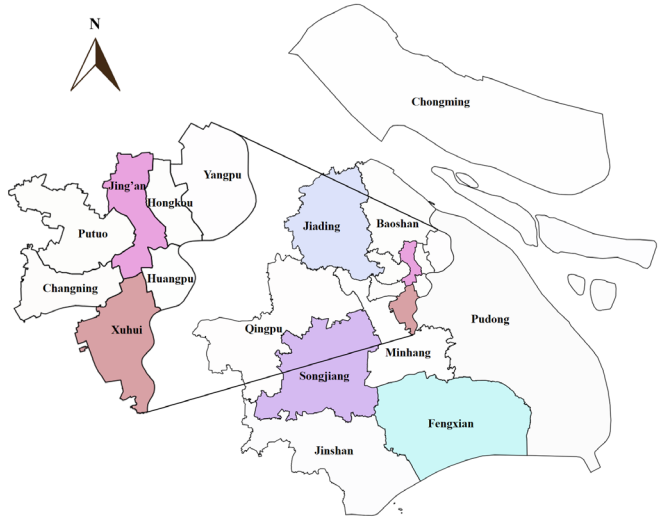

Figure 1 The geographical distribution of the selected districts for syndrome surveillance (all 16 districts) and pathogen surveillance (Xuhui, Jing'an, Jiading, Songjiang and Fengxian districts).

name, department visited and outpatient diagnoses and (c) hospitalisation records: name of the medical institution, admitting department, admission date and discharge summaries.

During the follow-up period, the community-based follow-up team encourages participants to actively report any infectious disease-related symptoms within their household through the Health Cloud app. For participants who do not report symptoms, supplementary weekly surveys are conducted by community staff. The weekly surveys will collect data on the occurrence and treatment of syndromic infectious diseases, results of nucleic acid and antigen testing for SARS-CoV-2 and symptom details, including the earliest onset date and resolution date, if applicable.

Shanghai has 16 districts, with seven central urban districts and nine suburban districts. For regional representation, cohort compliance and research feasibility, five districts are randomly selected to represent the entire city, including two central urban districts and three suburban districts. The selected districts are Xuhui, Jing'an, Jiading, Songjiang and Fengxian (figure 1 and table 1). Considering the syndrome surveillance cohort of the community population in these districts, one individual from each household will be selected to sample baseline oropharyngeal swabs, yielding a projected sample of 1507 participants.

Participants

Inclusion criteria: permanent residents of Shanghai who are currently residing in Shanghai for 6 months or longer and willing to participate in follow-up and sign the informed consent.

Exclusion criteria: (a) individuals with impaired autonomy or cognitive dysfunction; (b) those planning to leave Shanghai for an extended period exceeding 6 months during the study period; (c) participants with confirmed plans to relocate their primary residence during the study period and (d) participants who experienced ARI symptoms within 1 month prior to baseline sampling.

Enrolment

To ensure the cohort is representative of Shanghai's general population, we referenced data from China's Seventh National Population Census, which indicates that one-person, two-person and three-or-more-person households account for 28.4%, 34.5% and 37.1% of households, respectively. Recruitment was targeted to align with these proportions, as reflected in the baseline characteristics (table 3). Community and street management workers mobilised and informed potential participants whose households meet the study's sample size requirements. After mobilising potential participants, community workers checked whether they met the inclusion criteria and enrol the entire households. Eligible potential participants signed a paper-based informed consent form, which was uploaded to the 'Shanghai Health Cloud'. A baseline survey and information entry followed this. Initial oropharyngeal swab samples were obtained from selected districts, preserved at −80°C, and matched with samples collected during the symptomatic period for metagenomic sequencing.

Study procedures and follow-up

Community health service personnel conducted surveys at the time of enrolment. Personal data and health behaviour indicators were recorded using the 'Health Cloud' system. Cross-referencing records from city-wide medical facilities documented information regarding

Table 1 Projected inclusion of study participants in the sampling of baseline oropharyngeal swabs by district

Area type	District	Number of households	Number of streets/towns	Number of participants
Central Urban	Xuhui	241	14	241
	Jing'an	211	14	211
Suburban	Jiading	396	12	396
	Songjiang	412	18	412
	Fengxian	247	12	247
Total	All districts	1507	68	1507

vaccinations, outpatient visits and hospitalisations. Following this, teams of community follow-up staff prompt individuals to report symptoms of infectious diseases affecting their family members through the 'Health Cloud' platform. Community follow-up staff conduct weekly follow-up surveys for individuals who do not report any symptoms, covering the presence of infectious disease symptoms, medical consultations and results of pathogen examinations. In cases of symptom reporting, the onset and duration of symptoms are accurately documented. If symptom reports meet defined criteria, residents can schedule a sampling appointment via the 'Health Cloud' platform. Throughout the symptomatic period, residents may visit the community health service centre for sampling or conduct self-sampling using provided materials after prior notification of standardised sampling protocols. Recommended samples include oropharyngeal swabs and sputum. After sampling, the specimens are temporarily stored at 4°C and collected the same day by sampling collection personnel from the community health service centre or the resident's designated location. The samples are transported to the laboratory under refrigerated conditions. Each patient sample undergoes pathogen testing, while the remaining samples, preserved at −80°C, serve as specimens in subsequent metagenomic sequencing. Results of pathogen detection are shared with participants via the 'Health Cloud' platform within 24 hours. Participant investigation, symptom monitoring and treatment process can be seen in figures 2 and 3.

Outcomes

ARI was defined in alignment with the WHO Regional Office for Europe as the presence of an acute onset of at least one of the following respiratory symptoms: cough, sore throat, runny nose or difficulty breathing.[18] The sampling criteria for study participants required either (1) at least one acute infection symptom (eg, fever, chills, hypothermia or muscle aches) accompanied by at least one respiratory clinical symptom (eg, cough, sputum production, sore throat, dry throat, throat discomfort,

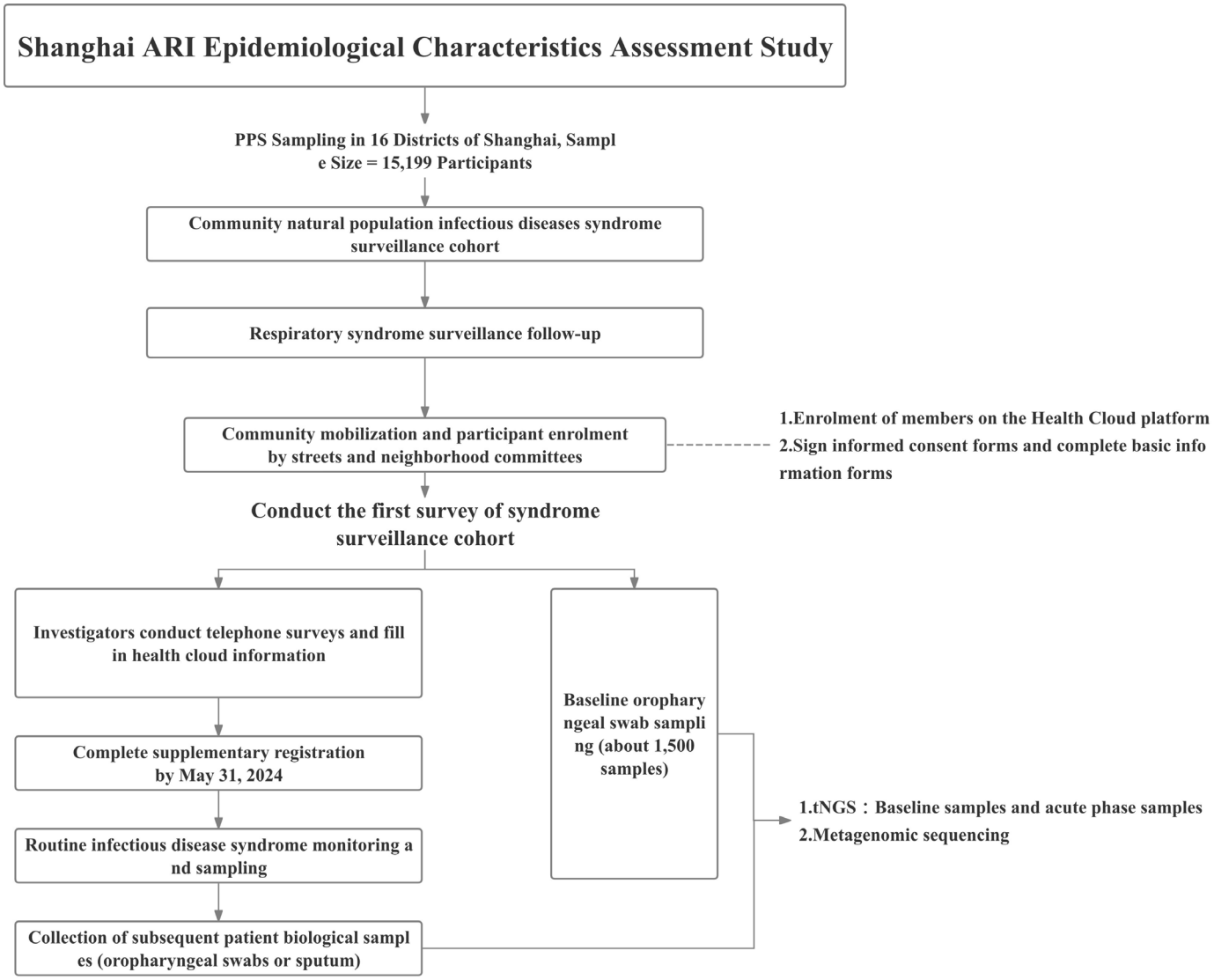

Figure 2 Participant flow through study investigations. ARIs, acute respiratory infections; PPS, probability proportional to size; tNGS, targeted next-generation sequencing.

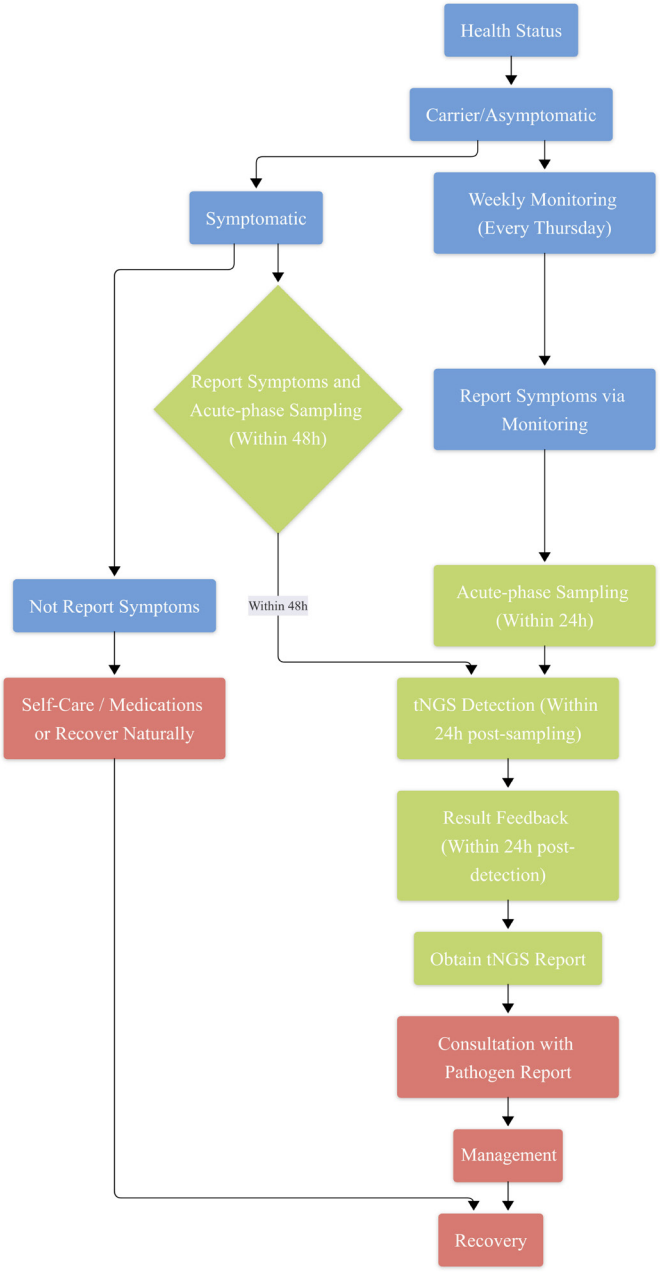

Figure 3 Flow diagram of participants who developed ARI symptoms. ARIs, acute respiratory infections; tNGS, targeted next-generation sequencing.

nasal congestion, runny nose, reduced sense of smell, reduced sense of taste, shortness of breath, chest pain, rapid breathing or difficult breathing) or (2) two or more respiratory symptoms from the list above. The occurrence of the primary outcome triggers the sampling process but does not interrupt the follow-up of the participants. Considering the multiaetiological nature of ARI, we record multiple ARI episodes for the same participant and calculate the incidence rate. It is important to note that the occurrence of related symptoms does not result in the termination of follow-up. Repeated infections with the same or multiple pathogens are systematically recorded and analysed.

Survey, specimen collection and analysis

Data sources

The primary data source includes face-to-face registration (baseline survey) and subsequent follow-up surveys conducted on symptom onset through the 'Health Cloud' platform, with case report forms securely stored on the 'Health Cloud' platform.

Collection, transport and storage of specimens

When participants report symptoms that meet the sampling criteria, either community health service personnel or the participants themselves use sterile long swabs to promptly collect samples from both tonsillar arches, pharynx and tonsil secretions. The swab is placed into a transport tube, securely capped, and this process is repeated four times. Participants who produce more than 2 mL of sputum (not saliva) directly provide sputum samples in a sterile container. Medical staff screen these samples, discard saliva and retain only the qualified sputum specimens. Participants unable to provide adequate sputum undergo induced sputum collection. 10 min before induction, participants inhale salbutamol (a bronchodilator), and forced expiratory volume (FEV1) is measured. Those with postbronchodilation FEV1 below 35% predicted value, unwilling or unable to participate are excluded from the sputum induction procedure. Individuals with postbronchodilation FEV1 >65% predicted value use 5% hypertonic saline nebulisation for 15–20 min, and those with FEV1 ≤65% use 3% hypertonic saline. The procedure stops if either (1) sufficient sputum (>2 mL) is produced or (2) intolerable respiratory discomfort occurs. Sputum samples will be screened to exclude saliva and retain qualified sputum specimens.

Microbiological analyses

Oropharyngeal swabs or sputum samples are analysed using targeted next-generation sequencing (tNGS) technology, adhering to the manufacturer's guidelines provided by KingMEed Diagnosis. This tNGS technology is a rapid and economical approach designed to target 198 respiratory pathogens and resistance genes using specific primers. An ultra-multiplex PCR library construction system enriches target sequences through amplification and is sequenced with second-generation sequencing technology, enabling comprehensive pathogen detection. This methodology encompasses over 95% of common respiratory pathogens with high detection efficiency.[19]

Respiratory microbiome laboratory analyses

Pathogen genomic DNA will be extracted from quality-controlled sputum plugs using the MolYsis Basic5 (Molzym) differential lysis method to eliminate human DNA. Subsequently, the QIAsymphony DSP Virus/Pathogen Kit (Qiagen, 937055) will be employed on the QIAsymphony SP automated nucleic acid extraction instrument to extract total nucleic acids. The construction of the DNA library includes endonuclease treatment for nucleic acid fragmentation, end repair, adapter ligation and purification, PCR amplification

Table 2 Overview of testing projects and required samples

Testing projects	Sample type	Number of samples	Description
Targeted next-generation sequencing	Baseline oropharyngeal swabs	150	Collected from participants in a healthy state for baseline comparison.
	Acute-phase oropharyngeal swabs	150	Collected during symptomatic episodes.
Metagenomic sequencing	Baseline oropharyngeal swabs	150	Corresponding baseline samples for metagenomic comparison.
	Acute-phase oropharyngeal swabs or sputum	150 (total)	Collected during symptomatic episodes, including optional sputum samples.

and purification. Additionally, library concentration and fragment quality control are included. Successfully constructed libraries should exhibit a Qubit concentration of at least 1 ng/μL, with the central peak of library fragments analysed by Agilent 2100 between 280 bp and 400 bp. The libraries will be pooled and mixed, and the pooled libraries will undergo concentration and fragment distribution quality control using Qubit and Qsep. Pooled libraries that meet quality control criteria will be normalised, denatured, diluted and sequenced using the Nextseq 500 (Illumina) high-throughput sequencer with a sequencing strategy of SE75. Postsequencing, low-quality sequences and those shorter than 35 bp will be discarded. High-quality sequences aligned to the human reference genome will be filtered out using the JinHong system. The remaining data, after the removal of low-complexity sequences, will be aligned to the specialised PDseq_db-V2.0 microbial database. The aligned data will then be classified and categorised into viruses, bacteria, fungi and parasites.

Respiratory microbiota bioinformatics

Throughout the study, we will calculate pathogen diversity measures, relative abundance and estimated concentration for each respiratory biological sample. Hierarchical clustering will delineate respiratory microbiota types using the entire sequencing dataset, assigning each sample to a specific type. The Shannon and Bray-Curtis dissimilarity indexes will be used to quantify microbiota components' α and β diversity, respectively. Concurrently, we will use a metagenomics approach to carry out gene function annotation for the collected samples, encompassing essential functional genes like drug resistance genes and virulence genes and analyse the changes in the respiratory microecological community before and after pathogen infection. Table 2 *summarises all testing projects and the respective samples required for analysis.*

Study timeline

Cohort 1 completed participant enrolment in May 2024, with baseline surveys and initial infectious disease syndrome follow-up assessments conducted on May 16. Subsequently, routine follow-ups were conducted weekly until a severe loss of follow-up forced the replacement of follow-up participants or a sudden public health event leading to this cohort being suspended or changed (figure 4).

Statistical analysis

We will present the number of participants registered and available at each follow-up in a flowchart. Descriptive tables detailing sociodemographic, behavioural

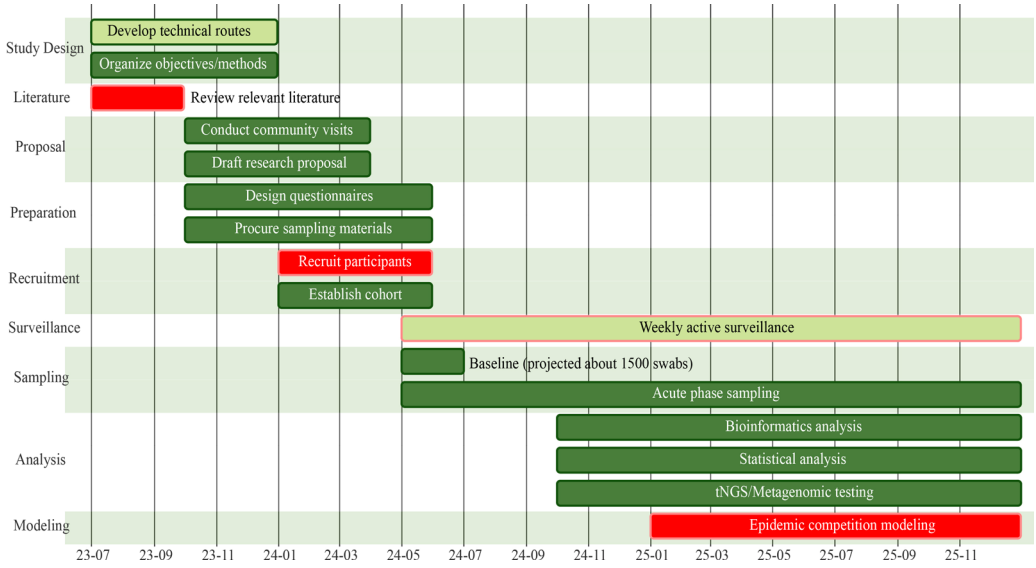

Figure 4 Timeline of Acute Respiratory Infection Epidemiological Characteristics Assessment Study (ARI-ECAS) activities in Shanghai, China, between June 2023 and December 2025. tNGS, targeted next-generation sequencing.

and clinical characteristics will be included. Additionally, comparisons will be made between participants who completed the follow-ups and those who were lost to follow-up.

Descriptive statistics

We will summarise the baseline characteristics of the study population by employing means and SDs for continuous variables and frequencies and percentages for categorical variables. Additionally, we will present follow-up data regarding person-weeks of observation and the incidence of ARI events.

Incidence of ARI and description of its transmission characteristics among the natural population in Shanghai

We will collect the characteristics of ARI disease syndromes among participants from 16 districts through the 'Health Cloud' platform. Utilising the 198-pathogen multiplex PCR of KingMed Diagnostics, we will detect pathogens in oropharyngeal swabs and provide feedback to participants within 24 hours. We will use these data to calculate the cumulative incidence of ARI with 95% CIs throughout the study period and to describe its transmission characteristics.

Association of different pathogens' infections with the respiratory microbial landscapes

Linear mixed-effects models will be used to analyse the association between the microbial landscape (tNGS and metagenomics results: α and β diversity, respiratory microbiota composition, microbial absolute abundance and functional genes) and ARI incidence. Adjustments will be made for age, sex, clinical symptoms and comorbidities. The association will be expressed as regression coefficients with 95% CIs.

Software

All analyses will be conducted using R (V.4.2.3; R Core Team). Significance will be set at a two-tailed p value <0.05.

Data management and confidentiality

Data management

Each participant is assigned a unique identifier on enrolment. The naming convention adheres to the format 'Sample Unit Number+Household Number+Household Member Number'. The Survey Point Number is encoded as a '9-digit Street (Town) National Standard Code - Sample Unit Serial Number', starting with 01. For example, the first survey point in Nanjing East Road Street, Huangpu District, would be designated '310101002–01'. The Household Number is a sequential number within the survey point, such as J01, J02. Household members are numbered sequentially, with the head of the household assigned 01, followed by 02, 03 and so on. An example of a complete participant identifier is '310110012–01J0502'.

If a household remains uncontactable for three consecutive follow-ups, the participants will be considered lost to follow-up. Participants who die or voluntarily withdraw will also be classified as lost to follow-up. Every January and July, community health service centres will compile a list of lost participants and upload the basic registration forms of newly added replacement households to the 'Health Cloud' platform.

Confidentiality

The research team, trained in protecting human research participants and adhering to good clinical practice guidelines, will ensure the confidentiality of all participants. Interviews will be conducted in private settings, and data will not be presented in a manner that could identify individual participants in reports and publications.

Quality control

In the nearly one and a half years of research in the early stage,[20] we have greatly trained our entire research team, making the information collection and sampling processes very reasonable and efficient. Moreover, we have replaced some research subjects with poor compliance and those who were lost to follow-up. These are the essential guarantees to carry out the subsequent research as planned.

The family doctors responsible for information collection using the 'Health Cloud' platform have received professional training. The standard sampling processes of oropharyngeal swabs and sputum have also been ensured to be notified to the doctors or individual participants.

Patient and public involvement

Patients or the public were not involved in developing the research questions or study methods. The research findings will be shared through open-access publications and dissemination meetings with local stakeholders, healthcare providers and communities.

ETHICS AND DISSEMINATION

Ethical considerations (including informed consent)

This study was approved by the Medical Research Ethics Committee of the Shanghai Center for Disease Control and Prevention (KY-2024–19). All patients were asked to provide their written informed consent to participate. All procedures of this study are performed following the principles laid down in the Declaration of Helsinki.

Dissemination and data availability

The results will be disseminated in peer-reviewed journals. Authorship will follow the International Committee of Medical Journal Editors' recommendations, and professional writers will not be used. On completion of the study, the anonymised dataset will be available to

Table 3 Baseline characteristics of responders compared with Shanghai's registered permanent population

Characteristic	ARI-ECAS (n=15 199)	Shanghai's registered permanent population (n=14 390 900)
Sex, n (%)		
Male	7388 (48.61)	7 423 900 (49.37)
Female	7811 (51.39)	7 614 300 (50.63)
Age group, n (%)		
0–4 years	240 (1.58)	521 386 (3.62)
5–17 years	1525 (10.03)	1 068 980 (7.43)
18–29 years	1266 (8.33)	1 429 450 (9.93)
30–39 years	2762 (18.17)	2 239 138 (15.56)
40–49 years	2961 (19.48)	1 870 848 (13.00)
50–59 years	2282 (15.01)	2 073 942 (14.41)
60–69 years	2305 (15.17)	2 936 918 (20.41)
70–79 years	1500 (9.87)	1 445 660 (10.04)
80+ years	358 (2.36)	804 578 (5.59)
Level of education, n (%)		
High school and below	7774 (51.15)	9 191 781 (63.87)
Associate and Bachelor's degrees	6602 (43.44)	5 199 119 (36.13)
Master's degree and above	309 (2.03)	
Unknown	514 (3.38)	–
Occupation, n (%)		
Retirees	4191 (27.57)	–
Civil servants and administrative staff	4162 (27.38)	–
Students	1719 (11.31)	–
Workers and artisans	1426 (9.38)	–
Medical staff	984 (6.47)	–
Preschool children	261 (1.72)	–
Non-institutionalised children	106 (0.7)	–
Educators	89 (0.59)	–
Service industry	41 (0.27)	–
Freelancers and self-employed	25 (0.16)	–
Technical professionals	23 (0.15)	–
Others	2172 (14.29)	–
Smoking status, n (%)		
Never smoked	11 684 (76.87)	–
Used to smoke, but not now	938 (6.17)	–
Smoke now, but not daily	698 (4.59)	–
Smoke daily now	1365 (8.98)	–
Unknown	514 (3.38)	–
Alcohol consumption history, n (%)		
Never	11 265 (74.12)	–
Occasionally (1–3 days/month)	2782 (18.3)	–
Frequently (1–6 days/week)	402 (2.64)	–
Daily	236 (1.55)	–
Unknown	514 (3.38)	–
Underlying diseases, n (%)		

Continued

Table 3 Continued

Characteristic	ARI-ECAS (n=15 199)	Shanghai's registered permanent population (n=14 390 900)
No	11 933 (78.51)	–
Yes	3266 (21.49)	–
Household size, n (%)		
One person	1812 (27.73)	1 627 462 (28.4)
Two persons	2127 (32.55)	1 977 022 (34.5)
Three or more persons	2595 (39.72)	2 126 015 (37.1)

ARI-ECAS, Acute Respiratory Infection Epidemiological Characteristics Assessment Study.

both members of the ARI-ECAS collaboration and other external researchers. They will be available from the chief investigator on reasonable request.

FINDINGS TO DATE

Table 3 reveals moderate variations in demographic characteristics between the study cohort (n=15 199) and Shanghai's permanent residents. The cohort demonstrated lower proportions of paediatric populations (0–4 years: 1.58% vs 3.62%) and octogenarians (2.36% vs 5.59%), with moderate over-representation in middle-aged subgroups (30–49 years: 37.65% vs 28.56%). Educational attainment showed a marginally higher prevalence of tertiary education (associate/bachelor's: 43.44% vs 36.13%) than municipal averages. Household composition patterns remained generally aligned with city-wide statistics (single-person households: 27.73% vs 28.4%; households with three or more persons: 39.72% vs 37.1%). While occupational distribution reflected expected workforce patterns, health behaviour metrics (76.87% never-smokers, 74.12% non-drinkers) and comorbidity rates (21.49%) provided cohort-specific insights without established municipal comparators. These findings suggest the cohort maintains reasonable demographic congruence with Shanghai's population in core parameters while exhibiting expected minor deviations in age stratification and educational profile.

Figure 5 illustrates the 8-month surveillance outcomes (9 May 2024–31 January 2025) in the ARI Epidemiological Characteristics Assessment Cohort. The cohort demonstrated high surveillance compliance, with 85.47% (12 991/15 199) remaining asymptomatic throughout follow-up. Among symptomatic participants (10.96%, 1666/15 199), 78.57% experienced single ARI episodes, while 21.43% exhibited recurrent symptomatology (2–5+ episodes). Attrition analysis revealed 3.57% loss to follow-up (542/15 199), with mortality accounting for 2.39% (13/544) of attrition cases.

Healthcare-seeking behaviour analysis of 2132 symptomatic episodes showed predominant self-management (85.09%, 1814 episodes) versus formal medical consultation (14.91%, 318 episodes). Pathogen detection via tNGS testing (53.52% of symptomatic episodes, 1141/2132) identified complex microbial profiles: 31.72% single-pathogen detection (362/1141), 27.96% dual-pathogen codetection (319/1141) and 33.01% polymicrobial infections (≥3 pathogens). Notably, 8.06% (92/1141) of tested episodes showed no detectable pathogens. Regarding sample types, while oropharyngeal swabs were the primary specimen, sputum samples were collected where possible (15/1141). These serve as supplementary specimens for multi-omics analysis when swab samples are insufficient. These findings characterise the cohort's infection dynamics and healthcare utilisation patterns during longitudinal surveillance.

DISCUSSION

This study aims to gain a comprehensive understanding of the epidemiology and mechanisms of ARI in Shanghai, focusing on seasonal variations and their impact on the spread of ARI. Integrating data from questionnaire surveys, pathogen detection and metagenomic sequencing within a cohort study of the natural population can explore the mechanisms of ARI transmission among the population and the interactions between multiple pathogens and the host.

At present, there have been a considerable number of cohort studies on chronic diseases worldwide, including well-known large-scale chronic disease databases, such as the UK Biobank (UKB)[21] and the China Kadoorie Biobank.[22] Specifically, the UKB database also integrates the research objects with electronic hospitalisation records,[23] allowing us to conduct cohort studies on ARIs.[24] However, data on infectious diseases based on hospitalisation records reflect a significant admission rate bias. Many patients with milder symptoms do not require hospitalisation, leading to incomplete tasks such as sample collection and contact investigation, hindering continuous in-depth research. Given the nature of infectious diseases, particularly ARIs, characterised by rapid onset, a short symptom window and frequent recurrence, long-term and frequent short-interval follow-up efforts (at least one follow-up per week) are essential. Establishing and maintaining the cohort under these circumstances presents substantial challenges, necessitating a

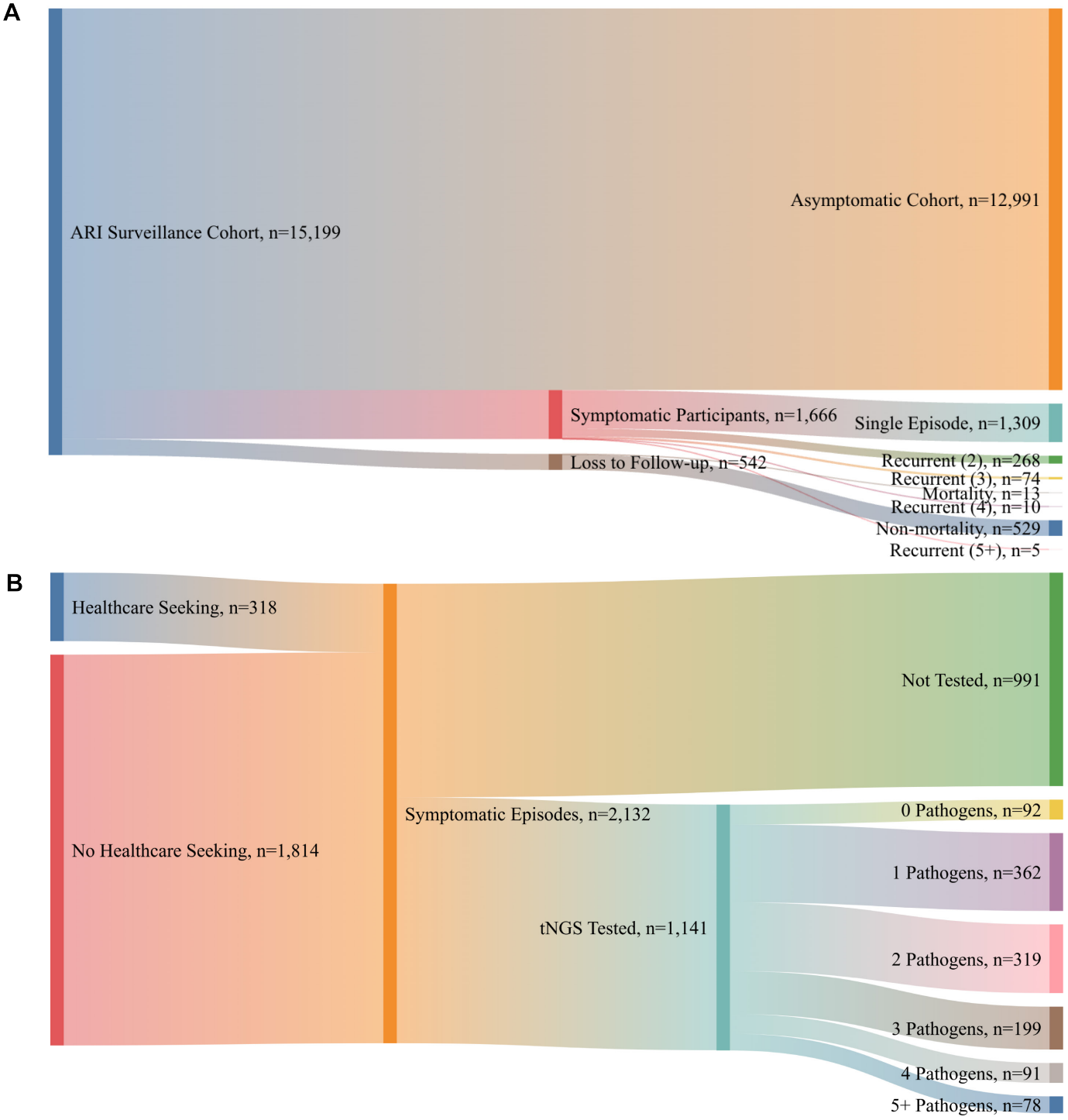

Figure 5 ARI-ECAS cohort follow-up and tNGS results from 9 May 2024 to 31 January 2025. ARI-ECAS, Acute Respiratory Infection Epidemiological Characteristics Assessment Study; tNGS, targeted next-generation sequencing.

considerable allocation of manpower, material resources and financial support.

Although there are currently some studies that have established cohorts for subsequent prognostic follow-up studies based on patients with respiratory infections[25] and have also established cohort studies on respiratory pathogen infections based on newborn infants,[26] there are currently no regular ARI cohort studies based on the natural population conducted worldwide. A previous study[27] evaluated the variations in the respiratory microbial landscape between ARI patients and healthy controls, but no research has conducted self-control comparisons before and after the onset in the same participants, which is crucial for establishing causal relationships. Some studies have utilised multi-omics techniques to target specific pathogens, such as SARS-CoV-2[27 28] and RSV,[29] to study the interaction between the pathogen and the host. However, there is still no study to explore the interaction

mechanism among multiple pathogens and between multiple pathogens and the host.

Strengths and limitations

Thus, our study results will provide valuable evidence of the epidemiological characteristics and mechanisms of multiple pathogen respiratory outbreaks in Shanghai and offer parameters for subsequent epidemic predictions and effectiveness evaluations of interventions and positively impacting prevention efforts in regions with similar demographic characteristics. Using tNGS technology allows for rapidly and precisely identifying a wide range of respiratory pathogens. This provides timely feedback to the residents and can be an authoritative test report for their subsequent medical actions. We also recognise that providing rapid pathogen detection results to participants may influence their healthcare-seeking behaviours, potentially reducing unnecessary hospital visits or facilitating targeted treatment. While this 'intervention' effect is difficult to quantify in the current phase, we plan to investigate the specific impact of diagnostic feedback on medical behaviour in our subsequent 'Respiratory Map'series of studies.

Despite the strengths of our study, several limitations should be acknowledged. First, regarding the sampling strategy, our current design involves a single baseline sample followed by sampling only during symptomatic episodes. We acknowledge that this limits our ability to monitor longitudinal asymptomatic carriage and seasonal fluctuations of the respiratory microbiome in the absence of symptoms. Second, the reliance on self-reported data for timing of onset and specific symptoms may introduce recall bias and affect the accuracy of our findings. Third, due to the long entire research cycle, the short follow-up interval and the complex investigation and sampling procedure, participants included in the study must have relatively high compliance. This will inevitably lead to volunteer bias and reduce the representativeness of the sample population, even though we try to select the study population randomly from the whole city as much as possible. Finally, the study's observational nature means that causal relationships between microbial landscape and ARI incidence cannot be definitively established. Future studies should consider these limitations and explore ways to address them, such as expanding the study to diverse settings and improving standardisation in sample handling procedures.

To mitigate the under-representation of sampling during asymptomatic periods, we plan to implement a follow-up protocol in future phases where symptomatic participants are resampled during the asymptomatic period of the same season in the following year to provide a more comparable seasonal control. To deeply understand transmission characteristics, we will integrate the tNGS-derived pathogen spectrum and seasonal trends with data from a parallel contact and mobility survey conducted on the same population. By linking symptomatic events with human contact patterns and mobility features, we aim to model the specific transmission dynamics of ARIs in Shanghai's high-density urban setting.

Our research highlights the critical role of continuous surveillance and advanced diagnostic techniques in managing respiratory infections. By comprehensively understanding the interaction between human behaviour, pathogen spread and clinical outcomes, we can formulate more robust and practical strategies to mitigate the impact of ARI in urban centres such as Shanghai.

Contributors XY, ZW, WZ, YZ, WW, HW and XC conceived the study. XY, ZW and WZ designed the protocol. WX, SL, HL and XD provided statistical support and advised on study design. JC, WW, HW and XC provided further feedback on the protocol. All authors have reviewed the protocol and approved the final manuscript. XC is the guarantor of the study.

Funding This work was supported by the Shanghai Municipal Science and Technology Major Project (grant no. ZD2021CY001), the National Natural Science Foundation of China (grant no. 82073612), the Three-year Action Program of Shanghai Municipality for Strengthening the Construction of Public Health System (2023–2025) (grant no. GWVI-1, GWVI-11.1-03 and GWVI-11.1-01) and Eastern Talent Plan Youth Project (2023).

Map disclaimer The depiction of boundaries on this map does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or its authorities. This map is provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. This study was approved by the Medical Research Ethics Committee of the Shanghai Center for Disease Control and Prevention (KY-2024-19). All patients will be asked to provide their written informed consent to participate. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iD

Weibing Wang https://orcid.org/0000-0002-4497-5251

REFERENCES

1 Roth GA, Mensah GA, Johnson CO, *et al*. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. *J Am Coll Cardiol* 2020;76:2982–3021.
2 Li ZJ, Yu LJ, Zhang HY, *et al*. Broad Impacts of Coronavirus Disease 2019 (COVID-19) Pandemic on Acute Respiratory Infections in China: An Observational Study. *Clin Infect Dis* 2022;75:e1054–62.
3 Pan L, Yuan Y, Cui Q, *et al*. Impact of the COVID-19 pandemic on the prevalence of respiratory viral pathogens in patients with acute respiratory infection in Shanghai, China. *Front Public Health* 2024;12:1230139.
4 Xu X, Zhang Y, Xu L, *et al*. Analysis of respiratory pathogen detection in hospitalized children with acute respiratory tract infections after ending the zero COVID policy. *Sci Rep* 2024;14:31784.
5 COVID-19 response further optimized with 10 new measures. Available: https://english.www.gov.cn/statecouncil/ministries/202212/08/content_WS63913c92c6d0a757729e4135.html?utm_source=chatgpt.com [Accessed 6 Jan 2025].

6 Conroy G. What's behind China's mysterious wave of childhood pneumonia? *Nature New Biol* 2023.
7 Getzzg. Effects of public health interventions and zero covid policy on paediatric diseases: a southern china study. JOGH; 2024. Available: https://jogh.org/2024/jogh-14-05011/ [Accessed 6 Jan 2025].
8 Li Z-J, Zhang H-Y, Ren L-L, *et al*. Etiological and epidemiological features of acute respiratory infections in China. *Nat Commun* 2021;12:5026.
9 Zhang J, Klepac P, Read JM, *et al*. Patterns of human social contact and contact with animals in Shanghai, China. *Sci Rep* 2019;9:15141.
10 McGowan CR, Takahashi E, Romig L, *et al*. Community-based surveillance of infectious diseases: a systematic review of drivers of success. *BMJ Glob Health* 2022;7:e009934.
11 Murarkar S, Gothankar J, Doke P, *et al*. Prevalence of the Acute Respiratory Infections and Associated Factors in the Rural Areas and Urban Slum Areas of Western Maharashtra, India: A Community-Based Cross-Sectional Study. *Front Public Health* 2021;9:723807.
12 Chew FT, Doraisingham S, Ling AE, *et al*. Seasonal trends of viral respiratory tract infections in the tropics. *Epidemiol Infect* 1998;121:121–8.
13 Sciaraffa R, Borghini A, Montuschi P, *et al*. Impact of air pollution on respiratory diseases in urban areas: a systematic review. *Eur J Public Health* 2017;27:117.
14 Sciaraffa R, BorghiniA, Montuschi P, *et al*. Impact of air pollution on respiratory diseases in urban areas: a systematic review: Daniele Ignazio La Milia. *Eur J Public Health* 2024.
15 Anugwom EE, Anugwom KN. Urbanization and the epidemiology of infectious diseases: towards the social framing of global responses. In: Rezaei N, ed. *Integrated Science of Global Epidemics*. Cham: Springer International Publishing 2023:307–28,
16 WHO mosaic respiratory surveillance framework. Available: https://www.who.int/initiatives/mosaic-respiratory-surveillance-framework [Accessed 6 Jan 2025].
17 Carducci A, Arzilli G, Atomsa NT, *et al*. Integrated environmental and clinical surveillance for the prevention of acute respiratory infections (ARIs) in indoor environments and vulnerable communities (Stell-ARI): Protocol. *PLoS One* 2024;19:e0309111.
18 Europe WHORO for. WHO regional office for europe guidance for sentinel influenza surveillance in humans. 2011.
19 Li S, Tong J, Liu Y, *et al*. Targeted next generation sequencing is comparable with metagenomic next generation sequencing in adults with pneumonia for pathogenic microorganism detection. *Journal of Infection* 2022;85:e127–9.
20 Yaxu Z, Xiao Y, Huanyu W, *et al*. Establishment and application of a prospective follow-up research method for acute infectious diseases in Shanghai community residents. *Shanghai Journal of Preventive Medicine* 2024;36:5–10.
21 Palmer LJ. UK Biobank: bank on it. *Lancet* 2007;369:1980–2.
22 Chen Z, Chen J, Collins R, *et al*. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011;40:1652–66.
23 UK Biobank Category. Summary diagnoses-hospital inpatient-health-related outcomes. 2002. Available: https://biobank.ctsu.ox.ac.uk/ukb/label.cgi?id=2002 [Accessed 3 Jan 2025].
24 Vu T-HT, Van Horn L, Achenbach CJ, *et al*. Diet and Respiratory Infections: Specific or Generalized Associations? *Nutrients* 2022;14:1195.
25 Cheetham NJ, Penfold R, Giunchiglia V, *et al*. The effects of COVID-19 on cognitive performance in a community-based cohort: a COVID symptom study biobank prospective cohort study. *EClinicalMedicine* 2023;62:102086.
26 Toivonen L, Hasegawa K, Waris M, *et al*. Early nasal microbiota and acute respiratory infections during the first years of life. *Thorax* 2019;74:592–9.
27 Li H, Wu X, Zeng H, *et al*. Unique microbial landscape in the human oropharynx during different types of acute respiratory tract infections. *Microbiome* 2023;11:157.
28 Wang C, Li X, Ning W, *et al*. Multi-omic profiling of plasma reveals molecular alterations in children with COVID-19. *Theranostics* 2021;11:8008–26.
29 de Steenhuijsen Piters WAA, Heinonen S, Hasrat R, *et al*. Nasopharyngeal Microbiota, Host Transcriptome, and Disease Severity in Children with Respiratory Syncytial Virus Infection. *Am J Respir Crit Care Med* 2016;194:1104–15.