Appendix 1: Description of Studies used to identify cases and controls.

We undertook a case-control study to evaluate the clinical and pulmonary function sequelae of African children hospitalized with RSV LRTI during infancy at Chris Hani Baragwanath Academic Hospital (CHBAH) from 01 April 2016 to 31 December 2019.

Inclusion Criteria

- Term infants hospitalized with severe and very severe RSV LRTI

Exclusion criteria

- Infants with a birth weight of less than 2.5kg
- Infants with any underlying congenital (ex. congenital cardiac disease, hydrocephalus), genetic (ex. trisomy 21) or medical diagnosis (ex. neurological disability, such as neuromuscular disease or cerebral palsy, hepatic abnormalities such as biliary atresia, and musculoskeletal disorders such as osteogenesis imperfecta) that may affect respiratory function
- A lower respiratory tract infection in the 4 weeks preceding the pulmonary function testing

Sample size calculation

Based on a 1:1 case-control ratio, we estimated that a sample size of 567 cases and 567 controls at the one-year visit would be required to detect a 20% difference in lung resistance between cases and controls with an 80% power, assuming a standard deviation of 1.2; and a sample size of 100 cases and 100 controls would be required to detect a 40% difference in reactance between cases and controls with an 80% power, assuming a standard deviation of 1.2.

We estimated that two consecutive RSV seasons (2016 and 2017) would be required to reach the resistance measurement targeted sample size. These numbers were not achieved over the two-year period and the study was extended to a portion of a third RSV season (2018) – a power calculation based on the enrolled number of participants was undertaken (Table 2.1). The enrolled sample size at one year of age testing was insufficiently powered (53% power) to detect a 20% difference in resistance between cases and controls, (sufficiently powered to detect a 19.9% difference), but sufficiently powered to detect a 40% difference between cases and controls for reactance (98% power).

Table 1: Sample size and power calculations

Number of cases	**Power to detect 20% difference in resistance***	**Power to detect 40% difference in reactance****
100	21.5%	81.9%
200	38.3%	98.3%
300	53.1%	99.9%
400	65.3%	
500	74.9%	
600	82.2%	
700	87.6%	

* based on a two-sample t test power calculation; delta: 0.2, standard deviation: 1.2, significance level: 0.05

** based on a two-sample t test power calculation; delta: 0.49, standard deviation: 1.2, significance level: 0.05

Identification of eligible cases

Two surveillance studies (undertaken by the Respiratory and Meningeal Pathogens Research Unit (RMPRU) at CHBAH from December 2014) of hospitalized children were used to identify RSV LRTI cases for eligibility into our study: Surveillance of pathogen-specific causes of pneumonia and diarrhoea hospitalization in children (HREC: 131109) and Surveillance of Severe Childhood Illness in Soweto, South Africa / Babies of Soweto study (HREC: 140904). Cases were defined as term infants hospitalised for severe or very severe RSV LRTI. Research nursing staff reviewed admission books for the admission diagnoses of LRTI in children under five years of age admitted to one of five paediatric wards. Enrolment of LRTI cases occurred seven days a week from 8:00 am until 4:00 pm. In August 2016, ward surveillance was reduced to five days per week from Monday through to Friday; an admission over the weekend that was still present on the Monday morning was screened for possible inclusion in the study.

Caregivers were approached by research staff and written informed consent was obtained. Upon caregiver consent, the infant's demographic, clinical and laboratory data were collected on a paper based case report form (CRF). Additionally, research staff reviewed and made copies of the patient's

hospital charts for auditing and quality control purposes. After enrollment, patient logs were periodically reviewed to capture any changes to the level of care, including admission to the intensive care unit. Additional laboratory data, which may not have been available at the time of enrollment, was accessed through the National Health Laboratory Service (NHLS) Trakcare® system. All patient CRFs were filled out by hand and delivered to the research unit on a daily basis for auditing. After auditing by a research clinician for accuracy and completeness, the CRFs were entered into a centralized and secure electronic database system using Microsoft Visual Studio forms application and the the Microsoft SQL server database platform housed at the RMPRU.

All children with a LRTI enrolled under the surveillance studies had a nasopharyngeal swab (NPS), nasopharyngeal aspirate (NPA) or induced sputum (IS) taken for the identification of respiratory pathogens. Specimens were collected as soon as possible after arrival to the hospital, preferably within the first 72 hours (max 168 hours) of symptom onset and acute phase of illness to obtain a sample when viral shedding was greatest. The NPS was a flocked swab with a plastic shaft (FLOQS, Copan Flock Technologies, Brescia, Italy). The technique for sample collection was as follows: the swab was placed into the patient's nostril and gently advanced in the direction of the infant's ear until it was estimated to be at the mid-turbinate point or halfway between the opening of the nostril and the ear. It was then rotated several times, gently removed and placed into a collection tube with universal transport medium (UTM). Samples were placed on ice and transported to the RMPRU laboratories for processing and analysis. If the patient was intubated, a NPA was collected. Details of the date, time and name of the sampler were entered in the patient's CRF.

At the RMPRU laboratory total nucleic acid extraction using the automated NucliSENS® easyMAG® nucleic acid extraction machine was performed. A 1-Step multiplex polymerase chain reaction (PCR) assay developed at the RMPRU, which detects RSV A, RSV B, Human metapneumovirus, Influenza A, Influenza B and *Bordetella pertussis* (using IS481), was performed on the NPS. (Details of this SOP are found in appendix 2: Multiplex PCR for the detection of human Respiratory Syncytial Virus A and B, Human metapneumovirus, Influenza A and B and Bordetella species, Version 2). A cycle threshold (Ct) value of <37 was used as a cutoff for identifying positive RSV samples.

Identification of eligible controls

Controls were defined as term infants with no significant medical history and who were not hospitalized with a LRTI in the first year of life. Controls were matched at a 1:1 ratio for chronological age at the time of pulmonary function testing (±2 weeks of the case), gender and race. Controls were identified through a birth cohort study of 35,000 mother-newborn dyads (V98_28OB study; HREC: 140203) from 01 April 2016 until 31 December 2016, and Surveillance of Severe Childhood Illness in Soweto, South Africa / Babies of Soweto study (HREC: 140904) 01 January until 31 March 2019. There was a slight discrepancy between the number of cases and controls included in the study, this was due to the difference in the success rate of the pulmonary function testing and the inability to absolutely confirm cases and controls, and the success of the pulmonary function testing.

The V98_28OB study was a case-control study nested within a prospective, longitudinal cohort of mothers and their infants. This study aimed to establish a sero-correlate of protection, based on maternal and newborn serotype-specific levels of group B streptococcal (GBS) anti-capsular antibody, against invasive GBS disease in newborns. The study included a total of six enrolment sites and one delivery site. The infants were evaluated at six and 12 months of age by the study team to determine if hospitalization occurred in the first year of life. A list of well, non-hospitalized controls that met inclusion criteria was generated from the V98_28OB database and matched to cases. The sample function in R version 3.5.1. was used to randomly select 100 babies as a block, who were either 10-14 or 22-26 months of age and would therefore correspond to the ages of the included cases at time of pulmonary function testing. This list was worked through from the first to the last control listed before a further list with more possibilities would be generated. Enrolment for the V98_28OB ended on 31 December 2016 and therefore no further controls were available for inclusion into the study from V98_28OB after 31 December 2018.

After 31 December 2018 identification of community controls happened through Surveillance of Severe Childhood Illness in Soweto, South Africa / Babies of Soweto study. Daily surveillance of all births at CHBAH was performed by the RMPRU and lists generated. The details of the mother was then gathered via the hospital's Medicom system (permission was obtained). Infants previously hospitalized were not eligible for participation.