Association between disease severity and co-detection of respiratory pathogens in infants with RSV infection

Gu-Lung Lin¹, Simon B Drysdale¹,², Matthew D Snape¹, Daniel O’Connor¹, Anthony Brown³, George MacIntyre-Cockett⁴, Esther Mellado-Gomez⁴,⁵, Mariateresa de Cesare⁴,⁶, M Azim Ansari³,⁴, David Bonsall⁴,⁷, James E Bray⁸, Keith A Jolley⁸, Rory Bowden⁸,⁹,¹⁰, Jeroen Aerssens¹¹, Louis Bont¹², Peter J M Openshaw¹³, Federico Martinon-Torres¹⁴,¹⁵,¹⁶, Harish Nair¹⁷,¹⁸, Tanya Golubchik⁷,¹⁹, Andrew J Pollard¹,², on behalf of the RESCEU Investigators

¹Oxford Vaccine Group, Department of Paediatrics, University of Oxford, and the NIHR Oxford Biomedical Research Centre, Oxford, United Kingdom
²Centre for Neonatal and Paediatric Infection, Institute for Infection and Immunity, St George’s, University of London, London, United Kingdom
³Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, United Kingdom
⁴Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom
⁵Wellcome Sanger Institute, Hinxton, United Kingdom
⁶Human Technopole, Milan, Italy
⁷Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
⁸Department of Biology, University of Oxford, Oxford, United Kingdom
⁹The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia
¹⁰Department of Medical Biology, University of Melbourne, Melbourne, VIC, Australia
¹¹Translational Biomarkers, Infectious Diseases Therapeutic Area, Janssen Pharmaceutica NV, Beerse, Belgium
¹²Department of Pediatrics, Wilhelmina Children’s Hospital, University Medical Center Utrecht, Utrecht, Netherlands, and the ReSViNET Foundation, Zeist, Netherlands
¹³National Heart and Lung Institute, Imperial College London, London, United Kingdom
¹⁴Translational Pediatrics and Infectious Diseases, Pediatrics Department, Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Spain
¹⁵Genetics, Vaccines, Infectious Diseases and Pediatrics Research Group (GENVIP), Instituto de Investigación Sanitaria de Santiago, University of Santiago de Compostela, Santiago de Compostela, Spain
¹⁶Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Centre for Global Health, Usher Institute, Edinburgh Medical School, University of Edinburgh, Edinburgh, United Kingdom

MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Sydney Infectious Diseases Institute, School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

T. G. and A. J. P. contributed equally to this work.

Correspondence: G.-L. Lin, MD, DPhil, Oxford Vaccine Group, Centre for Clinical Vaccinology and Tropical Medicine, Churchill Hospital, Oxford OX3 7LE, UK (guling.lin.oxford@gmail.com).
Abstract

BACKGROUND: Respiratory syncytial virus (RSV) is the leading cause of hospitalisation associated with acute respiratory infection in infants and young children, with substantial disease burden globally. The impact of additional respiratory pathogens on RSV disease severity is not completely understood.

OBJECTIVES: The objective of this study was to explore the associations between RSV disease severity and the presence of other respiratory pathogens.

METHODS: Nasopharyngeal swabs were prospectively collected from two infant cohorts: a prospective longitudinal birth cohort study and an infant cross-sectional study recruiting infants <1 year of age with RSV infection in Spain, the UK, and the Netherlands during 2017–20 [part of the REspiratory Syncytial virus Consortium in EUrope (RESCEU) project]. The samples were sequenced using targeted metagenomic sequencing with a probe set optimised for high-resolution capture of sequences of over 100 pathogens, including all common respiratory viruses and bacteria. Viral genomes and bacterial genetic sequences were reconstructed. Associations between clinical severity and presence of other pathogens were evaluated after adjusting for potential confounders, including age, gestational age, RSV viral load, and presence of comorbidities.

RESULTS: RSV was detected in 433 infants. Nearly one in four of the infants (24%) harboured at least one additional non-RSV respiratory virus, with human rhinovirus being the most frequently detected (15% of the infants), followed by seasonal coronaviruses (4%). In this cohort, RSV-infected infants harbouring any other virus tended to be older (median age: 4.3 vs. 3.7 months) and were more likely to require intensive care and mechanical ventilation than those who did not. Moraxella, Streptococcus, and Haemophilus species were the most frequently identified target bacteria, together found in 392 (91%) of the 433 infants (S. pneumoniae in 51% of the infants and H. influenzae in 38%). The strongest contributors to severity of presentation were younger age and the co-detection of Haemophilus species alongside RSV. Across all age groups in both cohorts, detection of Haemophilus species was associated with higher overall severity, as captured by ReSVinet scores, and specifically with increased rates of hospitalisation and respiratory distress. In contrast, presence of Moraxella species was associated with lower ReSVinet scores and reduced need for intensive care and mechanical ventilation. Infants with and without Streptococcus species (or S. pneumoniae in particular) had similar clinical outcomes. No specific RSV strain was associated with co-detection of other pathogens.

CONCLUSION: Our findings provide strong evidence for associations between RSV disease severity and the presence of additional respiratory viruses and bacteria. The associations, while not indicating causation, are of potential clinical relevance. Awareness of coexisting microorganisms could inform therapeutic and preventive measures to improve the management and outcome of RSV-infected infants.
**Introduction**

Human respiratory syncytial virus (RSV) is the leading cause of hospitalisation associated with acute lower respiratory tract infection (LRTI) in infants and young children worldwide. The global RSV epidemic is estimated to cause 33 million LRTIs annually, leading to 3.6 million hospitalisations and around 100,000 RSV-attributable overall deaths in children under 5 years old, with mortalities predominantly in low- and middle-income countries\(^1\).

The standard of care for RSV infection is supportive management. Palivizumab, a monoclonal antibody for RSV prevention, has been used in young children at high risk for severe RSV LRTI, but it requires monthly administrations and its use is limited to higher-income settings due to prohibitive cost\(^2\). Nirsevimab, another monoclonal antibody with an extended half-life, has recently been authorised for use to prevent severe RSV LRTI with a single injection in neonates and infants during their first RSV season in the United Kingdom (UK) (licence numbers: PLGB 17901/0370–0371) and European Union\(^3\).

Infants normally harbour commensal microorganisms in the upper respiratory tract regardless of respiratory symptoms, including bacteria with pathogenic potential such as *Streptococcus pneumoniae* and *Haemophilus influenzae*\(^4\), and viruses such as enteroviruses and coronaviruses\(^5\)\(^-\)\(^8\). Simultaneous detection of multiple viruses is frequent in both healthy children\(^8\) and children with respiratory infections\(^8\)\(^-\)\(^14\), and virus-virus interactions can be either synergistic or antagonistic. A recent study showed that co-infection with influenza A virus and RSV can form hybrid viral particles in vitro, which evade neutralising antibodies and broaden receptor tropism\(^15\). In contrast, human rhinovirus (HRV) has been shown to interfere with influenza virus, reducing the chance of co-detection and co-circulation of both viruses at the individual and population levels, respectively\(^16\). Throughout the first year of life, there are constant changes in the respiratory microbiome with increasing biodiversity\(^5\)\(^,\)\(^17\). Colonisation with *Moraxella*, *Haemophilus*, or *Streptococcus* spp. and pneumococcal disease have been shown to be associated with viral respiratory infections\(^5\)\(^,\)\(^18\)\(^,\)\(^19\).

Molecular testing and typing of respiratory viruses and common bacterial pathogens have been increasingly used in the clinical setting. However, despite the high prevalence of colonisation and co-infection, the impact of other potential pathogens on the presentation of RSV infection (e.g., disease severity) remains to be understood.

In this study we applied high-throughput RNA-based targeted metagenomic sequencing to simultaneously recover over 100 bacterial and viral pathogens from respiratory samples prospectively collected from RSV-infected infants enrolled in two European multicentre studies. We sought to explore the associations between RSV disease severity and the presence of additional respiratory...
pathogens, focusing on bacteria and viruses. A better understanding of microbial factors associated with RSV disease severity will help direct therapeutic and preventive measures and improve the management and outcome of infants with RSV infection.

Methods

Study design and clinical data collection

Nasopharyngeal swabs were prospectively collected from infants under 1 year of age with primary RSV infection from the community and hospitals in Spain, the UK, and the Netherlands during the 2017–20 RSV seasons. These infants were enrolled in two clinical studies of the REspiratory Syncytial virus Consortium in EUrope (RESCEU) project, a European multicentre project investigating epidemiology, immunology, and virology of RSV infection. The two studies have been previously described in detail.20,21

In the longitudinal birth cohort study (ClinicalTrials.gov identifiers: NCT03627572), healthy term neonates (born at ≥37 weeks’ gestation) were enrolled. Key exclusion criteria were a history of major congenital defects (e.g., congenital heart and/or lung disease, genetic, immunological and/or metabolic disorders), acute severe medical conditions (e.g., sepsis or asphyxia), or receipt of immunoglobulin, monoclonal antibodies, or an investigational vaccine or medication against RSV. When participants developed respiratory symptoms during the RSV seasons before their first birthday, a nasopharyngeal swab was taken within 3 days of symptom onset and tested for RSV using point-of-care qualitative molecular testing on the Alere™ i RSV assay (Abbott, Illinois, United States). Regardless of the results, the swabs were immersed in M4RT® transport medium after collection, aliquoted, and frozen at –80°C until use.

In the infant cross-sectional study (ClinicalTrials.gov identifiers: NCT03756766), RSV-infected infants under 1 year of age were enrolled from the community within 4 days of symptom onset and from hospitals within 2 days of admission during the RSV seasons. Infants who were previously healthy and who had pre-existing medical conditions were both eligible for inclusion in this study. Key exclusion criteria were a history of RSV infection, receipt of immunoglobulin or monoclonal antibodies, or exposure to an RSV investigational vaccine or medication. RSV infection was diagnosed using the Alere™ i RSV assay in community settings or by routine antigen or polymerase chain reaction (PCR) tests at the hospital laboratory in hospital settings. A nasopharyngeal swab was collected at the time of enrolment for every participant. For hospitalised participants, daily nasopharyngeal swabs were also collected until hospital discharge. These swabs were also frozen at –80°C until use.
Demographic and clinical information was gathered in both studies. Clinical outcomes were evaluated by the ReSViNet score\textsuperscript{22}, presence of fever, and requirement for hospitalisation, intensive care (i.e., admission to a high dependency unit or an intensive care unit), respiratory support, or invasive mechanical ventilation. The ReSViNet score accounts for seven clinical parameters, including feeding intolerance, medical intervention, respiratory difficulty, respiratory frequency, apnoea, general condition, and fever. The scores range from 0 to 20, with higher scores representing a more severe disease. Fever was defined as at least one episode of a tympanic or rectal temperature of $\geq 38^\circ$C during the acute infection.

The studies were conducted in accordance with the provisions of the Declaration of Helsinki and were approved by the relevant authorities and ethics committees at each site: Hospital Clínico Universitario de Santiago de Compostela, Comité de Ética de la Investigación de Santiago-Lugo (no. 2017/395) in Spain; the University of Oxford, the Health Research Authority (no. 231136), the NHS National Research Ethics Service Oxfordshire Research Ethics Committee A (no. 15/SC/0335) and South Central – Hampshire A (no. 17/SC/0522) in the UK; and the Medical Ethical Committee, University Medical Center Utrecht (no. 17/563) in the Netherlands. The parents or guardians of all participants provided written, informed consent.

**Nucleic acid isolation and targeted metagenomic sequencing**

Nucleic acid isolation and sequencing were performed as described previously\textsuperscript{23-25}. The NucliSENS\textsuperscript{®} easyMAG\textsuperscript{®} system (BioMérieux, Marcy-l’Étoile, France) was used for automatic total nucleic acid extraction from 500 $\mu$L of each sample, following the manufacturer’s instructions. Sequencing libraries were constructed using the SMARTer\textsuperscript{®} Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara Bio USA, California, United States), following a modified veSEQ-HIV protocol\textsuperscript{23,24,26}. A 10-$\mu$L aliquot of each library was pooled together, and 750 ng of the pooled library was pulled down with a predesigned SureSelect RNA Target Enrichment multi-pathogen probe set (Agilent, California, United States). This probe set, Castanet, consisting of 120-mer oligonucleotides, was designed using the algorithm devised in the veSEQ protocol\textsuperscript{27}. It targeted more than 100 potentially pathogenic bacteria and viruses, including both RSV-A and RSV-B (Table S1)\textsuperscript{28}. The post-capture libraries were amplified with 16 cycles of PCR, and then purified using AMPure XP.

Sequencing was performed on the Illumina MiSeq platform (Illumina, California, US) or the Illumina NovaSeq 6000 system, generating paired-end reads. The MiSeq platform was used for a batch of $\leq 96$ samples, and the NovaSeq 6000 system was used for a batch of 384 samples. Each 96-well sequencing plate also included one RSV-negative sample (collected from participants with acute
respiratory symptoms, testing negative by the Alere™ i RSV assay) and one pure M4RT® transport medium as negative controls. These negative controls were processed together with RSV-positive samples.

**Viral load measurement**

Viral load was determined using reverse transcription quantitative PCR (RT-qPCR) assays performed at GSK as previously described. The primers of this duplex RT-qPCR assay targeted the N gene for both RSV-A and RSV-B. The limit of detection was 304 and 475 copies/mL for RSV-A and RSV-B, respectively.

**Viral genome assembly and phylogenetic reconstruction**

RSV genomes were reconstructed using shiver as previously described. Briefly, reads were trimmed to remove adapters, random primers, and low-quality bases using Trimmomatic (v0.39) (option: Adapters:2:10:7:1:true LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:50). Trimmed reads were assembled into contigs using IVA (v1.0.8) and metaSPAdes (v3.14.1), and mapped to genotype-specific RSV references using shiver, with Bowtie 2 (v2.4.1) as the mapper (option: --very-sensitive-local --maxins 2000 --no-discordant --no-unal). Properly paired reads were retained and duplicates removed using Picard MarkDuplicates (v2.18.14, https://broadinstitute.github.io/picard/). Consensus sequences were generated by shiver, where base calling was supported by a minimum of two unique reads per position. RSV consensus sequences covering at least 70% of the coding sequences were used to reconstruct phylogenetic trees. For infants with multiple samples collected, only the sequence with the highest coverage was included. To put the study strains in a global context, we also included a subset of contemporaneous global RSV strains downloaded from GenBank on 4th December 2020 with at least 70% coverage of the coding sequences, collected between 2015 and 2019. Genomic sequences were aligned using mafft (v7.490) with the FFT-NS-i method. RAxML (v8.2.12) was used to reconstruct the maximum-likelihood phylogenies with the general time reversible nucleotide substitution model and gamma-distributed rate heterogeneity among sites. The R package ggtree (v2.2.4) was used for tree visualisation. Patristic distances, used to measure phylogenetic distances between sample pairs, were computed using the cophenetic function in the R package stats (v4.0.2). Maximum-likelihood phylogenies of enterovirus and human herpesvirus 6 (HHV-6) were reconstructed using RAxML with the same models as above.
Respiratory viral genome reconstruction

The Castanet probe set\textsuperscript{28} is designed to cover the known genetic diversity of respiratory viruses, as well as phylogenetically informative sequences of respiratory-associated bacteria. It includes full genomes of viruses <40 kb in length, and 20 kb of sequence of longer genomes such as human herpesviruses (HHV), representing around 10\% of the genome from the U23 to U37 genes. Although designed prior to the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), this probe set includes sufficient coronavirus sequences to capture 23 kb of the 29-kb genome of SARS-CoV-2 (Fig. S1)\textsuperscript{39}. The Castanet pipeline (Golubchik, \url{https://github.com/tgolubch/castanet}) was used to determine coverage of all target viruses.\textsuperscript{28} Reads originating from human cells were removed using BBMap (Bushnell, \url{https://sourceforge.net/projects/bbmap/}). The remaining host-depleted reads were mapped using Bowtie 2 to the references that covered known sequence variability of the target viruses. Read depth, genome coverage, and consensus sequences were generated for each virus from binary alignment/map (BAM) files.

Based on comparison with negative controls, viruses with more than 20–50 uniquely mapped reads and \(\geq30\%\) genome recovery were considered present, and those with <10 unique reads were considered absent. Viruses fulfilling neither criterion were considered equivocal. Further criteria varied between viruses due to different efficiencies of the target enrichment and sequence specificities; detailed criteria for each virus are shown in Table S2.

Enterovirus genotyping was performed using the web-based Enterovirus Typing Tool\textsuperscript{40} (\url{https://www.rivm.nl/mpf/typingtool/enterovirus/}), based on phylogenetic analysis of the VP1 gene. The genotypes of HHV-6 were determined by phylogenetic analysis of the HHV-6 partial genome consensus sequences and reference strains. These references included HHV-6A (strains: U1102, GS, and AJ; accession numbers: NC_001664, KC465951, and KP257584), HHV-6B (strains: Z29 and HST; accession numbers: NC_000898 and AB021506), chromosomally integrated HHV-6A (accession numbers: MG894371 and KY315540), and chromosomally integrated HHV-6B (accession numbers: KY316051 and MG894376).

Respiratory bacterial reconstruction

The Castanet probe set also covered the known variation of 53 rps genes of the target bacteria. These genes encode the bacterial ribosome protein subunits, and their allele characterisation can be used to classify bacteria into groups at all taxonomic and most typing levels\textsuperscript{41}. A database containing catalogued rps gene variation has been developed for ribosomal multilocus sequence typing (rMLST)\textsuperscript{41}.
Reconstruction of the bacterial genetic sequences was based on the rMLST scheme. Contig sequences assembled by metaSPAdes were matched against all known alleles for each rps gene to find exact matches. The PubMLST multi-species isolate database, integrating curated allelic and species information, was then searched to identify the matching species (algorithm available at https://pubmlst.org/species-id). The PubMLST RESTful application programming interface was applied to efficiently analyse all samples.

Bacteria supported by at least two exact and unique allelic matches were considered present. A ‘unique’ allelic match is an allelic pattern that is linked to exactly one taxon at its lowest taxonomic rank or two bacteria with one being an unspecified species of the same genus as the other specified species (e.g., if Moraxella sp. and Moraxella catarrhalis were both linked to an allelic pattern, Moraxella catarrhalis was considered a unique match). Bacteria supported by only one exact and unique match were considered equivocal. The rest were considered absent. The number of reads assigned to a certain species was determined by taxonomic classification of the trimmed reads using Kraken 2 (v0.39). A Krona chart was used to visualise the composition of identified bacteria.

**Statistical analyses**

For infants with multiple samples collected, species that were detected in at least one sample collected while RSV was detectable or within 7 days of enrolment (whichever is longer) were considered present; species that were not detected in any sample collected while RSV was detectable or within 7 days of enrolment (whichever is longer) were considered absent; the rest were considered equivocal. Infants with an equivocal presence of a species were excluded from clinical outcome comparisons.

Phylogenetic signal for a co-detected species on RSV phylogenies was evaluated using the $D$ statistic based on a permutation testing framework. An equivocal presence status was considered absence in this analysis. A $D$ value of 0 indicates a strong phylogenetic signal under the Brownian motion model of evolution (i.e., conserved trait evolution), and a value of 1 indicates no phylogenetic signal (i.e., a random phylogenetic distribution). A negative $D$ value indicates that the binary trait is more conserved than the expectations of the Brownian motion model, whereas a value greater than 1 indicates phylogenetic overdispersion. The $D$ statistical tests were performed using the phylo.d function in the R package caper (v1.0.1).

When comparing clinical outcomes between different groups of infants, multiple linear regression (for continuous outcome variables), multivariable logistic regression (for dichotomous outcome variables), and proportional odds ordered logistic regression (for ordered outcome variables) were used to adjust for covariates. Covariates included age, gestational age, sex, comorbidity, sampling season...
and country, study, viral subgroup, peak RSV read count, and the duration between symptom onset and sampling. Models with different combinations of covariates were tested, and the model with the lowest Akaike information criterion (AIC) was selected. A post hoc adjustment for multiple comparisons with the Benjamini–Hochberg method was applied to determine false discovery rate–corrected Q values in all clinical outcome comparisons. The effect size of the presence of a pathogen on a clinical outcome was evaluated using Cohen’s $f^2$. An $f^2$ value between 0.02 and <0.15 represents a small effect size; a value between 0.15 and <0.35, medium; and a value of ≥0.35, large\(^{47}\).

All statistical analyses were performed using R (v4.0.2)\(^{48}\).

## Results

### Study and sample populations

Overall, 839 nasopharyngeal swabs from a total of 440 RSV-infected infants, enrolled from Spain, the UK, and the Netherlands during 2017–2020, were sequenced (Table S3). One infant’s sample could not be sequenced, and no RSV reads were recovered from six infants’ samples ($n = 7$), five of which had undetectable viral load. These seven infants and eight samples were excluded from further analysis. A total of 125 hospitalised infants in the infant cross-sectional study had more than one swab collected (mean ± SD, 4.2 ± 2.5).

The median age of the remaining 433 infants was 4.1 months [interquartile range (IQR), 1.9–7.5 months] at the time of the RSV infection. Females accounted for 45%. Nine percent of the infants had pre-existing medical conditions, including prematurity with or without bronchopulmonary dysplasia, ventricular septal defect, and other congenital abnormalities. Younger infants tended to have a higher ReSVinet score and were more likely to require hospitalisation, intensive care, respiratory support, and mechanical ventilation than older infants, whereas a greater proportion of older infants had a fever than younger infants (<3 months vs. 3 to <6 months vs. 6 to <12 months) (Table S4). Due to the differences in study design, infants in the infant cross-sectional study were likely to be younger and have more severe disease than those in the longitudinal birth cohort study (Table S5).

We previously showed\(^{23}\) that the number of uniquely mapped RSV reads from our protocol is highly correlated with RSV viral load (Pearson correlation, $R^2 = 0.78$, $P = 5 \times 10^{-197}$; Fig. S2). Thus, we used the number of unique RSV reads as a surrogate for viral load for the analyses that follow. Of the 831 samples where RSV sequencing was possible, all 597 (72%) samples with at least 70% of the RSV genome reconstructed had a median of 11,851 unique RSV reads (range, 285 to 810,113), corresponding to a median viral load of $1.2 \times 10^7$ copies/mL (range, 505 to $4.3 \times 10^9$).
Similar proportions of the 433 infants had RSV-A and RSV-B (220 vs. 204; the remaining 9 could not be genotyped due to evidence of mixed infection or contamination and were excluded from genomic analyses). RSV-B dominated the 2017–18 and 2018–19 RSV seasons, accounting for 71% and 55% of the infections, respectively, whereas RSV-A was the predominant subgroup during the 2019–20 RSV season, accounting for 66% of the infections (Table S6). Phylogenetic analyses classified all RSV-A strains into genotype ON1 and all RSV-B strains into genotype BA (Fig. S3). There was no evidence to support differences in severity between RSV-A and RSV-B infections (Table S7).

### RSV and co-detected respiratory viruses

Co-infection with multiple respiratory viruses, defined as co-detected presence of at least two respiratory viruses, was common among the infants in this study. At least one respiratory virus in addition to RSV was recovered in 106 (24%) of the 433 RSV-infected infants, including 10 infants with two additional viruses (Fig. 1). Non-RSV viruses were present at substantial viral load: in the quantitative sequencing method we used, over 90% of the non-RSV viruses had sufficient unique reads to enable reconstruction of at least 50% of their genome, consistent with viral loads in a range similar to that of RSV and suggestive of active replication/infection (Fig. S4). RSV-infected infants with any viral co-infection tended to be older than those without [median (IQR), 4.3 (2.3–8.2) vs. 3.7 (1.7–6.9) months; Mann–Whitney U test, P = 0.031] (Table S8 and Fig. S5). There was no correlation between the viral load of non-RSV viruses and RSV viral load or genotype, and likewise no correlation between viral load of non-RSV viruses and infant age, sex, or clinical outcome.

Human rhinovirus (HRV) was the virus most frequently found alongside RSV (63/433 infants, 15%), followed by seasonal human coronaviruses (HCoVs, 4%) and adenoviruses (2%) (Table S9). We did not find any SARS-CoV-2 in 76 swabs collected from 35 infants in January and February 2020. VP1 genotyping of the HRV detected in the 63 infants showed that HRV-A, HRV-B, and HRV-C accounted for 41% (26/63), 29% (18/63), and 30% (19/63), respectively, as well as three cases of enterovirus D68 (Fig. S6). Human herpesvirus 6 (HHV-6) was detected in seven (2%) infants, including two HHV-6A and five HHV-6B (Fig. S7). Phylgenetic segregation of HHV-6A was strongly suggestive of chromosomally integrated HHV-6A in our samples49.

We examined the impact of additional viruses on markers of disease severity after adjusting for covariates, including age, gestational age, RSV viral load, and presence of comorbidities (Fig. 2). There was strong evidence that infants with viral co-infection were more likely to require intensive care and mechanical ventilation than those without (Fig. 2, Table S8). Analysing each non-RSV virus
family separately (i.e., enterovirus, HRV, HCoV, adenovirus, and HHV-6) showed that infants co-infected with HCoV tended to be older than those without [median (IQR), 7.7 (3.84–10.0) vs. 4.0 (1.8–7.4) months; Mann–Whitney U test, P = 0.012]. There was also some evidence that infants co-infected with HHV-6 were more likely to require intensive care than those without (OR, 10.4; 95% CI, 1.7–86.9; P = 0.015; Cohen’s F² = 0.184), but limited sample numbers of HHV-6 preclude a robust comparison (not significant after adjusting for multiple comparisons).

Figure 3 shows the distribution of the non-RSV viruses on the RSV phylogenies. There was no significant phylogenetic signal for the presence of these viruses on the RSV phylogenies with D values ranging from 0.80 to 1.33, indicating no specific RSV clade was linked with the presence of other viruses (Table S10).

**RSV and co-detected bacterial pathogens**

We detected target bacterial species in the vast majority of infants (406/433, 94%). Among the 21/27 infants with no detected target bacteria for whom information on antibiotic usage was available, 10 did not receive any antibiotics during the RSV infection, 10 had received antibiotics before sample collection, and one received antibiotics without information on the timing of antibiotic dosing. These 27 infants had a mean ± SD peak RSV read count of 3.7 ± 1.2 log₁₀, falling in the lower quartile of the study population (Fig. S4), suggesting these may have been poorer quality swabs.

Excluding these 27 infants, the mean ± SD number of target bacterial species was 3.4 ± 2.0, with *Moraxella* (76% of all 433 RSV-infected infants), *Streptococcus* (68%), and *Haemophilus* (45%) being the most frequently found genera (Table S11, Figs. S8 and S9). These three genera together were found in 392/433 infants (91%). Among each of these genera, *M. catarrhalis*, *S. pneumoniae*, and *H. influenzae* were the most commonly identified species, and had read numbers consistent with substantial bacterial load (Fig. S4). Infants with any of these three bacterial genera were older than those without [median (IQR), 4.6 (1.9–7.7) months vs. 2.0 (1.1–3.4) months; Mann–Whitney U test, P = 0.002].

Five of the recovered bacterial species were further classified down to the ribosomal sequence type (rST) level in some samples: *M. catarrhalis*, *S. pneumoniae*, *H. influenzae*, *Escherichia coli*, and *Neisseria meningitidis*. Most rSTs were unique in the dataset, consistent with extensive bacterial genetic diversity at the population level. Only a few isolates shared the same rST—*M. catarrhalis* rST 105842 was found in two infants; each of the *S. pneumoniae* rSTs 644, 12187, 13906, and 104663 were found in two infants; and each of the *H. influenzae* rSTs 24162, 49634, 89110, and 133371 were also found in two infants (Table S12).
Similarly to the analysis of viral co-infections, we investigated the correlation between disease severity and the presence of *Moraxella*, *Haemophilus*, and *Streptococcus*, after adjusting for confounders as above. Compared with infants without any *Haemophilus* spp., infants with *Haemophilus* spp. were more likely to have a higher ReSVinet score and fever and require hospitalisation, respiratory support, and mechanical ventilation (Figs. 2 and 4A, Table S13). Although the effect size of *Haemophilus* presence on most of these clinical outcomes was small (Cohen’s $f^2$ between 0.02 and <0.15) (Table S13), presence of *Haemophilus* spp. correlated with around a 2.6-point increase in ReSVinet scores across all age groups, in both cohorts that made up the study population (Fig. 4A). As expected, ReSVinet scores correlated with age for the infants enrolled in the infant cross-sectional study (Pearson correlation, $P = 3.3 \times 10^{-9}$) but not for those in the longitudinal birth cohort study ($P = 0.44$) (Fig. 4A). This was because infants in the infant cross-sectional study were recruited on presentation with an RSV infection, while those in the longitudinal birth cohort study were prospectively followed up to detect RSV positivity and thus more likely to include mild cases.

In contrast to the observation of higher ReSVinet scores with *Haemophilus* spp., infants with *Moraxella* spp. tended to have a lower ReSVinet score (especially in infants under 6 months of age; Fig. 4B) and were less likely to require intensive care and mechanical ventilation than those without (Fig. 2). Infants with and without *Streptococcus* spp., or *S. pneumoniae* in particular, had similar clinical outcomes (Figs. 2 and 4C).

Given the different direction of effect of *Haemophilus* spp. vs. *Moraxella* spp. on disease severity, we hypothesised that antibiotic administration may have been higher among infants with *Haemophilus* spp., potentially indicating secondary bacterial infection. Information on the use of antibiotics was available in 361 RSV-infected infants, of which 25% received antibiotics during their RSV episode. Consistent with our hypothesis, a higher proportion of infants with *Haemophilus* spp. received antibiotics than those without (33% vs. 20%; Chi-square test, $P = 0.010$); and of those receiving antibiotics, 46% had received antibiotics before the first sample where *Haemophilus* spp. were identified. In contrast, a lower proportion of infants with *Moraxella* spp. received antibiotics than those without (19% vs. 43%; $P = 2.2 \times 10^{-5}$). The rates of receiving antibiotics were similar in infants with and without *Streptococcus* spp. (22% vs. 29%; $P = 0.363$), but fewer infants with *S. pneumoniae* received antibiotics than those without (19% vs. 30%; $P = 0.019$).

Figure S10 shows the distribution of commonly co-detected bacterial species (namely *M. catarrhalis*, *S. pneumoniae*, *H. influenzae*, and *Staphylococcus aureus*) on the RSV phylogenies. *S. pneumoniae* was non-randomly distributed across the RSV-B phylogeny ($P = 0.026$) with a $D$ value of 0.77, but the distribution was significantly different from the expectations under Brownian motion ($P < 0.001$), indicating a low phylogenetic signal (Table S10). There was no significant phylogenetic
signal between the RSV phylogenies and the presence of other commonly co-detected bacteria, with
$D$ values ranging from 0.76 to 1.01.

**Discussion**

Using RNA-based quantitative sequencing, we captured a broad range of respiratory pathogens in infants with RSV, with the aim of detecting additional viral and bacterial pathogens and examining correlates of disease severity. We found strong genomic evidence for presence of at least one additional virus in a high proportion (24%) of the 433 infants with active RSV infection, in keeping with previous reports of viral co-infection rates of 15–34% in children of all ages with respiratory infections. This is likely a conservative estimate of the true co-infection rate, as our stringent detection criteria would necessarily miss very low-viral load co-infections. We additionally assessed the presence of respiratory bacteria with high pathogenic potential, and found that the three most commonly detected genera—*Moraxella*, *Streptococcus*, and *Haemophilus*—were detected in an overwhelming majority (94%) of the tested samples.

In our study population, presence of respiratory viruses in addition to RSV correlated with increased need for intensive care and mechanical ventilation, but the overall impact of viral co-infection on severity was small. Likewise, the presence of most bacterial species had no negative effect on severity. Indeed, presence of *Moraxella*, the most commonly detected bacteria, was associated with less severe presentation, with reduced probability of requiring intensive care and mechanical ventilation. In contrast, there was strong evidence that the presence of *Haemophilus* spp. was associated with severe RSV disease independent of age, study cohort, or RSV genotype, and this effect was demonstrated by multiple clinical outcome variables.

There is a general perception that an inverse relationship exists between age and rates of viral co-infections—infants and children have a higher rate of viral co-infections than adults. However, at the lower extreme of age, under 12 months, we observed that older infants were more likely to have multiple respiratory viruses than younger infants. This may be due to presence of maternal antibodies in the youngest infants and a higher probability of social contacts (e.g., attending nursery) in older infants, exposing them to more respiratory viruses.

In our dataset, HRV was the most commonly co-infecting virus, found in 15% of the RSV-infected infants. The Pneumonia Etiology Research for Child Health (PERCH) study showed that HRV was the second most common virus causing severe pneumonia in children after RSV, and the most common virus in healthy children or those without severe pneumonia. Although we did not look at HRV in the general population, it has been shown that the presence of RSV reduces the chance of concurrent HRV...
infection\textsuperscript{56,57}, and the rate of HRV detection in our study was in accordance with that in previous studies which demonstrated interference between RSV and HRV\textsuperscript{56}. Other viruses were co-detected with RSV in a further 9% of the infants. These viruses were taxonomically diverse and no single virus apart from HRV was found in >5% of the infants.

Co-detected viruses could represent either active replication of an additional virus (either newly acquired or reactivated) or leftover nucleic acid from a previous infection; the additional virus may be either as a non-pathogenic ‘bystander’ or a true disease-causing agent. Although not definite, unique read count as a surrogate for viral load may provide a clue to differentiate between these possibilities. Co-detected viruses with low viral load may come from prolonged shedding from a previous infection, whereas those with high viral load are more likely to cause symptoms. In this study, we focused on viruses with evidence of substantial viral load based on read numbers, as being consistent with active replication and therefore possible co-infection.

Among the bacterial genera we tested, the presence of \textit{Haemophilus} spp. was significantly associated with severe RSV disease, consistent with the results of previous studies\textsuperscript{5,58,59}. A recent study also related high levels of \textit{Haemophilus} spp. with early recurrence of respiratory symptoms following a lower respiratory tract infection\textsuperscript{60}. In contrast, the presence of \textit{Streptococcus} spp. (or \textit{S. pneumoniae} specifically) was not associated with any of the tested clinical outcomes in our study. Acquisition, but not carriage, of \textit{S. pneumoniae} by infants has been associated with medically attended infectious illness\textsuperscript{61}. Contrary to our study, a previous study showed an association between \textit{Streptococcus} abundance and an increased likelihood of hospitalisation for RSV infection\textsuperscript{58}. \textit{Haemophilus} spp. may have caused secondary bacterial infection in the RSV-infected infants since a greater proportion had fever and received antibiotics than those without \textit{Haemophilus} spp. In contrast, the evidence for clinically suspected secondary bacterial infection with \textit{Streptococcus} spp. or \textit{Moraxella} spp. was insufficient. Overrepresentation of \textit{H. influenzae} and \textit{Streptococcus} spp. in RSV infection has been associated with enhanced Toll-like receptor (TLR) signalling\textsuperscript{58}, mucosal chemokine (C-X-C motif) ligand 8 (CXCL8, also known as IL-8) responses\textsuperscript{62}, IL-17A signalling\textsuperscript{58}, and inflammatory responses\textsuperscript{63}. CXCL8 and IL-17A signalling contribute to macrophage and neutrophil activation and recruitment\textsuperscript{58}, inducing bronchoalveolar neutrophil infiltration\textsuperscript{64}.

Despite being the most frequently found bacteria, \textit{Moraxella} spp. were associated with less severe RSV infection that did not require intensive care or mechanical ventilation in our study, suggesting a potential protective effect of these microorganisms on RSV infection or their role as a non-pathogenic ‘bystander.’ Previous studies also showed that \textit{Moraxella} spp. were more frequently found in outpatients than hospitalised children\textsuperscript{58} and children without severe pneumonia (including healthy children) than those with\textsuperscript{8}, and significantly associated with fever in infants with RSV lower
respiratory tract infections. Although fever is one of the parameters in the ReSVinet score, we demonstrated elsewhere that it did not correlate with the other clinical outcome measures in a subset of these RSV-infected infants. In this study, older infants were more likely to have a fever but generally presented with a milder RSV infection than younger infants. Further studies are required to evaluate the underlying immune mechanisms contributing to the potential protective effect of Moraxella spp.

Our study has several limitations. Due to regional and age-specific differences in respiratory flora, results generated from this study may not be generalisable to infants in low- and middle-income countries or different age groups (e.g., older adults). Although our method covers a comprehensive list of respiratory-associated viruses, co-detection does not imply capacity to initiate infection in the absence of RSV; for instance, we cannot determine whether HHV-6 detected in the infants was a primary respiratory pathogen or reactivation of a latent infection. More broadly, the associations observed in this study between disease severity and co-detection of respiratory pathogens in RSV-infected infants should be interpreted with caution and do not indicate causation. Such an indication would only come from a therapeutic trial. In addition, the rMLST scheme included in our method was able to classify bacteria to the species level or even the rST level, but the link between rST and serotype is currently unknown, which makes it difficult to classify detected S. pneumoniae and H. influenzae. Lastly, although multiple commensal and pathogenic viruses and bacteria were recovered, we did not attempt to characterise the whole microbial community. Microbiome analyses using samples collected in the RESCEU studies are ongoing.

This study deepens our understanding of the associations between RSV disease severity and the co-detected respiratory pathogens. Targeted metagenomic sequencing with the Castanet probe set enables simultaneous detection of combinations of bacterial and viral pathogens that may be of particular concern in RSV and other infections. Our results can assist in identifying patients at risk for severe RSV infection, directing the development of therapeutic and preventive measures, and improving the management and outcomes of individuals infected with RSV. Future studies evaluating the transcriptomic and immunological profiles of RSV infection should also consider the presence of other bacterial and viral pathogens.

RESCEU investigators

Harish Nair, Harry Campbell, Steve Cunningham (University of Edinburgh); Philippe Beutels (Universiteit Antwerpen); Louis Bont, Joanne Wildenbeest, Debby Bogaert (University Medical Centre Utrecht); Andrew Pollard, Matthew Snape, Simon Drysdale, Gu-Lung Lin, Daniel O’Connor.
Elizabeth Clutterbuck, Joseph McGinley, Hannah Robinson (University of Oxford); Peter Openshaw, Ryan Thwaites, Dexter Wiseman (Imperial College London); Federico Martinon-Torres, Alberto Gómez-Carballe, Carmen Rodriguez-Tenreiro, Irene Rivero-Calle, Ana Dacosta-Uribeta, Sara Pisschedda (Servicio Galego de Saude); Terho Heikkinen (University of Turku and Turku University Hospital); Adam Meijer (National Institute for Public Health and the Environment); Thea K Fischer (Statens Serum Institut); Maarten van den Berge (University of Groningen); Carlo Giaquinto (PENTA Foundation); Michael Abram (AstraZeneca); Kena Swanson (Pfizer); Bishopy Rizkalla (GSK); Charlotte Vernhes, Scott Gallichan (Sanofi Pasteur); Jeroen Aerssens, Deniz Öner (Janssen); Veena Kumar (Novavax); Eva Molero (Team-It Research).

Data availability

All raw sequencing reads of the samples included in this study are accessible at the European Nucleotide Archive under the study accession PRJEB34042 (https://www.ebi.ac.uk/ena/data/view/PRJEB34042). The RSV consensus sequences included in the phylogenetic analyses have been deposited in GenBank. The accession numbers can be found in Table S14.

Acknowledgements

This work was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre, the NIHR Thames Valley and South Midlands Clinical Research Network, the British Research Council, and the Respiratory Syncytial Virus Consortium in Europe (RESCEU) project. RESCEU has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (grant number 116019). This Joint Undertaking receives support from the European Union Horizon 2020 Research and Innovation Program and European Federation of Pharmaceutical Industries and Associations.

Author contributions

critically revised it. All authors have approved the submitted version and agreed to submit the manuscript. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy and integrity of any part of the work are appropriately investigated and resolved.

**Competing interests**

G.-L. L. is currently an employee of Mundipharma Research Ltd., a position he commenced subsequent to the completion of this study. S. B. D. has been an investigator for clinical trials of vaccines and antimicrobials for pharmaceutical companies, including AstraZeneca, Merck, Pfizer, Valneva, Iliad, Sanofi, and Janssen, and previously sat on RSV advisory boards for Sanofi and Merck. M. D. S. has been an investigator on behalf of the University of Oxford for studies funded or supported by vaccine manufacturers, including Pfizer, GSK, Novavax, MCM vaccines and Janssen. M. D. S. is currently an employee of Moderna Biotech UK, a position he commenced subsequent to the completion of this study. M. A. A. is supported by a Sir Henry Dale Fellowship, jointly funded by the Royal Society and Wellcome Trust (220171/Z/20/Z). J. E. B. and K. A. J. are funded on a Wellcome Trust Biomedical Resources Grant (218205/Z/19/Z, PubMLST: Disseminating and exploiting bacterial diversity data for public health benefit). J. A. is an employee of Janssen Pharmaceutica NV. P. J. M. O. has received honoraria from GSK, Pfizer Inc, Sanofi Pasteur, Seqirus, and Janssen for taking part in advisory boards and expert meetings and for acting as a speaker in congresses outside the scope of the submitted work. P. J. M. O. is also a National Institute for Health and Care Research (NIHR) Senior Investigator. F. M.-T. has received honoraria from GSK, Pfizer Inc, Sanofi Pasteur, MSD, Seqirus, and Janssen for taking part in advisory boards and expert meetings and for acting as a speaker in congresses outside the scope of the submitted work. F. M.-T. has also acted as principal investigator in randomised controlled trials of the above-mentioned companies as well as Ablynx, Regeneron, Roche, Abbott, Novavax, and MedImmune, with honoraria paid to his institution. F. M.-T. receives support for his research activities from the Instituto de Salud Carlos III (Proyecto de Investigación en Salud, Acción Estratégica en Salud): Fondo de Investigación Sanitaria (FIS;PI1601569/PI1901090/PI22/00406) del plan nacional de I+D+I and ‘fondos FEDER’. H. N. received grants from Innovative Medicine Initiative, NIHR, Bill and Melinda Gates Foundation, WHO, and Pfizer. H. N. also received consultancy or honoraria and speaker fees from Sanofi, Merck, Novavax, ReViral and GSK (all paid to institution). A. J. P. is chair of the UK Department of Health and Social Care’s Joint Committee on Vaccination and Immunisation. A. J. P. has also provided advice to Shionogi on COVID-19 vaccines and his institution receives research funding from NIHR, Bill & Melinda Gates Foundation, Wellcome Trust, Medical Research Council and AstraZeneca for vaccine
research. All other authors report no potential conflicts. The views expressed in this article are those of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the organizations with which the authors are employed/affiliated.
References


Figure 1. Genome coverage of all detected viruses in RSV-infected infants. Each row represents an individual infant. Dots on the Y axis are the 10 infants who had two other viruses identified in addition to RSV. Genome coverage is illustrated by the coloured gradient, defined as proportion of targeted genome that was reconstructed. Viruses are ordered in decreasing prevalence from left to right. RSV denotes respiratory syncytial virus; HRV, human rhinovirus; AdV, adenovirus; HCoV, human coronavirus; HBoV, human bocavirus; CMV, human cytomegalovirus; EV, enterovirus; HPeV-A, human parechovirus A; HPIV, human parainfluenza virus; and IFV-C, influenza C virus.
Figure 2. Clinical outcomes by presence and absence of pathogen groups. Odds ratios were adjusted for covariates including age, gestational age, sex, comorbidity, sampling season and country, study, RSV subgroup, peak RSV read count, and the duration between symptom onset and sampling using ordered logistic regression or multivariable logistic regression. Data in presence and absence are shown as no./total no. (%). Black dots represent the mean odds ratio, sized in proportion to Cohen’s.
$f^2$, ranging from 0.0009 (fever for *Streptococcus* spp.) to 0.0958 (mechanical ventilation for *Streptococcus* spp.). CI denotes confidence interval, and PICU paediatric intensive care unit.
Figure 3. Distribution of co-detected viruses on the (A) RSV-A and (B) RSV-B maximum-likelihood phylogenies. The phylogenies were reconstructed from samples in which at least 70% of the RSV coding sequences were recovered, using RAxML with the general time reversible nucleotide substitution model and gamma-distributed rate heterogeneity among sites. For infants with multiple samples collected, only the sample with the highest coverage was included. The trees were midpoint rooted. The sampling country of each strain is illustrated by tip colour. The scale bars represent the number of nucleotide substitutions per site. Human parechovirus A (HPeV-A) was not detected with any of the RSV-A strains; enterovirus B (EV-B) and human coronavirus 229E (HCoV-229E) were not detected with any of the RSV-B strains. HRV denotes human rhinovirus; AdV, adenovirus; HHV-6, human herpesvirus 6; HBoV, human bocavirus; CMV, human cytomegalovirus; and HPIV, human parainfluenza virus.
Figure 4. Relationships between ReSVinet score and age, and density plots of ReSVinet score, stratified by presence status of (A) *Haemophilus* spp., (B) *Moraxella* spp., and (C) *Streptococcus* spp. Results from the longitudinal birth cohort study and infant cross-sectional study are also shown separately in panel A. On the left panel, each dot represents an individual infant, and the lines are the regression lines with the shaded area representing the 95% confidence interval. On the right panel, dashed lines represent the mean ReSVinet score.