An investigation in the seasonal patterns of bacterial colonisation in childhood pneumonia

by

Marie Belle Kathrina Mendoza Auckloo

(student no. ACKMAR005)

Supervisor:

Professor Maia Lesosky

A dissertation submitted to the University of Cape Town in partial fulfilment of the requirement for the award of the degree of Master of Public Health (in Epidemiology & Biostatistics)

Division of Epidemiology and Biostatistics

School of Public Health and Family Medicine

Faculty of Health Sciences

University of Cape Town

South Africa

February 2022

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

PREAMBLE

DECLARATION

I, Marie Belle Kathrina Mendoza Auckloo (student no. ACKMAR005), hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 28/07/2022

ACKNOWLEDGEMENTS

To my supervisor, Professor Maia Lesosky, thank you for your support, guidance, and positivity during this research project. I appreciate all your efforts, feedback, and encouragement to get me through this minor dissertation.

To my husband, Siphesiihle Hlela, thank you for supporting and encouraging me to stay focused, motivated, and determined throughout this research project.

ABSTRACT

An investigation in the seasonal patterns of bacterial colonisation in childhood pneumonia

Marie Belle Kathrina M. Auckloo

(student no. ACKMAR005)

Under the supervision of Professor Maia Lesosky

Globally, pneumonia is a leading cause of morbidity and mortality in children younger than the age of 5 years, especially in low- and middle-income countries. The aetiology of paediatric pneumonia is complex, and its definitive determination remains challenging. *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, and *S. aureus* are among the most frequent bacterial causes of childhood pneumonia. Common to such recognised bacterial pathogens is the occurrence of asymptomatic bacterial colonisation or carriage of the nasopharynx, which in turn precedes disease development and progression. This study investigates the trends and patterns of bacterial carriage in the development of pneumonia in South African infants up to one year of life. It is hypothesised that respiratory bacterial carriage is considerably influenced by age and seasonality. Based on previously collected time-series data from the Drakenstein Child Health Study conducted in Paarl, South Africa, this study focuses on exploring the effects of season on nasopharyngeal carriage of pathogens occurring in the nasopharynx of young children with and without the occurrence of lower respiratory tract infection. Using logistic mixed effects models and taking into account the repeated measure structure of the data as well as seasonal components, we found that seasonal variations occur in the prevalence of nasopharyngeal carriage of respiratory pathogens in infants, with and without lower respiratory tract infection. With the inclusion of age-effects, these associations appeared to be highly complex. Understanding the factors that influence bacterial carriage, asymptomatic or not, is necessary to better understand the opportunities for and impact of intervention strategies against lower respiratory tract illness.

TABLE OF CONTENTS

PREAMBLE

DECLARATION i

ACKNOWLEDGEMENTS ii

ABSTRACT iii

ORGANISATION OF THESIS vi

LIST OF ABBREVIATIONS vii

LIST OF TABLES viii

LIST OF FIGURES ix

PART A: PROTOCOL

SYNOPSIS 1

RESEARCH PROTOCOL 3

1. Introduction 3
2. Research background 3
   - 2.1. Childhood pneumonia 3
   - 2.2. Pneumonia aetiology 4
3. Study rationale 5
   - 3.1. Research goal 6
   - 3.2. Research aim 6
   - 3.3. Research objectives 7
4. Research methods 7
   - 4.1. Study design and setting 7
   - 4.2. Characteristics of the study population 8
   - 4.3. Recruitment and enrolment 8
   - 4.4. Research procedures 8
   - 4.5. Data collection methods 8
   - 4.6. Data analysis 9
5. Description of risks and benefits 10
6. Informed consent process 11
7. Privacy and confidentiality 11
8. Research logistics 11
   - 8.1. Budget 11
   - 8.2. Workplan 11
9. Research dissemination 12
10. References 12

PART B: JOURNAL MANUSCRIPT

Title 1
Abstract 2
Introduction 4
Methods 6
Results 9
Discussion 22
Conclusions 25
References 27

PART C: APPENDICES

Appendix A: Journal manuscript supplementary material 1
Appendix B: Study ethics approval letter 12
Appendix C: Author Guidelines for the International Journal of Epidemiology 14

ORGANISATION OF THESIS

The Master of Public Health (MPH) (in Epidemiology & Biostatistics) degree consist of a coursework component followed by a minor dissertation. As per the MPH dissertation guidelines, this thesis is divided into three sections: Part A, B, and C.

Part A

The research protocol outlining the study proposal including research background, purpose of study, aims, objectives, proposed methodology, and ethical considerations.

Part B

A journal manuscript of a research paper formatted for submission to a named peer-reviewed academic journal as per the selected journal's instructions for authors. The manuscript provides a logical flow to the research study from background summary, research methods used, results reporting and evaluation to study conclusions.

Part C

The appendices including ethics approval for the research study, journal-specific author instructions and guidelines for submission, and any supplementary material to the journal manuscript.

LIST OF ABBREVIATIONS

BOBYQA	bound optimisation by quadratic approximation
CI	confidence interval
DCHS	Drakenstein Child Health Study
GBD	Global Burden of Diseases, Injuries, and Risk Factors
Hib	Haemophilus influenzae type b
HIV	human immunodeficiency virus
HREC	Human Research Ethics Committee
IQR	interquartile range
LMIC	low- and middle-income countries
LRTI	lower respiratory tract infection
ML	maximum likelihood
MPH	Master of Public Health
NHLS	National Health Laboratory Services
NP	nasopharyngeal
NTHi	non-typeable Haemophilus influenzae
OR	odds ratio
RBP	recognised bacterial pathogen
SD	standard deviation
SES	socio-economic status
UCT	University of Cape Town

LIST OF TABLES

Part A: Protocol

Table 1 List of variables included in this secondary analysis

Part B: Journal manuscript

Table 1 Characteristics of the infants (n=851) included in this study

Table 2 Prevalence of bacterial carriage in infants (n=851) at the time of collection according to infant age group. Note: An infant can have nasopharyngeal carriage for more than one organism at any given time

Table 3 Different mixed effects logistic regression models fitted for the association of nasopharyngeal carriage (from birth to 1 year of life) with age (in days) and season (summer, autumn, winter, spring)

Table 4 Mixed effects logistic regression model fitted for the association of nasopharyngeal carriage (from birth to 1 year of life) with potential risk factors

Part C: Appendices

Table S1 Prevalence of bacterial respiratory pathogens in nasopharyngeal samples (n=16 860) across infant age groups

LIST OF FIGURES

Part A: Protocol

Figure 1 Proposed work plan for this research study

Part B: Journal manuscript

Figure 1 Prevalence of bacterial carriage in infants (n=851) over time (infant age at collection in days). Note: An infant can have nasopharyngeal carriage for more than one organism at any given time

Figure 2 Prevalence of bacterial carriage in infants (n=851) at the time of collection according to calendar months. Note: An infant can have nasopharyngeal carriage for more than one organism at any given time

Figure 3 Season-related carriage interactions of bacterial pathogens in the nasopharynx of infants

Figure 4 LRTI-related carriage interactions of bacterial pathogens in the nasopharynx of infants.

Figure 5 LRTI-related carriage interactions of bacterial pathogens in the nasopharynx of infants.

Part C: Appendices

Figure S1 Prevalence of bacterial respiratory pathogens in nasopharyngeal samples (n=16 860)

Figure S2 Distribution of positive nasopharyngeal samples (n=16 860) over time (infant age at collection in days) across calendar months for each pathogen

Figure S3 Prevalence of bacterial carriage in infants (n=851) at the time of collection according to seasons. Note: An infant can have nasopharyngeal carriage for more than one organism at any given time

Figure S4 Regression outputs for the computed models with *S. pneumoniae* carriage as the outcome

Figure S5 Regression outputs for the computed models with *H. influenzae* carriage as the outcome

Figure S6 Regression outputs for the computed models with *M. catarrhalis* carriage as the outcome

Figure S7 Regression outputs for the computed models with *S. aureus* carriage as the outcome

PART A: PROTOCOL

An investigation in the seasonal patterns of bacterial colonisation in childhood pneumonia

Marie Belle Kathrina M. Auckloo

(student no. ACKMAR005)

SYNOPSIS

Background

Globally, pneumonia is a leading cause of morbidity and mortality in children younger than the age of 5 years. Although most deaths attributed to pneumonia in children are mostly in low- and middle-income countries (LMICs), the burden of disease is substantial, and there are significant healthcare-associated costs related to pneumonia in the developed world. Pneumonia in South African children is a common disease yet determining its aetiology remains elusive.

Primary study

The parent study uses secondary data from a number of sources to develop statistical models for longitudinal data. This proposed work in this project builds on analyses carried out for the parent research. In particular, this analysis will work on data from the Drakenstein Child Health Study (DCHS) which carried out detailed investigations on the aetiology of lower respiratory tract infection on nasopharyngeal specimens in children. A subset of the original DCHS birth cohort (called the intensive cohort) were followed every two weeks for the first year of life to enable very regular monitoring of respiratory symptoms and bi-monthly collection of nasopharyngeal samples to assess carriage of bacterial organisms and patterns that precede the development of lower respiratory tract infection (LRTI).

Rationale

Factors influencing bacterial colonisation are not fully understood as of yet, however many studies investigated adhesion to mucosal receptors and immune responses, as well as bacterial properties and resistance dynamics. Some bacterial pathogens are more likely to colonise the nasopharynx of children prone to recurrent otitis media, where impaired local immunity and repeated exposure to respiratory organisms are additional factors. The exact dynamic nature of nasopharyngeal colonisation remains unclear to date in part due to inadequate statistical models, but the timing, seasonal pattern, and co-colonisation of bacterial agents can be influenced by external and intrinsic risk factors, medical treatment, or vaccination. This work will apply advanced models developed by the parent study to a specific application.

Aim

The proposed research aims to perform a secondary analysis to apply longitudinal seasonal models to understand the nonlinear associations between risk factors influencing respiratory pathogen colonisation. The objectives are to describe the prevalence and seasonal patterns of bacterial carriage for four selected pathogenic respiratory organisms in the development of childhood pneumonia. To this end, his study attempts to add to the body of literature on the trends and patterns of bacterial colonisation, especially seasonally, in the development of pneumonia in South African children.

Methodology

This is an analysis of secondary data, with no participants recruited or enrolled and no study measures taken. The data being used is already collected, de-identified and prepared for analysis. The original data was collected during planned routine visits (active surveillance) and after hospital discharge, using clinical assessments and standardised questionnaires, all of which were administered by trained research investigators including clinicians and nurses, academic staff, and student researchers. The participants will not be subject to any health interventions and no additional information will be collected for the purpose of this proposed study.

Ethics

The confidentiality of the primary study shall be maintained throughout the duration of the secondary study. The data made available for this study is anonymised and contains no personal identifiers related to the original participants. It will be stored in an access-controlled database, to which only the student researcher and supervisors will have access.

RESEARCH PROTOCOL

1. Introduction

Globally, pneumonia is a leading cause of morbidity and mortality in children younger than the age of 5 years. Although most deaths attributed to pneumonia in children are in low- and middle-income countries (LMICs), the burden of disease globally is substantial, and there are significant healthcare-associated costs related to pneumonia in the developed world. Pneumonia in South African children is a common disease yet determining its aetiology remains elusive. The Drakenstein Child Health Study (DCHS) conducts detailed investigations on the aetiology of lower respiratory tract infection (LRTI) on nasopharyngeal specimens in children. A subset of the original DCHS birth cohort (called the intensive cohort) were followed every two weeks for the first year of life to enable very regular monitoring of respiratory symptoms and the collection of nasopharyngeal samples to assess carriage of organisms and patterns that precede the development of LRTI. The proposed research aims to perform a secondary analysis of the original data collected towards investigating the risk factors influencing respiratory pathogen colonisation and specifically to describe seasonal patterns of bacterial carriage. The objectives are to describe the prevalence and patterns of bacterial carriage for four selected pathogenic respiratory organisms.

2. Research background

2.1. Childhood pneumonia

Lower respiratory tract infections (LRTIs), typically defined as pneumonia or bronchiolitis, are a major cause of morbidity and mortality worldwide. According to the Global Burden of Diseases, Injuries, and Risk Factors (GBD) Study, LRTIs ranked 6th as a leading cause of mortality of all ages accounting for over two million deaths in 2016.[1] Among children aged 5 years or less (under-5s), pneumonia is a major single cause of death outside the neonatal period, causing approximately around 1.4 million deaths every year, accounting for 18% of all deaths of under-5s children globally.[2] Recently, a global decline in the incidence and disease severity of childhood pneumonia has been reported in multiple epidemiological studies due to more effective prevention strategies.[2] As a result, childhood morbidity and mortality due to LRTIs, especially among the under-5s are decreasing. However, pneumonia still remains a significant health burden for young children in LMICs.[3]

LMICs, especially African nations, experience high disease incidence and severity rates of childhood pneumonia.[2,4,5] In Africa, the under-5s paediatric population constitutes only a quarter of all live births globally. [2,6] Yet, for children under five, higher mortality countries are concentrated in African countries. Better living conditions, improved food nutrition and more effective vaccines contributed to the overall steady decline in pneumonia incidence, severe morbidity and mortality in LMICs.[7,8] Further, there is a very high prevalence of risk factors for severe disease in these settings including malnutrition and lack of breastfeeding, biomass or cigarette smoke exposure, HIV (human immunodeficiency virus) infection, poverty and poor living conditions.[9] With increasing availability of effective, improved vaccines for LRTI, especially pneumococcal conjugate vaccine (PCV) and *H. influenzae* type b (Hib) conjugate vaccine, the incidence, severity and causes of LRTI have been changing.[10–13] Case control studies from high income countries have shown that viral pathogens, predominantly respiratory syncytial virus (RSV), contribute an increasing proportion of the aetiology, but that multiple pathogens are common.[14,15]

In South Africa, the epidemiology and aetiology of childhood pneumonia is changing although it remains a major cause of illness and hospitalisation.[16,17] In 2017, reports observed about 300 000 episodes of childhood pneumonia and over 4000 under-5s deaths.[18] Improved diagnostic methods have highlighted the importance of multiple pathogens contributing to co-infection in the aetiology of respiratory illness, and the importance of organism interactions.[16] There are several determinants of pneumonia severity and mortality including infancy (particularly those <4 months), premature birth, incomplete immunisation, maternal smoking or household tobacco smoke exposure, indoor air pollution, low birthweight, malnutrition, lack of exclusive breastfeeding and overcrowding.[9,16,19,20]

2.2. Pneumonia aetiology

A wide spectrum of pathogens, ranging from viral and bacterial to fungal and parasitic organisms, is implicated in pneumonia pathogenesis, and disease may be due to multiple organisms.[16] If children are hospitalised with pneumonia, bacterial-viral co-infection is often the norm.[21,22] In the current era of conjugate vaccines targeting *H. influenzae* type b (Hib) and *S. pneumoniae*, there has been a shift in the spectrum of pathogens causing pneumonia, with viruses contributing to a greater proportion of severe LRTI episodes and non-typeable *H. influenzae* (NTHi) and S. aureus emerging as important bacterial pathogens.[23–29]

In understanding the aetiology of pneumonia in children, two important aspects are considered, the age of the population of interest and the type of pathogen under study.[30] Excluding neonates, all paediatric age groups are commonly affected by viral pathogens, with *S. pneumoniae* being the most commonly identified bacterial organism.[31] Neonates are at risk for bacterial pathogens present in the birth canal, such as *E. coli*, *Klebsiella*, *L. monocytogenes*, and *group B streptococci*.[32,33] However, *S. pneumoniae*, *S. pyogenes*, and *S. aureus* have previously been identified in late-onset neonatal pneumonia.[34] In infants and toddlers (between 30 days and 2 years old) as well as young children (between 2-5 years of age), viruses are the primary aetiological agent in causing pneumonia.[35,36] Older children (between 5-13 years old) commonly experience mycoplasma pneumonia with S. pneumoniae as the most commonly present bacterial pathogen.[36] Adolescents usually have the same infectious risks as adults.[37]

The pathogenesis of bacterial pneumonia begins with the microaspiration of recognised bacterial pathogens (RBPs) such as *S. pneumoniae*, *H. influenzae*, and *S. aureus* into the lower airways after they have colonised the nasopharynx, a common upper airway portal for microorganisms' entry into the respiratory tract.[38–40] In the absence of good clearance mechanisms and effective innate or acquired immune responses, such aspiration may be followed by the development of pneumonia. In children, the nasopharyngeal flora becomes established during the first year of life.[41] Studies have shown that nasopharynx colonisation by RBPs is established early in childhood, although rates vary greatly according to locality, sampling frequency, as well as individual and social factors.

3. Study rationale

The aetiology of paediatric pneumonia is complex, and its definitive determination remains challenging. In LMICs, there is a high prevalence of risk factors (host and environmental) for severe disease and high incidence of pneumonia in children, however, compared to the developed world, comprehensive aetiological research is limited.[42] *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, and *S. aureus* are among the most frequent bacterial causes of childhood pneumonia. Common to such RBPs and other invading organisms of the respiratory tract is the occurrence of asymptomatic bacterial colonisation, or carriage of the nasopharynx, which in turn precede disease development and progression.[43]

Factors influencing bacterial colonisation are not fully understood as of yet, however many studies investigated adhesion to mucosal receptors and immune responses, as well as bacterial properties and resistance dynamics.[41] Some bacterial pathogens are more likely to colonise the nasopharynx of children prone to recurrent otitis media, where impaired local immunity and repeated exposure to respiratory organisms are additional factors.[41] The exact dynamic nature of nasopharyngeal colonisation remains unclear to date, but the timing, pattern, and co-colonisation of bacterial agents can be influenced by external and intrinsic risk factors, medical treatment, or vaccination.

The nasopharyngeal colonisation process typically begins shortly after birth. The asymptomatic carriage of bacterial pathogens presents an opportunity for disease development, with the nasopharynx serving as a reservoir and source for pathogenic transmission.[44] Various demographic and clinical characteristics such as overcrowding, poverty, smoking, infancy, childcare, family size, and antibiotics use have been associated with bacterial carriage.[41,45,46] Nasopharyngeal specimens detect the presence of a pathogen in the upper respiratory tract. Although the presence of organisms in the nasopharynx does not necessarily lead to infection and hence cause pneumonia, the identification of the pathogen in the upper respiratory tract can provide circumstantial evidence on the dysbiosis of the nasopharyngeal flora which may predispose to pneumonia. Furthermore, nasopharyngeal isolates do reflect the epidemiological changes caused by the multiple bacterial pathogens present within the community.[47,48]

3.1. Research goal

This study attempts to add to the body of literature on the trends and patterns of bacterial colonisation in the development of pneumonia in South African children.

3.2. Research aim

The aim of this study is to investigate the trends, patterns, and risk factors of colonisation of selected pathogens occurring in the nasopharynx of young children (up to 1 year of life) with and without the occurrence of LRTI.

3.3. Research objectives

The research objectives of this study are:

1. To identify, describe, and quantify patterns (frequency, seasonality, age effects) of colonization of selected key bacterial pathogens colonising the nasopharynx in young children up to one year of age.
2. To compare the patterns of bacterial colonisation with respect to LRTI occurrence in the cohort of infants.
3. To evaluate potential risk factors associated with bacterial carriage in nasopharyngeal swabs specimens collected from the intensive cohort.

4. Research methods

The parent study (BiostatsLAB Africa HREC:485/2019) (Appendix A) uses secondary data from a number of sources to develop statistical models for longitudinal data. This proposed work in this project builds on analyses carried out for the parent research. In particular, this secondary analysis will work on data from the DCHS which carried out detailed investigations on the aetiology of lower respiratory tract infection on nasopharyngeal specimens in children. The following sections describe the procedures to be undertaken for the proposed secondary data analysis.

4.1. Study design and setting

The parent study to this proposed secondary analysis study, is a multi-year, population-based birth cohort study located in the peri-urban Drakenstein district in Paarl, outside of the Cape Town Metropole, South Africa.[49] It investigates early-life health determinants of LRTIs in South African children. From 2012, the DCHS study enrolled over 1200 mother-child pairs at two public primary healthcare clinics, TC Newman (predominantly serving Black African individuals) and Mbekweni (mainly serving individuals of mixed ancestry).[49] To date, a substantial amount of data on the enrolled infants has been collected during the DCHS research activities and the follow-up study visits are currently ongoing. A subset of the original birth cohort (called the intensive cohort) was followed every two weeks for the first year of life to thoroughly evaluate aetiological factors preceding LRTI development in infants.[50] A secondary longitudinal analysis will be conducted using previously collected data on infants from the intensive cohort.

4.2. Characteristics of the study population

As per the parent DCHS study's research protocol, and according to the Western Cape Government Department of Social Development, the population of Drakenstein local municipality was over 270,000 individuals in 2018.[51] This low socio-economic populace typically live in crowded conditions and informal settlements. Multiple social and health determinants such as poverty and unemployment greatly contribute to the high burden of infectious diseases, including pneumonia, tuberculosis, and HIV within this community.[49]

4.3. Recruitment and enrolment

As per the parent DCHS study's protocol, pregnant women aged 18 years or older were enrolled antenatally in the at 20-28 weeks gestation (2nd trimester).[49] Convenience sampling was carried out at both clinics (Mbekweni and TC Newman) during a 3-year period (03/2012 - 03/2015). Eligibility included attendance at one of the two health clinics for antenatal care and residence in the district for 1 year or more. Women enrolled in the study were followed through pregnancy and childbirth (all occurred at Paarl hospital). Postnatally, infants and their mothers were followed up until the children turned 5 years old, with some neonates enrolled into the intensive early life cohort for in-depth sampling.[50] Children receive primary healthcare and immunisation at clinics.

4.4. Research procedures

As per the parent DCHS study's protocol, antenatal and postnatal visits were at primary healthcare clinics, while birth, 6-week and annual study visits occur at Paarl hospital. Fathers, as identified by mothers, were invited to participate in an antenatal study visit. Infants attended study visits synchronised with the national programme where feasible at 6, 10 and 14 weeks, and 6, 9, 18, 30, 36, 42, 54 and 60 months. Two home visits (antenatally and 4 months postnatally) were done to investigate environmental risk factors.

4.5. Data collection methods

As per the parent DCHS study's protocol, comprehensive data including biomedical, environmental, psychosocial, demographic, and physical risk factors are longitudinally measured for the mother, father, and child. Environmental exposures (e.g., air composition) are measured using monitors placed at home visits. Lung function is measured in children at 6 weeks, annually and during LRTI episodes and 4-6 weeks thereafter. Chronic respiratory

disease measurements include clinical symptoms, patient records, lung function, chest X-ray and ultrasound (during an LRTI event). Neurodevelopment is assessed longitudinally using brain MRI. Microbiological investigations including microbiome and multiplex PCR measures are done longitudinally and at LRTI episodes. All children have six monthly nasopharyngeal swabs (NPs) and stool specimens collected, while the subset intensive cohort have two weekly NPs and monthly stool samples in the first year.

All the data generated during the DCHS research study was collected during planned routine visits (active surveillance) and after hospital discharge, using clinical assessments and standardised questionnaires, all of which were administered by trained research investigators including clinicians and nurses, academic staff, and student researchers.

The proposed study is a secondary investigation in the data previously produced by the DCHS research. The participants will not be subject to any health interventions and no additional information will be collected for the purpose of this proposed study.

4.6. Data analysis

For this proposed secondary data analysis, the following factors will be examined as possible risk factors of NP carriage prevalence and patterns in children: age, time of birth, gender, ethnicity, socioeconomic status (SES), season, passive smoking, number of siblings, LRTI occurrence, antibiotic use, and immunisation (Table 1).

Table 1 List of variables included in this secondary analysis

Construct	Variables
Biomedical	▪ LRTI event/episode/occurrence ▪ Pathogen/organism type
Psychosocial	• Family socio-economic status
Demographic	• Infant gender/sex • Infant ethnicity • Household structure (overcrowding) • Household smokers • Infant age at nasopharyngeal collections • Maternal antenatal alcohol use • Parental illness/comorbidities
Physical	• Birth characteristics (prematurity, birthweight, feeding type) • Vaccinations/immunisation • Clinic attended

Note: LRTI, lower respiratory tract infection

The raw data will be collated using a Microsoft® Excel spreadsheet. Data analysis will be conducted in the latest version of R statistical software®. Descriptive statistics (univariate analysis) will be used to summarise the characteristics of children with and without pneumonia. Continuous data will be expressed as mean (standard deviation [SD]) if normally distributed or median (interquartile range [IQR]) if not normally distributed. Categorical data will be summarised using counts, frequencies, percentages, and proportions, as appropriate. The prevalence of colonisation for all four pathogenic respiratory bacteria will be described.

Statistical chi-square test and Fisher's exact probability test (bivariate analysis) will be applied for comparing categorical variables between two groups and odds ratio (OR) and 95% confidence interval (CI) will be reported, where appropriate. The Mann-Whitney U test (Wilcoxon rank sum test), a non-parametric method, will be carried out when comparing continuous (numerical) variables without normal distribution between two groups. For comparisons among more than three groups, p-values will be adjusted by Bonferroni correction for multiple comparisons. Kruskal-Wallis test will be used to investigate associations between categorical and continuous variables.

Statistical modelling (multivariate analysis) will be performed to identify risk factors associated with bacterial colonisation (NP carriage). Log-binomial regression will be used to compare the overall prevalence of colonisation between groups and adjusted for relevant characteristics found to differ significantly ($p<0.05$) between groups. Multivariate logistic regression will be conducted to model significant variable effects on the occurrence of pneumonia adjusted for confounding covariates. Binomial models for repeated measures will be used to estimate the time varying association (seasonality) of colonisation with LRTI, adjusted for confounders.

5. Description of risks and benefits

There are no direct risks and benefits for the researchers and participants as this proposed study is a retrospective secondary analysis of previously collected data. There are potential indirect benefits for the participants in the future based on the public health related information produced by this study. A breach in confidentiality and privacy during the study presents a potential indirect risk for the participants.

6. Informed consent process

As per the parent DCHS research's protocol, expecting mothers provided written informed consent at enrolment and provided consent again after the first year. They were reconsented annually following childbirth. Ethics approval for the DCHS study was obtained from the University of Cape Town Faculty of Health Sciences Research ethics Committee (HREC ref: 401/2009), and the Provincial Research committee. The parent statistical study, BiostatsLAB Africa HREC:485/2019, which is the source of this protocol's secondary data, was also granted ethical approval from the UCT HREC office.

7. Privacy and confidentiality

The confidentiality of the primary study shall be maintained throughout the duration of the secondary study. The data made available for this study is anonymised and contains no personal identifiers related to the original participants. It will be stored in an access-controlled database, to which only the student researcher and supervisors will have access.

8. Research logistics

8.1. Budget

This research study is a secondary data analysis, therefore, there will be no anticipated direct or indirect costs associated with the preparation, development, and submission of this work.

8.2. Workplan

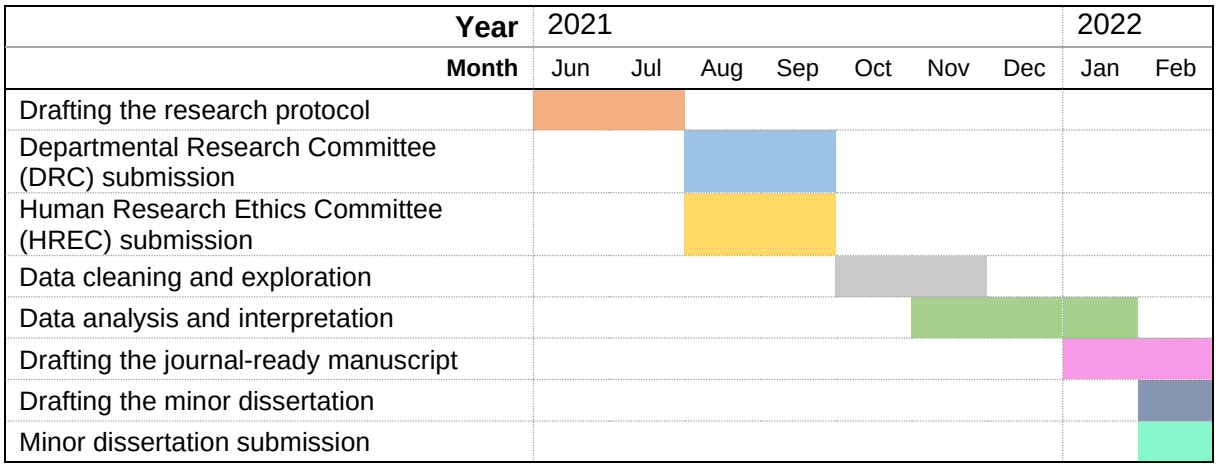


Figure 1 Proposed work plan for this research study

9. Research dissemination

The research findings may be important and useful to other researchers, governmental and non-governmental entities involved in understanding and characterising childhood pneumonia in South Africa. If deemed suitable for scientific publication, the study results will be published in a peer-reviewed journal.

10. References

1. Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017;390(10100):1151–210.
2. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet 2015;385(9966):430–40.
3. Walker CLF, Rudan I, Liu L, Nair H, Theodoratou E, Bhutta ZA, et al. Global burden of childhood pneumonia and diarrhoea. Lancet 2013;381(9875):1405–16.
4. Zar HJ, Barnett W, Myer L, Nicol MP. Childhood pneumonia – the Drakenstein Child Health Study. South African Med J 2016;106(7):642.
5. Chopra M, Mason E, Borrazzo J, Campbell H, Rudan I, Liu L, et al. Ending of preventable deaths from pneumonia and diarrhoea: an achievable goal. Lancet 2013;381(9876):1499–506.
6. Zar HJ, Madhi SA, Aston SJ, Gordon SB. Pneumonia in low- and middle-income countries: progress and challenges. Thorax 2013;68(11):1052–6.
7. Marangu D, Zar HJ. Childhood pneumonia in low-and-middle-income countries: An update. Paediatr Respir Rev 2019;32:3–9.
8. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Swartz S, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Infect Dis 2017;17(11):1133–61. 1
9. Jackson S, Mathews KH, Pulanić D, Falconer R, Rudan I, Campbell H, et al. Risk factors for severe acute lower respiratory infections in children – a systematic review and meta-analysis. Croat Med J 2013;54(2):110–21.
10. Zar HJ, Polack FP. Childhood pneumonia: the role of viruses. Thorax 2015;70(9):811–2.
11. Simonsen L, Taylor RJ, Schuck-Paim C, Lustig R, Haber M, Klugman KP. Effect of 13-valent pneumococcal conjugate vaccine on admissions to hospital 2 years after its introduction in the USA: a time series analysis. Lancet Respir Med 2014;2(5):387–94.
12. Angoulvant F, Levy C, Grimprel E, Varon E, Lorrot M, Biscardi S, et al. Early Impact of 13-Valent Pneumococcal Conjugate Vaccine on Community-Acquired Pneumonia in Children. Clin Infect Dis 2014;58(7):918–24.

13. Greenberg D, Givon-Lavi N, Ben-Shimol S, Ziv JB, Dagan R. Impact of PCV7/PCV13 introduction on community-acquired alveolar pneumonia in children <5 years. Vaccine 2015;33(36):4623–9.

14. Jain S, Williams DJ, Arnold SR, Ampofo K, Bramley AM, Reed C, et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Children. N Engl J Med 2015;372(9):835–45.

15. Rhedin S, Lindstrand A, Hjelmgren A, Ryd-Rinder M, Öhrmalm L, Tolfvenstam T, et al. Respiratory viruses associated with community-acquired pneumonia in children: matched case–control study. Thorax 2015;70(9):847–53.

16. Moore DP, Green RJ, Cohen C, Itzikowitz G, Andronikou S, Argent AC, et al. Epidemiology and aetiology of community-acquired pneumonia in children: South African Thoracic Society guidelines (part 1). South African Med J 2020;110(7):583–7.

17. le Roux DM, Nicol MP, Myer L, Vanker A, Stadler JAM, von Delft E, et al. Lower Respiratory Tract Infections in Children in a Well-vaccinated South African Birth Cohort: Spectrum of Disease and Risk Factors. Clin Infect Dis 2019;69(9):1588–96.

18. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1736–88.

19. Cohen C, Walaza S, Moyes J, Groome M, Tempia S, Pretorius M, et al. Epidemiology of Viral-associated Acute Lower Respiratory Tract Infection Among Children <5 Years of Age in a High HIV Prevalence Setting, South Africa, 2009–2012. Pediatr Infect Dis J 2015;34(1):66–72.

20. Sonego M, Pellegrin MC, Becker G, Lazzerini M. Risk Factors for Mortality from Acute Lower Respiratory Infections (ALRI) in Children under Five Years of Age in Low and Middle-Income Countries: A Systematic Review and Meta-Analysis of Observational Studies. PLoS One 2015;10(1):e0116380.

21. Michelow IC, Olsen K, Lozano J, Rollins NK, Duffy LB, Ziegler T, et al. Epidemiology and Clinical Characteristics of Community-Acquired Pneumonia in Hospitalized Children. Pediatrics 2004;113(4):701–7.

22. Blyth CC, Webb SAR, Kok J, Dwyer DE, van Hal SJ, Foo H, et al. The impact of bacterial and viral co-infection in severe influenza. Influenza Other Respi Viruses 2013;7(2):168–76.

23. Scott JAG, English M. What Are the Implications for Childhood Pneumonia of Successfully Introducing Hib and Pneumococcal Vaccines in Developing Countries? PLoS Med 2008;5(4):e86.

24. Cohen C, Walaza S, Moyes J, Groome M, Tempia S, Pretorius M, et al. Epidemiology of Severe Acute Respiratory Illness (SARI) among Adults and Children Aged ≥5 Years in a High HIV-Prevalence Setting, 2009–2012. PLoS One 2015;10(2):e0117716.

25. Jain S, Finelli L, CDC EPIC Study Team. Community-acquired pneumonia among U.S. children. N Engl J Med 2015;372(22):2167–8.

26. Zar HJ, Barnett W, Stadler A, Gardner-Lubbe S, Myer L, Nicol MP. Aetiology of childhood pneumonia in a well vaccinated South African birth cohort: a nested case-control study of the Drakenstein Child Health Study. Lancet Respir Med 2016;4(6):463–72.

27. Bénet T, Sánchez Picot V, Messaoudi M, Chou M, Eap T, Wang J, et al. Microorganisms Associated With Pneumonia in Children <5 Years of Age in Developing and Emerging Countries: The GABRIEL Pneumonia Multicenter, Prospective, Case-Control Study. Clin Infect Dis 2017.

28. le Roux DM, Zar HJ. Community-acquired pneumonia in children — a changing spectrum of disease. Pediatr Radiol 2017;47(11):1392–8.

29. O'Brien KL, Baggett HC, Brooks WA, Feikin DR, Hammitt LL, Higdon MM, et al. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. Lancet 2019;394(10200):757–79.

30. Rudan I, O'Brien KL, Nair H, Liu L, Theodoratou E, Qazi S, et al. Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries. J Glob Health 2013;3(1):10401.

31. Popovsky EY, Florin TA. Community-Acquired Pneumonia in Childhood. In: Reference Module in Biomedical Sciences. Elsevier; 2020.

32. Chen JC, Jenkins-Marsh S, Flenady V, Ireland S, May M, Grimwood K, et al. Early-onset group B streptococcal disease in a risk factor-based prevention setting: A 15-year population-based study. Aust New Zeal J Obstet Gynaecol 2019;59(3):422–9.

33. Al Hazzani AA, Bawazeer RAB, Shehata AI. Epidemiological characterization of serotype group B Streptococci neonatal infections associated with interleukin-6 level as a sensitive parameter for the early diagnosis. Saudi J Biol Sci 2018;25(7):1356–64.

34. Arif F. Updated Recommendations Of Rcog On Prevention Of Early Onset Neonatal Group B Streptococcus Infection. J Ayub Med Coll Abbottabad 2018;30(3):490.

35. Verhoeven D. Influence of Immunological Maturity on Respiratory Syncytial Virus-Induced Morbidity in Young Children. Viral Immunol 2019;32(2):76–83.

36. Troeger C, Blacker B, Khalil IA, Rao PC, Cao J, Zimsen SRM, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Infect Dis 2018;18(11):1191–210.

37. Ebeledike C, Ahmad T. Pediatric Pneumonia [Updated 2020 Nov 21]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021.

38. Mandell LA, Niederman MS. Aspiration Pneumonia. N Engl J Med 2019;380(7):651–63.

39. Rudan I. Epidemiology and etiology of childhood pneumonia. Bull World Health Organ 2008;86(5):408–16.

40. Cevey-Macherel M, Galetto-Lacour A, Gervaix A, Siegrist C-A, Bille J, Bescher-Ninet B, et al. Etiology of community-acquired pneumonia in hospitalized children based on WHO clinical guidelines. Eur J Pediatr 2009;168(12):1429–36.

41. Garcia-Rodriguez JA. Dynamics of nasopharyngeal colonization by potential respiratory pathogens. J Antimicrob Chemother 2002;50(90003):59–74.

42. Zar HJ, Ferkol TW. The global burden of respiratory disease-Impact on child health. Pediatr Pulmonol 2014;49(5):430–4.

43. Siegel SJ, Weiser JN. Mechanisms of Bacterial Colonization of the Respiratory Tract. Annu Rev Microbiol 2015;69(1):425–44.

44. Case for Carriage. Vaccine 2013;32(1):180–203.

45. Bogaert D, van Belkum A, Sluijter M, Luijendijk A, de Groot R, Rümke H, et al. Colonisation by Streptococcus pneumoniae and Staphylococcus aureus in healthy children. Lancet 2004;363(9424):1871–2.

46. Jounio U, Juvonen R, Bloigu A, Silvennoinen-Kassinen S, Kaijalainen T, Kauma H, et al. Pneumococcal carriage is more common in asthmatic than in non-asthmatic young men. Clin Respir J 2010;4(4):222–9.

47. Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG. Clonal Relationships between Invasive and Carriage Streptococcus pneumoniae and Serotype- and Clone-Specific Differences in Invasive Disease Potential. J Infect Dis 2003;187(9):1424–32.

48. Mackenzie GA, Leach AJ, Carapetis JR, Fisher J, Morris PS. Epidemiology of nasopharyngeal carriage of respiratory bacterial pathogens in children and adults: cross-sectional surveys in a population with high rates of pneumococcal disease. BMC Infect Dis 2010;10(1):304.

49. Zar HJ, Barnett W, Myer L, Stein DJ, Nicol MP. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. Thorax 2015;70(6):592 LP – 594.

50. Zar HJ. Bacterial and viral pneumonia: New insights from the Drakenstein Child Health Study. Paediatr Respir Rev 2017;24:8–10.

51. Socio-Economic Profile (SEP): Drakenstein Municipality. 2017.

PART B: JOURNAL MANUSCRIPT

An investigation in the seasonal patterns of bacterial colonisation in childhood pneumonia

Marie Belle Kathrina M. Auckloo

(student no. ACKMAR005)

Target Journal

International Journal of Epidemiology

Title

An investigation in seasonal patterns of bacterial colonisation in childhood pneumonia

Author

Marie Belle Kathrina Mendoza Auckloo[1]

Affiliation

[1] Division of Epidemiology and Biostatistics, School of Public Health and Family Medicine, Faculty of Health Sciences, University of Cape Town, South Africa

Corresponding author

Corresponding author: Marie Belle Kathrina Mendoza Auckloo, Anzio Road, Observatory, Cape Town, South Africa. E-mail: kathrina.auckloo@uct.ac.za

Word count:

Abstract: 250 (max: 250 words)

Main text: 3912 (max: 3000 words)

[1] This journal-ready manuscript is formatted for the *International Journal of Epidemiology* journal as per the requirements set out in the author guidelines (Part C, Appendix B). According to the Master of Public Health (MPH) dissertation guidelines, and for the purpose of this mini-dissertation, the student is the sole and first author, and the contributions of co-authors are recognised in the acknowledgements section of the dissertation.

Abstract

Background: In low- and middle-income countries, pneumonia is an important health burden in children, especially during their first year of life when they are at greatest risk for developing severe upper and lower respiratory tract infections. *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Staphylococcus aureus* are common infectious respiratory pathogens involved in the pathogenesis of childhood pneumonia. Given that the nasopharyngeal carriage of respiratory pathogens precedes bacterial colonisation and disease development, nasopharyngeal samples are often used to investigate colonisation patterns.

Methods: The study aimed to assess the seasonal variations in the prevalence of bacterial carriage in infants (up to 1 year old) using nasopharyngeal samples collected and analysed during the parent research, the Drakenstein Child Health Study conducted in Paarl, South Africa. A secondary statistical data analysis was carried out on the longitudinal time-series data obtained. Mixed effects logistic regression models were fitted to estimate the effect of seasonality on bacterial carriage, while adjusting for age effects.

Results: All four respiratory pathogens were prevalent in infants (25% – 55%). Generally, the prevalence of nasopharyngeal carriage increased with increasing infant age, except for S. aureus whose carriage decreased over time. Carriage was greatly influenced by age, season, and the experience of lower respiratory tract infections; however, these associations and interactions appeared highly complex. Another potential risk factor for carriage was the size of the infant's household.

Conclusions: The marked differences in the seasonality of respiratory bacterial carriage must be considered when investigating carriage as a proxy for bacterial colonisation towards clinical disease.

Keywords:

bacterial carriage, respiratory pathogen, seasonality, colonisation, lower respiratory tract infection, nasopharyngeal carriage

Key Messages:

- Respiratory pathogens *S. pneumonia*, *H. influenzae*, *M. catarrhalis*, and *S. aureus* are commonly carried in the nasopharynx of young infants.
- The pattern of those respiratory bacterial carriage in infants was significantly seasonal with marked differences between pathogens.
- Carriage of *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, and *S. aureus* varied across age in infants up to one year of life.
- An association was noted between nasopharyngeal carriage of respiratory bacteria and lower respiratory tract infections.
- Complex interactions exist between age, season and lower respiratory tract infection which effect bacterial carriage in infants.

Introduction

Acute respiratory tract infections are classified as upper respiratory tract infections or lower respiratory tract infections (LRTIs). LRTIs in children is a spectrum of common illness that may affect peripheral airways (e.g., bronchiolitis) or the alveoli (e.g., pneumonia), but these can occur sequentially or simultaneously. Childhood LRTIs account for over 700 000 deaths in children under five each year.[1] Due to differences in specific aetiologies and risk factors, the severity of those infections is worse in developing countries, resulting in higher case-fatality rate. [2]

Pneumonia can be caused by many infectious agents, including viruses, bacteria, or fungi. Bacterial pneumonia is most often caused by *Streptococcus pneumoniae* or *Haemophilus influenzae*, mostly type b (*Hib*), and occasionally by *Staphylococcus aureus*. [2] In low- and middle-income countries (LMICs), pneumonia remains a considerable health burden for young children, especially those under 6 months of age, who remain at greatest risk for developing severe LRTI. [3,4] The nasopharynx is the primary hub for recognised bacterial pathogens (RBPs) such as *S. pneumoniae*, *H. influenzae*, and *Moraxella catarrhalis*, [5–7] and a secondary one for *S. aureus*. [8]

The nasopharyngeal carriage of these bacterial respiratory pathogens is usually harmless and asymptomatic. However, under specific conditions like poor clearance mechanisms and ineffective immune responses, they can enter the lungs and cause severe illness. Importantly, the nasopharynx is also a reservoir where respiratory pathogens are transmitted to other individuals, therefore asymptomatic nasopharyngeal carriage is related to disease development and the spread of the pathogen. [5,8–10] The prevalence of nasopharyngeal carriage

of respiratory pathogens varies according to age, geographical area, crowding, concomitant respiratory disease, and antibiotic use. [11]

In young children, there are several determinants of bacterial pneumonia severity and mortality including infancy (particularly those under 4 months of age), premature birth, incomplete immunisation, maternal smoking or household tobacco smoke exposure, indoor air pollution, low birthweight, malnutrition, lack of exclusive breastfeeding and overcrowding. [12–14] A large body of literature exist on the variables related to the dynamics of nasopharyngeal carriage but the role of seasonal variation on the prevalence and patterns of nasopharyngeal carriage remains unclear.

This study investigated the influence of season on nasopharyngeal carriage of respiratory pathogens in a cohort of South African infants. To further the understanding on seasonal trends of bacterial colonisation in the development of pneumonia in children, we aimed to determine the impact of seasonality on bacterial carriage of selected pathogens occurring in the nasopharynx of young children (up to 1 year of life) with and without the occurrence of LRTI in Cape Town, South Africa – a subtropical country with four seasons, summer, fall, winter, and spring. To this end, this research work is a secondary analysis of longitudinal research data obtained from the Drakenstein Child Health Study (DCHS), an ongoing South African population-based birth cohort study. [15]

Methods

Study design and participants

The Drakenstein Child Heath Study, undertaken in Paarl, South Africa, is a multi-year population-based birth cohort study investigating early-life health determinants of LRTIs in children. [15] From 2012, the study enrolled over 1200 mother-child pairs at two public primary healthcare clinics, TC Newman and Mbekweni. These two study sites are about 5 km apart and serve a combined population of roughly 200 000 individuals. Based on socioeconomic and demographic characteristics, the DCHS study population is considered to be representative of many South Africa communities and other LMICs. [16] Pregnant women (20-28 weeks' gestation) were enrolled and followed up through pregnancy and childbirth. Enrolment criteria was broad to ensure generalisability and that the cohort would be representative of the general population. All births occurred at Paarl hospital and the children were followed up through early childhood. [17] The DCHS study obtained ethics approval from the University of Cape Town, Faculty of Health Sciences, Human Research Ethics Committee (HREC ref. no 401/2009 and 651/2013), and the Provincial Child Health Research committee. provided ethical approval. All enrolled mothers provided written informed consent with a renewal on an annual basis.

Study procedures

The DCHS research work collected a substantial amount of information on the enrolled mother-child dyads, and follow-up visits are still ongoing. A subset of the original birth cohort (called the intensive cohort) was followed every two weeks for the first year of life to thoroughly evaluate aetiological factors preceding LRTI development in infants.[18] All children have six monthly nasopharyngeal swabs (FLOQSwabs, Copan Diagnostics, Murrieta, CA, USA) collected, while the subset intensive cohort have two weekly nasopharyngeal swabs

samples in the first year. At each LRTI event, a nasopharyngeal swab was also obtained. The collected nasopharyngeal swabs were immediately placed into 1 mL skim milk-tryptone-glucose-glycerol (STGG), transported at 4°C to the National Health Laboratory Services (NHLS) in Cape Town within 2 hours of collection and frozen at -80°C for later batch culture of bacteria. Presumptive pneumococcal isolates were identified by colony morphology, α-haemolysis, optochin disk susceptibility (Oxoid, Basingstoke, UK) and confirmed using *lyt*A PCR. Serotyping was performed on a single morphologically distinct pneumococcal colony per sample by sequetyping and confirmation using the Quellung method. Longitudinal measurement of active surveillance risk factors (environment, nutrition, maternal and paternal variables, and vaccination) was collected during planned routine visits and at case presentation using clinical assessments and standardised questionnaires.

Statistical analysis

This current study is a secondary longitudinal analysis of previously obtained DCHS data on infants from the intensive cohort, with data collected up to 1 year of life. Infant age in days was measured and categorised as 0-3, 3-6, and 6-12 months old. Nasopharyngeal carriage was defined as positive or negative (i.e., dichotomous), as per the laboratory detection results. Carriage in infants was defined as any positive nasopharyngeal sample collected at a given time. An infant can have positive nasopharyngeal carriage for more than one respiratory bacterial pathogen at any given time. Time at collection of nasopharyngeal samples was defined in calendar months and seasons, which were categorised as summer (December to February), autumn (March to May), winter (June to August), and spring (September to November). LRTI was defined as positive if an infant ever experienced an episode from birth to 1 year of age. Summary statistics (median and IQR for quantitative variables and count percentage, n (%) for categorical variables) were estimated and described.

Mixed effects logistic regression models, estimated using ML and BOBYQA optimizer, including a random intercept for each infant, were fitted to quantify the association of nasopharyngeal carriage with seasonality, according to LRTI status, reporting odd ratios (OR) and their 95% confidence intervals (95% CI) while adjusting for age, a strong confounder. Age variable (in days) was scaled (z-scores) to improve the model fit for complex seasonal and interaction models. Seasonality was fitted using a truncated Fourier series. To test whether nasopharyngeal carriage is seasonally distributed, sine (point in time) and cosine (amplitude) functions were introduced into the regression model to capture the complexity of seasonality/cyclicality. To assess the seasonal effect over the year, the date of sample collection was used to create a *day of the* year variable, which in turn was used to create new dummy seasonal variables $sinday = \sin(\frac{2\pi day}{365})$, $cosday = \cos(\frac{2\pi day}{365})$, $sinday2 = \sin(\frac{4\pi day}{365})$, and $cosday2 = \cos(\frac{4\pi day}{365})$ (annual and biannual). These seasonal components were enters as independent variables in the logistic mixed models.

It is hypothesised that the seasonal pattern under study follows a cosine function with variable amplitude and horizontal shift. Standardised parameters were obtained by fitting the models on a standardized version of the dataset. 95% CIs and p-values were computed using the Wald approximation. Data were analysed using the R statistical programming software for Windows, version 3.6.3 (R core team, 2022).

Results

Characteristics of study population

Infants recruited into the intensive cohort at enrolment of the primary study undertook two-weekly follow-up with nasopharyngeal sampling throughout their first year of life. Altogether, 851 infants were included in this ancillary analysis, providing a total of 16 860 nasopharyngeal samples with a median of 23 samples per infant (IQR: 14 – 26). The infants were born between June 2012 and September 2015, and the collection period for nasopharyngeal samples occurred between June 2012 and August 2016. The cohort demographic, physical, biomedical, and psychosocial characteristics are shown in Table 1.

Table 1 Characteristics of the infants (*n*=851) included in this study

Study characteristics	Overall (*n*=851)
Sex	
male	425 (49.9%)
female	426 (50.1%)
Ethnicity	
black african	591 (69.4%)
mixed ancestry	255 (30.0%)
missing	5 (0.6%)
Clinic site	
Mbekweni	603 (70.9%)
TC Newman	248 (29.1%)
Preterm	
early term	133 (15.6%)
full term	248 (29.1%)
missing	1 (0.1%)
Length at birth (z-score)	
mean (sd)	-0.0858 (1.56)
median (range)	0.00273 (-7.09, 4.63)
missing	22 (2.6%)
Weight at birth (z-score)	
mean (sd)	-0.554 (1.18)
median (range)	-0.510 (-5.33, 6.33)
HIV exposure	
no	617 (72.5%)
yes	232 (27.3%)
missing	2 (0.2%)
BCG vaccination	
no	6 (0.7%)
yes	743 (87.3%)
missing	102 (12.0%)
Household size	
mean (sd)	4.78 (2.60)
median (range)	4 (1, 17)
missing	2 (0.2%)
Household adult size	
mean (sd)	2.99 (1.66%)
median (range)	2 (1, 13)

Exclusive breastfeeding	
no	371 (43.6%)
yes	444 (52.2%)
missing	36 (4.2%)
Smoking exposure at birth	
active smoking	212 (24.9%)
non smoking	217 (25.5%)
passive smoking	389 (45.7%)
missing	33 (3.9%)
House type	
flat	93 (10.9%)
house	419 (49.2%)
shack	251 (29.5%)
wendy house	78 (9.2%)
other	3 (0.4%)
missing	7 (0.8%)
Socio-economic status (in quartiles)	
high ses	184 (21.6%)
moderate-high ses	207 (24.3%)
moderate ses	0 (0.0%)
low-moderate ses	230 (27.0%)
low ses	229 (26.9%)
missing	1 (0.1%)
Prenatal alcohol exposure	
no	97 (11.4%)
yes	701 (82.4%)
missing	1 (0.1%)
Maternal HIV status	
negative	618 (72.6%)
positive	232 (27.3%)
missing	1 (0.1%)
Maternal history of asthma	
no	824 (96.8%)
yes	15 (1.8%)
missing	12 (1.4%)

Note: sd, standard deviation; HIV, human immunodeficiency virus

Prevalence of selected bacterial respiratory pathogens in nasopharyngeal swabs

Of the 16 860 nasopharyngeal samples, 54.5%, 34.3%, 40.2%, and 25.0% were positive for *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, and *S. aureus*, respectively (supplementary Figure S1). S. pneumoniae was most prevalent (n=9 193, 54.5%) while S. aureus was the least commonly detected (n=4 212, 25.0%). At the time of collection, 36.7%, 26.7%, and 36.6% nasopharyngeal samples were collected in infants aged 0-3, 3-6, and 6-12 months, respectively (supplementary Table S1). *S. aureus* was predominantly detected as positive (14.3%) among the 0-3 months olds while *S. pneumoniae* was most prevalent among both 3-6 months (17.4%) and 6-12 months (24.3%) old infants. Overall, the prevalence of nasopharyngeal carriage increases over time,

with the exception of *S. aureus* whose prevalence decreased as the infants grow older (supplementary Figure S2).

Nasopharyngeal carriage in infants with lower respiratory tract infection

Figure 1 shows the prevalence of nasopharyngeal carriage of each bacterial respiratory pathogen over time (infant age at collection in days). The prevalence of *S. aureus* carriage was highest immediately after birth at above 50%and decreased rapidly as the infants grow older. On the other hand, the prevalence of *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis* was low at birth, increased rapidly over time, and reached a plateau around 2-3 months of age until the end of the follow-up at 1 year of age. The prevalence of carriage in young infants is highest for *S. aureus*, followed by *S. pneumoniae*, *M. catarrhalis*, and *H. influenzae*. This trend changes over time with *S. pneumoniae* carriage being most common, followed by *M. catarrhalis*, *H. influenzae*, and *S. aureus* in older infants.

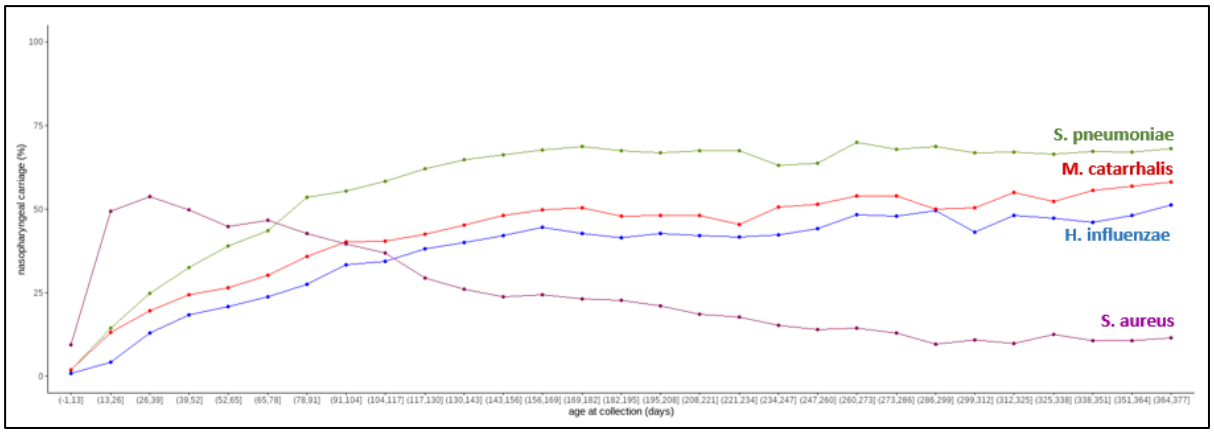


Figure 1 Prevalence of bacterial carriage in infants (*n*=851) over time (infant age at collection in days). Note: An infant can have nasopharyngeal carriage for more than one organism at any given time.

Overall, nasopharyngeal carriage of all four bacterial respiratory pathogen was higher among infants who never had an LRTI episode (Table 2). In infants who ever (one or more episodes) experienced LRTI, the prevalence of bacterial carriage was highest for *S. pneumoniae* (65.7%), followed by *H. influenzae* (60.4%), *M. catarrhalis* (54.3%), and *S. aureus* (36.2%). This trend

differed slightly in infants who never experienced an LRTI episode with *S. pneumoniae* (92.7%) being most prevalent, followed by *M. catarrhalis* (91.3%), *H. influenzae* (84.4%), and *S. aureus* (82.8%). The pattern of bacterial carriage in infants with and without LRTI shows some differences with increasing age groups (Table 2). However, irrespective of LRTI experience, S. pneumoniae remained highly prevalent across all age groups.

Table 2 Prevalence of bacterial carriage in infants (*n*=851) at the time of collection according to infant age group. Note: An infant can have nasopharyngeal carriage for more than one organism at any given time.

All infants					
age (mths)	**number of infants with bacterial carriage, *n* (%)**				**total number of infants, *n* (%)**
	S. pneumoniae **795 (93.4)**	***H. influenzae*** **730 (85.8)**	***M. catarrhalis*** **777 (91.3)**	***S. aureus*** **709 (83.3)**	
0-3	631 (74.1)	456 (53.6)	606 (71.2)	613 (72.0)	848 (99.6)
3-6	699 (82.1)	580 (68.2)	667 (78.4)	424 (49.8)	774 (91.0)
6-12	653 (76.7)	572 (67.2)	631 (74.1)	342 (40.2)	698 (82.0)
Infants with LRTI					
age (mths)	**number of infants with bacterial carriage, *n* (%)**				**total number of infants, *n* (%)**
	S. pneumoniae **174 (65.7)**	***H. influenzae*** **160 (60.4)**	***M. catarrhalis*** **144 (54.3)**	***S. aureus*** **96 (36.2)**	
0-3	61 (7.2)	57 (6.7)	50 (5.9)	56 (6.6)	127 (14.9)
3-6	73 (8.6)	70 (8.2)	54 (6.3)	27 (3.2)	107 (12.6)
6-12	87 (10.2)	72 (8.5)	67 (7.9)	18 (2.1)	128 (15.0)
Infants without LRTI					
age (mths)	**number of infants with bacterial carriage, *n* (%)**				**total number of infants, *n* (%)**
	S. pneumoniae **788 (92.7)**	***H. influenzae*** **717 (84.4)**	***M. catarrhalis*** **776 (91.3)**	***S. aureus*** **704 (82.8)**	
0-3	625 (73.4)	436 (51.2)	603 (70.9)	609 (71.6)	848 (99.6)
3-6	686 (80.6)	564 (66.3)	660 (77.6)	419 (49.2)	769 (90.4)
6-12	635 (74.6)	553 (65.0)	620 (72.9)	336 (39.5)	677 (79.6)

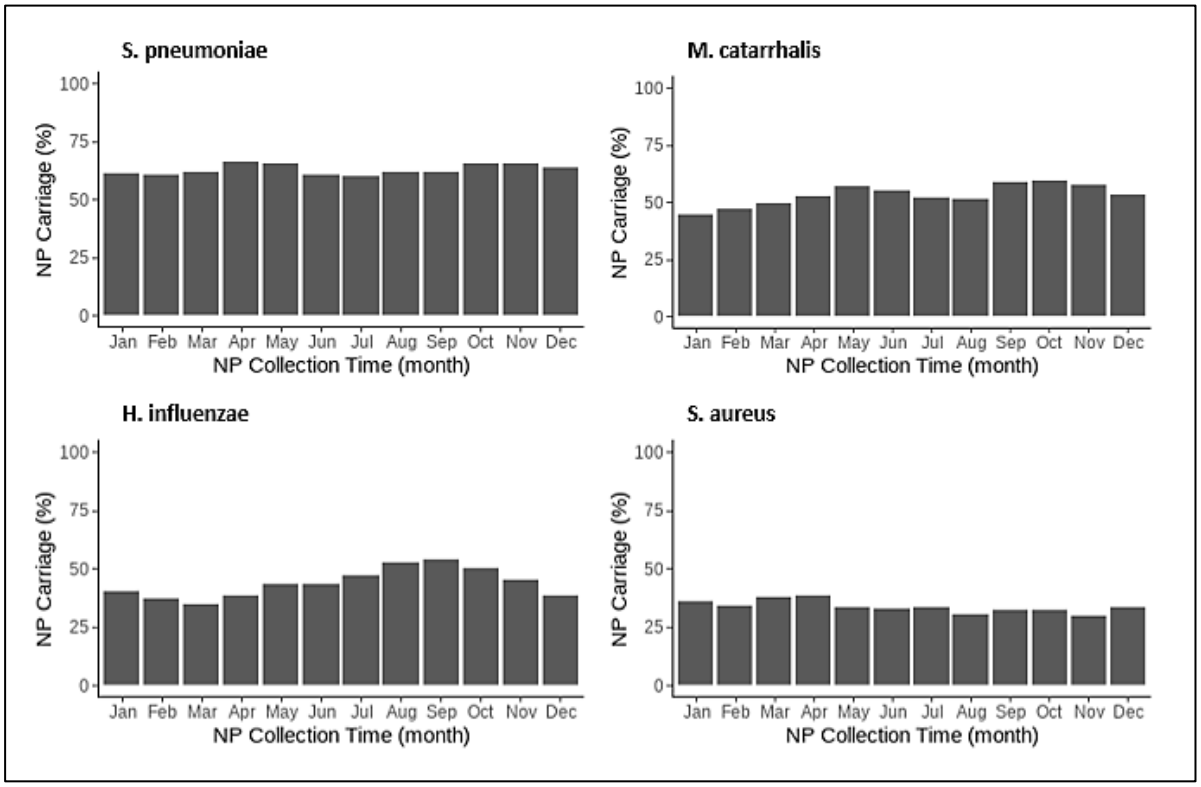
Note: mths, months

Seasonal patterns of respiratory bacterial carriage in infants

Nasopharyngeal swabs were collected and analysed each month between 2012 – 2016 from the infants enrolled in the intensive cohort, starting from birth until their first birthday. The prevalence of bacterial carriage in the infants varied between the study months (Figure 2). The

proportion of *S. pneumoniae* carriers remained fairly consistent through the months, being highest in the months of autumn and spring (supplementary Figure S3). There was a marked seasonal pattern of *H. influenzae* carriage across the months, being most prevalent in September (early spring). For *M. catarrhalis*, infant carriers were most and least common in spring and summer, respectively. Carriage of *S. aureus* did not show considerable differences across the study months, and overall, its prevalence was lower than the other respiratory pathogens.

Figure 2 Prevalence of bacterial carriage in infants (*n*=851) at the time of collection according to calendar months. Note: An infant can have nasopharyngeal carriage for more than one organism at any given time.



Effect of season and age on the dynamics of bacterial carriage

This study analysed longitudinal data using appropriate statistical modelling tools to determine whether the nasopharyngeal carriage of bacterial respiratory pathogens exhibit prominent differences in seasonality. To this end the study months were partitioned into seasons to infer on the season-specific patterns of bacterial carriage. In all the models, bacterial

carriage (for *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, and *S. aureus*) was the outcome as a binary variable. The repeated-measure structure of the data was accounted for by including a random effect term for the infant participant.

The effect of seasonality on the dynamics of bacterial carriage are depicted in table 3 and supplementary figures S4(a), S5(a), S6(a), and S7(a). For each pathogen, model A estimated bacterial carriage with infant age. There is strong evidence that nasopharyngeal carriage of all four organisms is significantly associated with infant age. Therefore, the probability of carrying the respiratory bacteria increases as the infant grows older. Model B fitted bacterial carriage against infant age and season as predictors. The effect of age is statistically significant and positive across organisms. *S. pneumoniae* carriage was significantly more prevalent in winter (Figure S4(a)). Carriage of both *H. influenzae* and *S. aureus* in infant nasopharynx were significantly more common in winter and spring (Figure S5(a), Figure S7(a)). For *M. catarrhalis*, carriage in infants was least probable in summer (Figure S6(a)). Model C estimated the seasonal effects of bacterial carriage over the study year with age and seasonal components as predictors. In *S. pneumonia* and *M. catarrhalis* carriage, all of the seasonal variables were statistically significant and positive, indicating a strong seasonal effect (semesterly and quarterly) (Figure S4(a), Figure S6(a)). In *H. influenzae* and *S. aureus* carriage, most and half of the Fourier terms were statistically significant, respectively (Figure S5(a), Figure S7(a)). This supports the hypothesis that season significantly influence respiratory bacterial carriage in infants and that the patterns are different for each organism.

Interactions influencing bacterial carriage in infants

Model D estimated the effect of an age-season interaction on bacterial carriage (Table 3). Infant age was statistically significant and positive across all organisms. The interaction effect of

season on age in *S. pneumoniae* and *M. catarrhalis* carriage were significant in spring and winter, respectively (Figure S4(b), Figure S6(b)). Seasonality has a strong effect on *H. influenzae* carriage depending on how old the infant is in their first year of life (Figure S5(b)). The interaction effect between age and season was not significant for *S. aureus* carriage (Figure S7(b)). Figure 3 illustrates the interactions between age and season effects on bacterial carriage. Young infants are most likely to carry *S. pneumoniae* in spring, and as they grow older, this likelihood shifts to autumn months and a higher probability. For both *H. influenzae* and *M. catarrhalis*, the odds of nasopharyngeal carriage are highest in spring and winter, immediately after birth and just before their first birthday, respectively. There is a general negative relationship between *S. aureus* carriage and infant age with carriage being most prevalent in autumn in the early months of life, and this decreases with age.

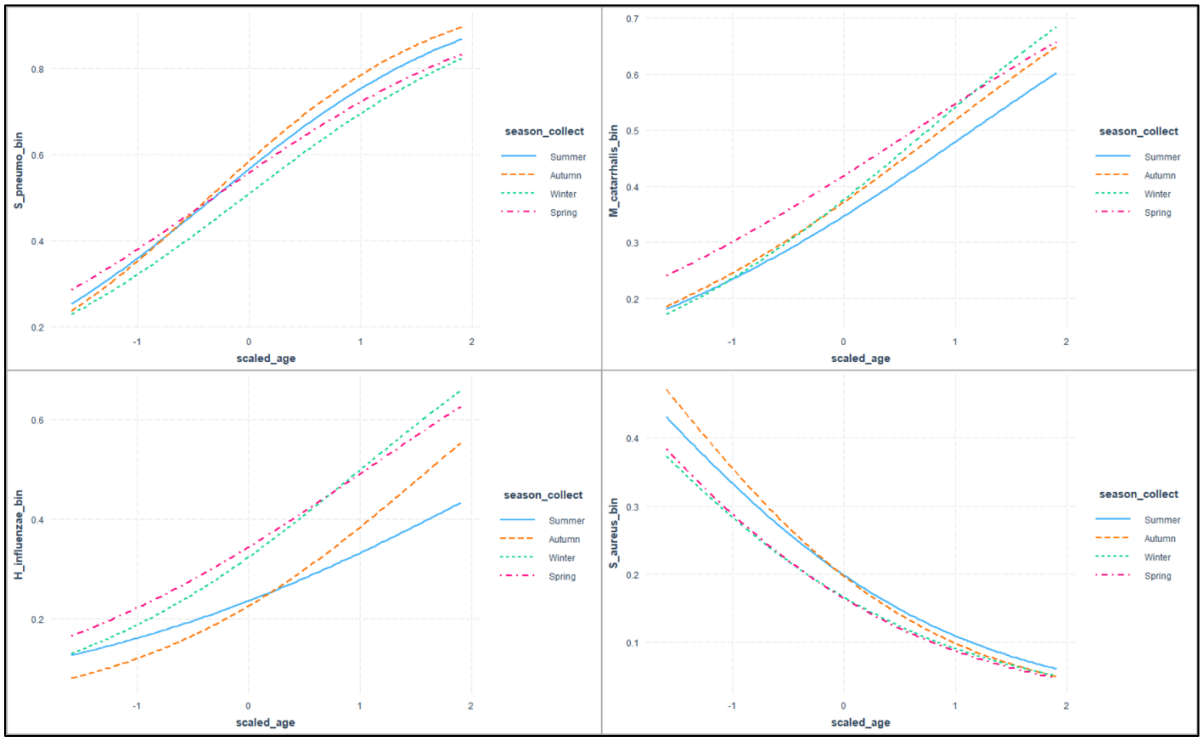


Figure 3 Season-related carriage interactions of bacterial pathogens in the nasopharynx of infants.

Table 3 Different mixed effects logistic regression models fitted for the association of nasopharyngeal carriage (from birth to 1 year of life) with age (in days) and season (summer, autumn, winter, spring).

Outcome:	***S. pneumoniae*** **OR (95% CI)**	***P*-value**	***H. influenzae*** **OR (95% CI)**	***P*-value**	***M. catarrhalis*** **OR (95% CI)**	***P*-value**	***S. aureus*** **OR (95% CI)**	***P*-value**
Model A: *carriage ~ age*								
age (z-score)	2.26 (2.17 – 2.35)	**<0.001**	1.90 (1.83 – 1.97)	**<0.001**	1.79 (1.73 – 1.85)	**<0.001**	0.48 (0.46 – 0.50)	**<0.001**
Model B: *carriage ~ age + season*								
age (z-score)	2.27 (2.18 – 2.36)	**<0.001**	1.91 (1.84 – 1.99)	**<0.001**	1.79 (1.72 – 1.86)	**<0.001**	0.48 (0.46 – 0.51)	**<0.001**
season: summer (ref)	-	-	-	-	-	-	-	-
season: autumn	1.07 (0.97 – 1.19)	0.191	0.99 (0.89 – 1.11)	0.849	1.12 (1.01 – 1.24)	**0.025**	1.01 (0.90 – 1.14)	0.833
season: winter	0.80 (0.72 – 0.89)	**<0.001**	1.62 (1.45 – 1.80)	**<0.001**	1.15 (1.05 – 1.27)	**0.004**	0.80 (0.71 – 0.90)	**<0.001**
season: spring	0.97 (0.88 – 1.08)	0.590	1.75 (1.58 – 1.95)	**<0.001**	1.36 (1.24 – 1.50)	**<0.001**	0.80 (0.71 – 0.90)	**<0.001**
Model C: *seasonal model*								
age (z-score)	2.27 (2.18 – 2.36)	**<0.001**	1.92 (1.85 – 2.00)	**<0.001**	1.79 (0.57 – 0.64)	**<0.001**	0.48 (0.46 – 0.51)	**<0.001**
sinday	1.08 (1.02 – 1.13)	**0.005**	0.68 (0.65 – 0.72)	**<0.001**	0.89 (0.84 – 0.93)	**<0.001**	1.15 (1.09 – 1.22)	**<0.001**
cosday	1.12 (1.06 – 1.18)	**<0.001**	0.82 (0.77 – 0.86)	**<0.001**	0.92 (0.88 – 0.97)	**0.002**	1.09 (1.02 – 1.15)	**0.006**
sinday_v2	0.90 (0.85 – 0.95)	**<0.001**	1.08 (1.03 – 1.14)	**0.003**	0.89 (0.85 – 0.94)	**<0.001**	1.04 (0.99 – 1.10)	0.126
cosday_v2	0.94 (0.89 – 0.99)	**0.012**	0.96 (0.91 – 1.01)	0.110	0.93 (0.89 – 0.98)	**0.005**	0.98 (0.93 – 1.04)	0.478
Model D: *age * season interaction model*								
age (z-score)	2.33 (2.14 – 2.54)	**<0.001**	1.61 (1.48 – 1.75)	**<0.001**	1.73 (1.60 – 1.87)	**<0.001**	0.50 (0.45 – 0.55)	**<0.001**
season: summer (ref)	-	-	-	-	-	-	-	-
season: autumn	1.08 (0.97 – 1.20)	0.172	0.94 (0.84 – 1.06)	0.315	1.12 (1.01 – 1.23)	**0.032**	0.99 (0.88 – 1.12)	0.888
season: winter	0.79 (0.72 – 0.88)	**<0.001**	1.55 (1.39 – 1.73)	**<0.001**	1.14 (1.03 – 1.26)	**0.011**	0.80 (0.71 – 0.91)	**<0.001**
season: spring	0.96 (0.87 – 1.07)	0.494	1.70 (1.53 – 1.89)	**<0.001**	1.36 (1.23 – 1.50)	**<0.001**	0.80 (0.71 – 0.89)	**<0.001**
age (z-score) * season: autumn	1.11 (0.98 – 1.25)	0.099	1.33 (1.18 – 1.50)	**<0.001**	1.05 (0.94 – 1.17)	0.386	0.90 (0.79 – 1.02)	0.108
age (z-score) * season: winter	0.94 (0.84 – 1.06)	0.314	1.29 (1.15 – 1.45)	**<0.001**	1.13 (1.01 – 1.26)	**0.029**	1.01 (0.89 – 1.16)	0.828

age (z-score) * season: spring	0.88 (0.79 – 0.99)	**0.032**	1.14 (1.02 – 1.28)	**0.021**	0.97 (0.87 – 1.07)	0.518	0.98 (0.86 – 1.12)	0.758
Model E: *age * LRTI interaction model*								
age (z-score)	2.29 (2.20 – 2.38)	**<0.001**	1.91 (0.36 – 0.42)	**<0.001**	1.81 (1.74 – 1.87)	**<0.001**	0.48 (0.46 – 0.51)	**<0.001**
no LRTI (ref)	-	-	-	-	-	-	-	-
LRTI	1.23 (0.98 – 1.54)	0.072	2.88 (2.31 – 3.61)	**<0.001**	1.19 (0.96 – 1.47)	0.107	1.05 (0.81 – 1.37)	0.703
age (z-score) * LRTI	0.63 (0.51 – 0.79)	**<0.001**	0.69 (0.56 – 0.86)	**0.001**	0.73 (0.59 – 0.90)	**0.003**	0.83 (0.62 – 1.12)	0.218
Model F: *age * season * LRTI interaction model*								
age (z-score)	2.37 (2.17 – 2.58)	**<0.001**	1.63 (1.50 – 1.78)	**<0.001**	1.75 (1.62 – 1.89)	**<0.001**	0.50 (0.45 – 0.55)	**<0.001**
season: summer (ref)	-	-	-	-	-	-	-	-
season: autumn	1.07 (0.96 – 1.20)	0.202	0.93 (0.83 – 1.04)	0.200	1.13 (1.02 – 1.25)	**0.021**	1.01 (0.89 – 1.13)	0.932
season: winter	0.78 (0.70 – 0.87)	**<0.001**	1.52 (1.36 – 1.70)	**<0.001**	1.14 (1.03 – 1.26)	**0.009**	0.81 (0.71 – 0.91)	**<0.001**
season: spring	0.95 (0.85 – 1.05)	0.323	1.69 (1.51 -1.88)	**<0.001**	1.36 (1.24 – 1.50)	**<0.001**	0.80 (0.71 – 0.91)	**<0.001**
no LRTI (ref)	-	-	-	-	-	-	-	-
LRTI	0.89 (0.52 – 1.53)	0.676	2.49 (1.43 – 4.32)	**0.001**	1.65 (0.99 – 2.74)	0.053	1.52 (0.84 – 2.76)	0.166
age (z-score) * season: autumn	1.11 (0.98 – 1.26)	0.095	1.31 (1.16 – 1.48)	**<0.001**	1.05 (0.94 – 1.17)	0.394	0.91 (0.80 – 1.04)	0.181
age (z-score) * season: winter	0.93 (0.83 – 1.05)	0.256	1.29 (1.14 – 1.46)	**<0.001**	1.12 (1.01 – 1.25)	**0.037**	1.02 (0.89 – 1.16)	0.807
age (z-score) * season: spring	0.88 (0.78 – 0.99)	**0.033**	1.13 (1.01 – 1.27)	**0.039**	0.97 (0.87 – 1.07)	0.508	0.98 (0.86 – 1.12)	0.788
age (z-score) * LRTI	0.56 (0.32 – 0.98)	**0.041**	0.46 (0.26 – 0.82)	**0.008**	0.57 (0.33 – 0.96)	**0.036**	0.90 (0.47 – 1.71)	0.744
season: autumn * LRTI	1.14 (0.57 – 2.28)	0.714	1.18 (0.59 – 2.39)	0.637	0.59 (0.31 – 1.13)	0.113	0.56 (0.25 – 1.25)	0.157
season: winter * LRTI	1.65 (0.83 – 3.27)	0.152	1.50 (0.75 – 2.99)	0.254	0.69 (0.36 – 1.32)	0.260	0.84 (0.39 – 1.80)	0.658
season: spring * LRTI	1.66 (0.84 – 3.28)	0.143	1.05 (0.53 – 2.08)	0.896	0.72 (0.38 – 1.36)	0.311	0.54 (0.25 – 1.20)	0.133
age (z-score) * season: autumn * LRTI	1.12 (0.56 – 2.24)	0.756	1.86 (0.91 – 3.79)	0.088	1.36 (0.71 – 2.61)	0.354	0.58 (0.24 – 1.43)	0.237
age (z-score) * season: winter * LRTI	1.37 (0.68 – 2.77)	0.380	1.13 (0.55 – 2.31)	0.734	1.42 (0.73 – 2.75)	0.302	1.09 (0.48 – 2.48)	0.844
age (z-score) * season: spring * LRTI	1.02 (0.50 – 2.07)	0.965	2.19 (1.04 – 4.63)	**0.039**	1.26 (0.65 – 2.47)	0.496	1.06 (0.42 – 2.66)	0.906

Note: OR, odds ratio; CI, confidence interval; LRTI, lower respiratory tract infection

Model E estimated the interaction effect of age and LRTI on bacterial carriage (Table 3). Infant age was statistically significant and positive across all organisms. There was a marked LRTI effect on bacterial carriage for *H. influenzae* only (Table 3). However, significant age-LRTI interaction effects were observed for *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis* carriage (Figures S4(b) – S6(b)). The odds of carriage for both *S. pneumonia* and *M. catarrhalis* were significantly higher in young infants who experienced LRTI (Figure 4). With increasing age, the odds of carriage in those without LRTI were significantly higher. *H. influenzae* carriage is significantly more probable in early infants with LRTI compared to those without LRTI (Figure 4). With increasing age, the same trend was observed, however the difference in the odds of carriage decreases between those with and without LRTI. No signifant age-LRTI interaction effects were observed for *S. aureus* carriage.

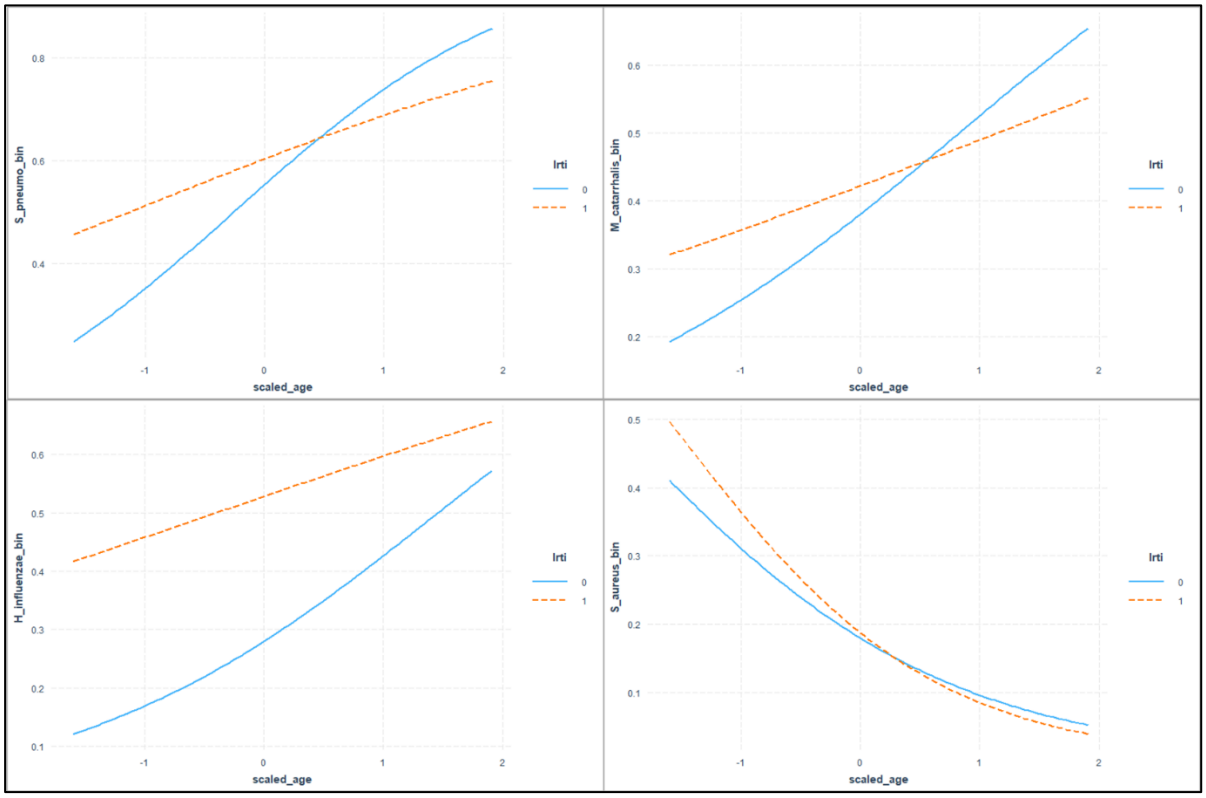


Figure 4 LRTI-related carriage interactions of bacterial pathogens in the nasopharynx of infants.

Model F fitted a three-way interaction effect between age, season, and LRTI on bacterial carriage (Table 3). Infant age was statistically significant and positive across all organisms. There was no marked effect of a season-LRTI interaction in carriage for all bacterial pathogens (Figures S4(b) – S7(b)). A significant age-season-LRTI interaction effect was observed for *H. influenzae* only, where bacterial carriage over time (i.e., with increasing age) is significantly influenced by season when the infants had at an LRTI episode (Figure 5).

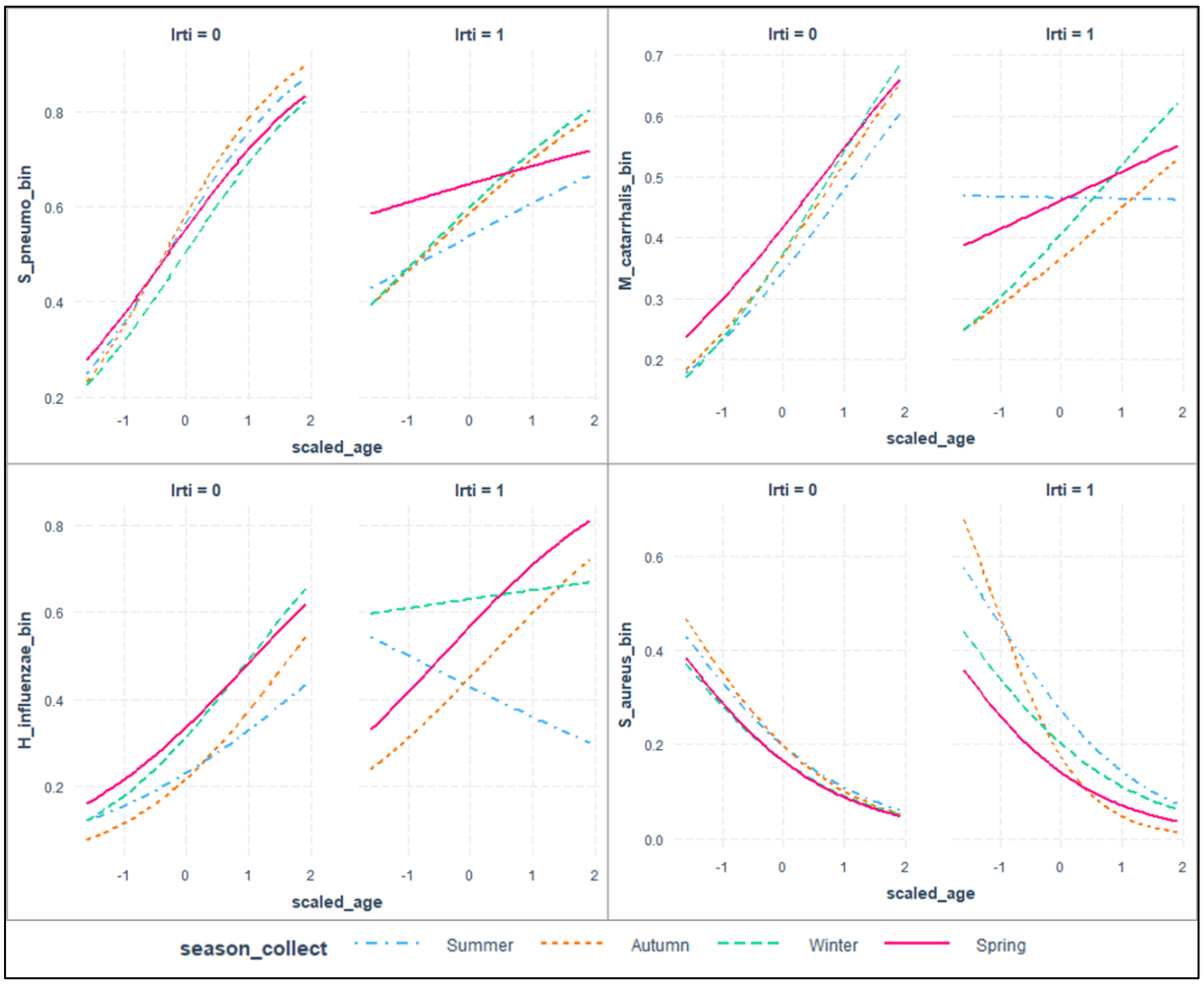


Figure 5 LRTI-related carriage interactions of bacterial pathogens in the nasopharynx of infants.

Risk factors associated with respiratory bacterial carriage in infants

A risk factor model was fitted to potential identify risk factors for bacterial carriage other than age, season and LRTI (Table 4). A marked effect was observed for the household size in all four organisms. The number of adults in the infant's household also appeared to be related to the odds of carriage for all organisms except *M. catarrhalis*. Interestingly, for *S. aureus*, the infant ethnic background seemed to be associated with carriage; this was not observed for the other pathogens.

Table 4 Mixed effects logistic regression model fitted for the association of nasopharyngeal carriage (from birth to 1 year of life) with potential risk factors.

Outcome:	*S. pneumoniae* OR (95% CI)	*P*-value	*H. influenzae* OR (95% CI)	*P*-value	*M. catarrhalis* OR (95% CI)	*P*-value	*S. aureus* OR (95% CI)	*P*-value
sex: female (ref)	-	-	-	-	-	-	-	-
sex: male	0.94 (0.79 – 1.12)	0.497	0.95 (0.79 – 1.13)	0.550	0.85 (0.75 – 0.97)	**0.015**	1.06 (0.87 – 1.28)	0.580
ethnicity: black african (ref)	-	-	-	-	-	-	-	-
ethnicity: mixed ancestry	1.01 (0.81 – 1.27)	0.898	0.90 (0.71 – 1.14)	0.380	0.95 (0.81 – 1.13)	0.573	0.75 (0.58 – 0.97)	**0.026**
preterm: early term (ref)	-	-	-	-	-	-	-	-
preterm: full term	1.06 (0.81 – 1.38)	0.695	1.10 (0.83 – 1.45)	0.526	0.92 (0.75 – 1.12)	0.398	0.93 (0.69 – 1.25)	0.619
household size	1.13 (1.07 – 1.19)	**<0.001**	1.14 (1.08 – 1.21)	**<0.001**	1.06 (1.02 – 1.10)	**0.002**	090 (0.85 – 0.96)	**0.001**
household adult size	0.86 (0.79 – 0.93)	**<0.001**	0.89 (0.81 – 0.96)	**0.005**	0.95 (0.90 – 1.01)	0.102	1.10 (1.01 – 1.21)	**0.032**
HIV exposure: no (ref)	-	-	-	-	-	-	-	-
HIV exposure: yes	0.90 (0.73 – 1.11)	0.327	0.99 (0.79 – 1.23)	0.896	0.89 (0.76 – 1.04)	0.149	0.85 (0.67 – 1.08)	0.185
length at birth (z-score)	1.02 (0.95 – 1.09)	0.666	1.08 (1.00 – 1.16)	0.053	1.04 (0.99 – 1.09)	0.159	0.99 (0.91 – 1.06)	0.724
weight at birth (z-score)	0.97 (0.89 – 1.07)	0.581	0.97 (0.88 – 1.07)	0.503	0.94 (0.87 – 1.00)	0.062	0.95 (0.85 – 1.05)	0.312
exclusive breastfeeding: no	-	-	-	-	-	-	-	-
exclusive breastfeeding: yes	1.06 (0.89 – 1.26)	0.538	1.05 (0.87 – 1.27)	0.586	1.07 (0.94 – 1.22)	0.312	0.88 (0.72 – 1.07)	0.214
smoking exposure at birth: active (ref)	-	-	-	-	-	-	-	-
smoking exposure at birth: passive	0.96 (0.76 – 1.20)	0.696	0.97 (0.76 – 1.22)	0.781	1.00 (0.85 – 1.18)	0.987	1.20 (0.93 – 1.55)	0.154
smoking exposure at birth: none	0.84 (0.64 – 1.09)	0.187	0.76 (0.58 – 1.01)	0.055	0.96 (0.79 – 1.16)	0.653	1.30 (0.97 – 1.74)	0.084
BCG vaccination: no (ref)	-	-	-	-	-	-	-	-
BCG vaccination: yes	2.38 (0.67 – 8.51)	0.181	1.07 (0.28 – 4.12)	0.918	0.63 (0.24 – 1.62)	0.336	1.50 (0.36 – 6.25)	0.576

Note: OR, odds ratio; CI, confidence interval; HIV, human immunodeficiency virus; BCG, bacille Calmette-Guerin

Discussion

This study assessed the seasonal effects of bacterial (*S. pneumoniae*, *H. influenzae*, *M.catarrhalis*, and *S. aureus*) nasopharyngeal carriage in infants over their first year of life. The longitudinal time-series data was collected as part of the Drakenstein Child Health Study in South Africa, between 2012 and 2016. The prevalence of bacterial carriage was evaluated in a cohort of 851 infants who were followed for one year after birth, with nasopharyngeal samples collected every two weeks. Our main findings were that the prevalence and patterns of bacterial carriage differed remarkably according to age and season. There is an association between bacterial carriage and LRTI, but the variations observed are highly complex.

We report that the nasopharyngeal carriage of respiratory bacteria in the nasopharynx reservoir of infants is prevalent in South Africa. A high prevalence of nasopharyngeal carriage with *S. pneumoniae* (54.5%), *H. influenzae* (34.3%), *M. catarrhalis* (40.2%), and *S. aureus* (25.0%) were observed in a cohort of South African infants. This is consistent with previous studies describing respiratory bacterial carriage in young children. For instance, a systematic review in 2014 reported the prevalence of pneumococcal carriage in children less than 5 years of age at 63.2%. [19] In addition, the early acquisition of such recognised bacterial pathogens in the nasopharynx of neonates and infants have previously been described in other African, Asian, and European communities. [20–23]

The probability of bacterial carriage was dependent on age over the follow up period for all four infectious agents. Carriers of *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis* increased with increasing age and reached a constant in older infants, while *S. aureus* showed a different age-carriage trend with a dramatic rapid increase reaching the highest prevalence among neonates, followed by a gradual decrease before reaching a plateau. The literature has previously

reported on the rapid and highly dynamic colonisation of the nasopharynx by such organisms supporting the notable influence of age-effects on carriage, as well as co-carriage in infants. [21,24–27]

This study present strong evidence to suggest that nasopharyngeal carriage of all four respiratory pathogens exhibit different seasonality among infants over time. *S. pneumoniae* and *S. aureus* carriage peaked in autumn months while *H. influenzae* and *M. catarrhalis* carriage were highest in spring. Some studies, such as Marchisio *et al*. 2001, reported similar seasonal fluctuations in the nasopharyngeal carriage of such organisms in healthy Italian young children. [11] Other studies, such as Bojang *et al.* 2015, have found different seasonal patterns of bacterial carriage, with *S. pneumoniae* carriage being more common during the dry season (June to October), while *S. aureus* carriage is not associated with season in Gambian infants. [21,28] Differences in seasonal patterns of nasopharyngeal carriage of respiratory bacteria may be due to various factors not limited to the study population size, the geographical study location, the study design (e.g., longitudinal follow-up versus cross-sectional), and the susceptibility of study participants to respiratory tract illnesses. For instance, viral upper respiratory tract infections, which are susceptible to seasonal effects, may predispose to bacterial pneumonia by facilitating growth of pathogens in the nasopharynx. However, high concentrations of RBPs in the nasopharynx does not necessarily precede the development of LRTI.

We showed that nasopharyngeal carriage of respiratory bacteria is associated with lower respiratory tract infections. This relationship was significant in *H. influenzae* carriage only, indicating that having lower respiratory tract infections increases the odds of carrying the bacteria in the nasopharyngeal hub. Interestingly, when simultaneously accounting for age and LRTI effects, there is a strong evident association with *S. pneumonia*, *H. influenzae*, and *M.*

catarrhalis carriage probabilities. This finding suggests that age is an influential factor in the relationship between bacterial carriage and the occurrence of lower respiratory tract infections. When simultaneously accounting for season and LRTI effects, no evidence association was observed between with nasopharyngeal carriage. However, when investigating age, season, and LRTI altogether, a strong interaction effect was observed with *H. influenzae* carriage. It is important to note that many other underlying factors that may play a role in the association between nasopharyngeal carriage of respiratory bacterial and the incidence of lower respiratory tract infections must be considered for interpretations, for instance, the clinical health characteristics of the infants are crucial determinants of bacterial colonisation. [29,30]

Literature suggests that many individual, socio-economic, and environmental factors contribute to the risk of nasopharyngeal carriage and hence bacterial colonisation, which is a precursor to respiratory disease. [31,32] In this study, in addition to age, season, and LRTI, we identified the size of the infant household, and the number of adults present in the household as strong risk factors associated with nasopharyngeal carriage of bacteria in infants up to 1 year of age. These findings may have important implications on clinical practice and immunisation strategies.

A major strength of this study is the availability of individual time-series repeated data which allowed for longitudinal regression modelling. However, the present study had some limitations indicating that cautious interpretations should be carried out on the results findings. Firstly, the interpretation of seasonality is limited to 1 year of follow-up after birth. Secondly, carriage was measured as binary outcome based on collected and analysed nasopharyngeal swabs only. The combination of nasopharyngeal and oropharyngeal swabbing may provide a slightly different picture of the main findings. Thirdly, the study did

not investigate co-carriage and other dynamic characteristics of bacterial colonisation such re-acquisition which may provide further insights in seasonal patterns due to age-effects in the presence or absence of LRTI. Lastly, although nasopharyngeal carriage of respiratory bacteria can be harmless and asymptomatic, in this study, it is used a proxy for illness development and progression.

Despite these limitations, the study suggests that investigating changes in nasopharyngeal carriage of respiratory bacterial pathogens, using longitudinal data analyses, may provide a better understanding of their health determinants in the context of respiratory illness. Our results underline the importance of considering potential external factors such as seasonality, when evaluating nasopharyngeal carriage information in infants.

Conclusions

The nasopharyngeal carriage of infectious agents such as *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, and *S. aureus* occurs prior to respiratory disease development and progression and is a source of transmission. In this study, we provide some insights into the age-effects and non-linear seasonality of nasopharyngeal carriage as a proxy for colonisation in young South African infants with and without lower respiratory tract infection. The complex relationships observed suggest that nasopharyngeal carriage may be a risk factor for diseases such as pneumonia in young children. Additionally, longitudinal studies on bacterial carriage in infants may identify external changes that are key to understanding dynamics of transmission and carriage to inform intervention strategies.

Supplementary data

Supplementary data are available in Part C: Appendix A.

Ethics approval

This research study was approved by the University of Cape Town (UCT) Human Research Ethics Committee (HREC) in Cape Town, South Africa (No. 757/2021) (Part C: Appendix B).

Funding

None

Data availability

The data underlying this article are available in the article and in its online supplementary material.

Conflict of interest

None declared.

References

1. Roux DM le, Nicol MP, Myer L, et al. Lower Respiratory Tract Infections in Children in a Well-vaccinated South African Birth Cohort: Spectrum of Disease and Risk Factors. *Clin Infect Dis*. 2019 Oct 15;**69**(9):1588–1596.
2. Simoes EAF, Cherian T, Chow J, Shahid-Salles SA, Laxminarayan R, John TJ. Acute Respiratory Infections in Children. *Dis Control Priorities Dev Ctries*. 2006. p. 483–497.
3. Walker CLF, Rudan I, Liu L, et al. Global burden of childhood pneumonia and diarrhoea. *Lancet* [Internet]. 2013 Apr;**381**(9875):1405–1416.
4. Williams DJ, Zhu Y, Grijalva CG, et al. Predicting Severe Pneumonia Outcomes in Children. *Pediatrics*. 2016;**138**(4):e20161019.
5. Weiser JN, Ferreira DM, Paton JC. Streptococcus pneumoniae: transmission, colonization and invasion. *Nat Rev Microbiol*. 2018 Jun 29;**16**(6):355–367.
6. Agrawal A, Murphy TF. Haemophilus influenzae Infections in the H. influenzae Type b Conjugate Vaccine Era. *J Clin Microbiol*. 2011 Nov;**49**(11):3728–3732.
7. Murphy TF, Parameswaran GI. Moraxella catarrhalis, a Human Respiratory Tract Pathogen. *Clin Infect Dis*. 2009 Jul;**49**(1):124–131.
8. Sakr A, Brégeon F, Mège J-L, Rolain J-M, Blin O. Staphylococcus aureus Nasal Colonization: An Update on Mechanisms, Epidemiology, Risk Factors, and Subsequent Infections. *Front Microbiol*. 2018 Oct 8;**9**.
9. Martin RM, Bachman MA. Colonization, Infection, and the Accessory Genome of Klebsiella pneumoniae. *Front Cell Infect Microbiol*. 2018 Jan 22;**8**.
10. Simell B, Auranen K, Käyhty H, et al. The fundamental link between pneumococcal carriage and disease. *Expert Rev Vaccines*. 2012 Jul;**11**(7):841–55.
11. Marchisio P, Gironi S, Esposito S, Schito GC, Mannelli S, Principi N. Seasonal variations in nasopharyngeal carriage of respiratory pathogens in healthy Italian children attending day-care centres or schools. *J Med Microbiol*. 2001 Dec;**50**(12):1095–1099.
12. Jackson S, Mathews KH, Pulanić D, et al. Risk factors for severe acute lower respiratory infections in children – a systematic review and meta-analysis. *Croat Med J*. 2013 Apr;**54**(2):110–121.
13. Moore DP, Green RJ, Cohen C, et al. Epidemiology and aetiology of community-acquired pneumonia in children: South African Thoracic Society guidelines (part 1). *South African Med J*. 2020;**110**(7):583–587.
14. Sonego M, Pellegrin MC, Becker G, Lazzerini M. Risk Factors for Mortality from Acute Lower Respiratory Infections (ALRI) in Children under Five Years of Age in Low and Middle-Income Countries: A Systematic Review and Meta-Analysis of Observational Studies. Sankoh OA, editor. *PLoS One*. 2015 Jan 30;**10**(1):e0116380.
15. Zar HJ, Barnett W, Myer L, Stein DJ, Nicol MP. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. *Thorax*. 2015 Jun 1;**70**(6):592 LP – 594.
16. Donald KA, Hoogenhout M, Plooy CP du, et al. Drakenstein Child Health Study (DCHS): investigating determinants of early child development and cognition. BMJ Paediatr Open. 2018 Jun 13;2(1):e000282.
17. Zar HJ, Barnett W, Stadler A, Gardner-Lubbe S, Myer L, Nicol MP. Aetiology of childhood pneumonia in a well vaccinated South African birth cohort: a nested case-control study of the Drakenstein Child Health Study. Lancet Respir Med. 2016 Jun;4(6):463–472.
18. Zar HJ. Bacterial and viral pneumonia: New insights from the Drakenstein Child Health Study. Paediatr Respir Rev. 2017 Sep;24:8–10.
19. Usuf E, Bottomley C, Adegbola RA, Hall A. Pneumococcal Carriage in Sub-Saharan Africa—A Systematic Review. Trotter CL, editor. PLoS One. 2014 Jan 20;9(1):e85001.
20. Kwambana BA, Barer MR, Bottomley C, Adegbola RA, Antonio M. Early acquisition and high nasopharyngeal co-colonisation by Streptococcus pneumoniae and three respiratory pathogens amongst Gambian new-borns and infants. BMC Infect Dis. 2011 Dec 20;11(1):175.

21. Usuf E, Bojang A, Hill PC, Bottomley C, Greenwood B, Roca A. Nasopharyngeal colonization of Gambian infants by Staphylococcus aureus and Streptococcus pneumoniae before the introduction of pneumococcal conjugate vaccines. New Microbes New Infect. 2016 Mar;10:13–18.

22. Labout JAM, Duijts L, Arends LR, et al. Factors Associated with Pneumococcal Carriage in Healthy Dutch Infants: The Generation R Study. J Pediatr. 2008 Dec;153(6):771-776.e1.

23. Kinabo GD, Ven A van der, Msuya LJ, et al. Dynamics of nasopharyngeal bacterial colonisation in HIV-exposed young infants in Tanzania. Trop Med Int Health. 2013 Mar;18(3):286–95.

24. Darboe MK, Fulford AJ, Secka O, Prentice AM. The dynamics of nasopharyngeal streptococcus pneumoniae carriage among rural Gambian mother-infant pairs. BMC Infect Dis. 2010 Dec 5;10(1):195.

25. Gómez E, Moore A, Sánchez J, et al. The epidemiology of Haemophilus influenzae type b carriage among infants and young children in Santo Domingo, Dominican Republic. Pediatr Infect Dis J. 1998 Sep;17(9):782–6.

26. Hill PC, Akisanya A, Sankareh K, et al. Nasopharyngeal Carriage of Streptococcus pneumoniae in Gambian Villagers. Clin Infect Dis. 2006 Sep 15;43(6):673–679.

27. Mackenzie GA, Leach AJ, Carapetis JR, Fisher J, Morris PS. Epidemiology of nasopharyngeal carriage of respiratory bacterial pathogens in children and adults: cross-sectional surveys in a population with high rates of pneumococcal disease. BMC Infect Dis. 2010 Dec 23;10(1):304.

28. Bojang A, Jafali J, Egere UE, et al. Seasonality of Pneumococcal Nasopharyngeal Carriage in Rural Gambia Determined within the Context of a Cluster Randomized Pneumococcal Vaccine Trial. Borrow R, editor. PLoS One. 2015 Jul 1;10(7):e0129649.

29. Syrjänen RK, Kilpi TM, Kaijalainen TH, Herva EE, Takala AK. Nasopharyngeal Carriage of Streptococcus pneumoniae in Finnish Children Younger Than 2 Years Old. J Infect Dis. 2001 Aug 15;184(4):451–459.

30. Tenenbaum T, Franz A, Neuhausen N, et al. Clinical characteristics of children with lower respiratory tract infections are dependent on the carriage of specific pathogens in the nasopharynx. Eur J Clin Microbiol Infect Dis. 2012 Nov 1;31(11):3173–3182.

31. Dunne EM, Choummanivong M, Neal EFG, et al. Factors associated with pneumococcal carriage and density in infants and young children in Laos PDR. Chia J-S, editor. PLoS One. 2019 Oct 29;14(10):e0224392.

32. Ueno M, Ishii Y, Tateda K, et al. Prevalence and risk factors of nasopharyngeal carriage of Streptococcus pneumoniae in healthy children in Japan. Jpn J Infect Dis. 2013;66(1):22–5.

PART C: APPENDICES

An investigation in the seasonal patterns of bacterial colonisation in childhood pneumonia

Marie Belle Kathrina M. Auckloo

(student no. ACKMAR005)

Appendix A: Journal manuscript supplementary material

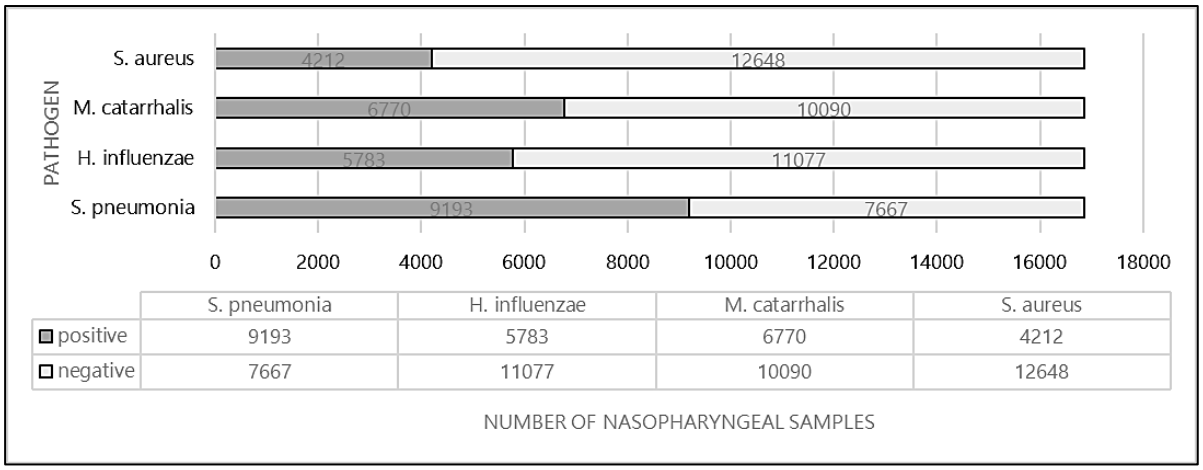


Figure S1 Prevalence of bacterial respiratory pathogens in nasopharyngeal samples (*n*=16 860).

Out of all nasopharyngeal samples (*n*=16 860), *S. pneumoniae* was predominantly detected as positive (*n*=9 193, 54.5%), followed by *H. influenzae* (*n*=5 783, 34.3%), *M. catarrhalis* (*n*=6 770, 40.2%), and *S. aureus* (*n*=4 212, 25.0%). For each bacterial respiratory pathogen, there were more negative nasopharyngeal samples than positive ones, with the exception of *S. pneumoniae* which was mostly positive across all samples (54.5%). Note: One nasopharyngeal sample may be positive for more than one organism at any given time.

Table S1 Prevalence of bacterial respiratory pathogens in nasopharyngeal samples (n=16 860) across infant age groups.

age (mths)	number of positive samples, *n* (%)				number of samples collected, *n* (%)
	S. pneumoniae **9 193 (54.5)**	***H. influenzae*** **5 783 (34.3)**	***M. catarrhalis*** **6 770 (40.2)**	***S. aureus*** **4 212 (25.0)**	
0-3	2 165 (12.8)	1 182 (7.0)	1 533 (9.1)	2 404 (14.3)	6185 (36.7)
3-6	2 938 (17.4)	1 833 (10.9)	2 088 (12.4)	1 067 (6.3)	4503 (26.7)
6-12	4 090 (24.3)	2 768 (16.4)	3 149 (18.7)	741 (4.4)	6172 (36.6)

Note: mths, months

Out of the nasopharyngeal samples collected from the cohort of infants, *S. aureus* was predominantly detected as positive (14.3%) among the 0-3 months old. *S. pneumoniae* was most prevalent among both 3-6 months (17.4%) and 6-12 months (24.3%) old infants. For each bacterial respiratory pathogen, the prevalence of nasopharyngeal carriage increases with age, with the exception of *S. aureus* whose prevalence decreased as the infants grow older. Note: One nasopharyngeal sample may be positive for more than one organism at any given time.

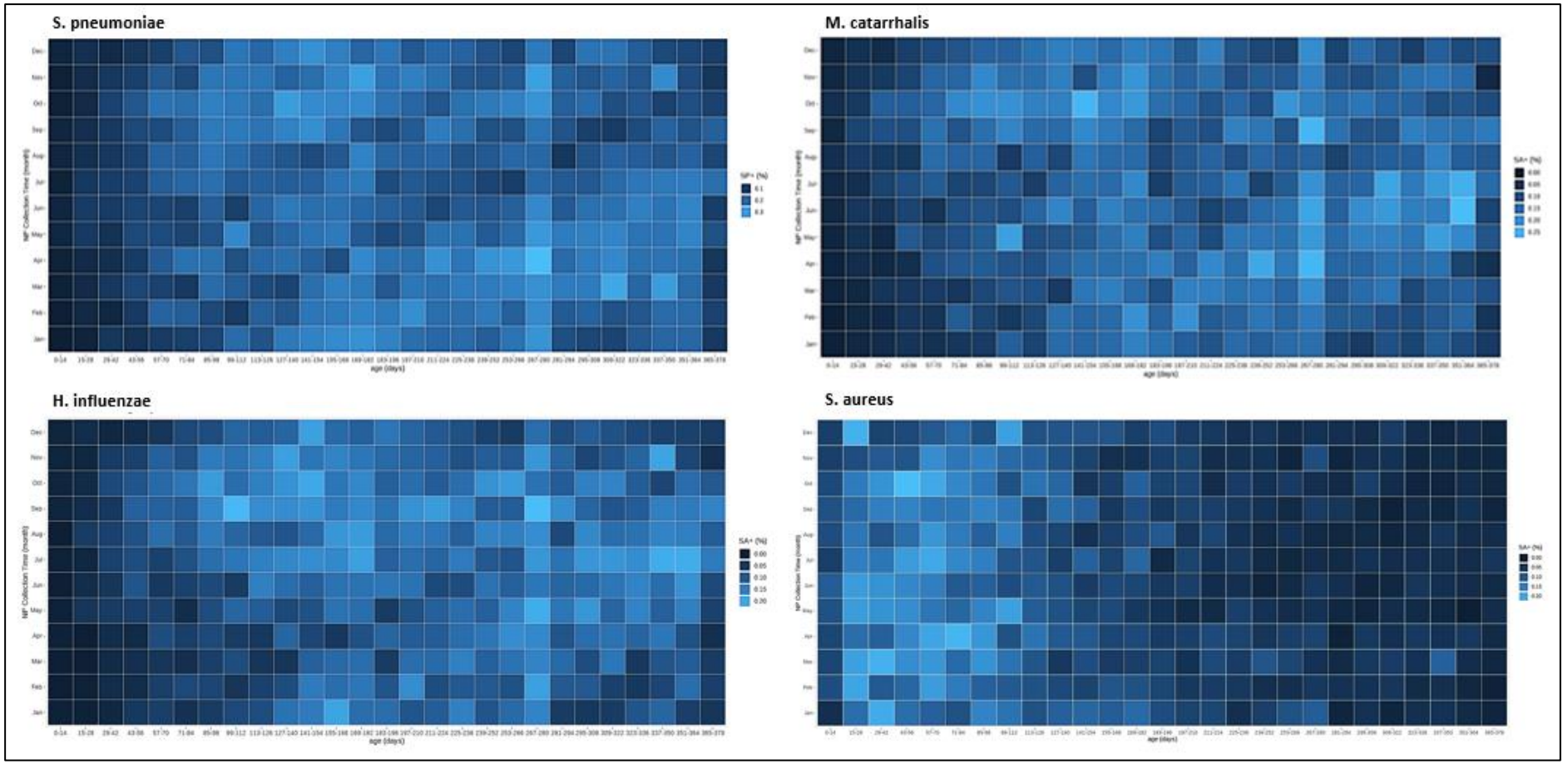


Figure S2 Distribution of positive nasopharyngeal samples (*n*=16 860) over time (infant age at collection in days) across calendar months for each pathogen.

Nasopharyngeal carriage increases over time for *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*. For *S. aureus*, nasopharyngeal carriage is most prevalent in the early months and decreases over time. Note: One nasopharyngeal sample may be positive for more than one organism at any given time.

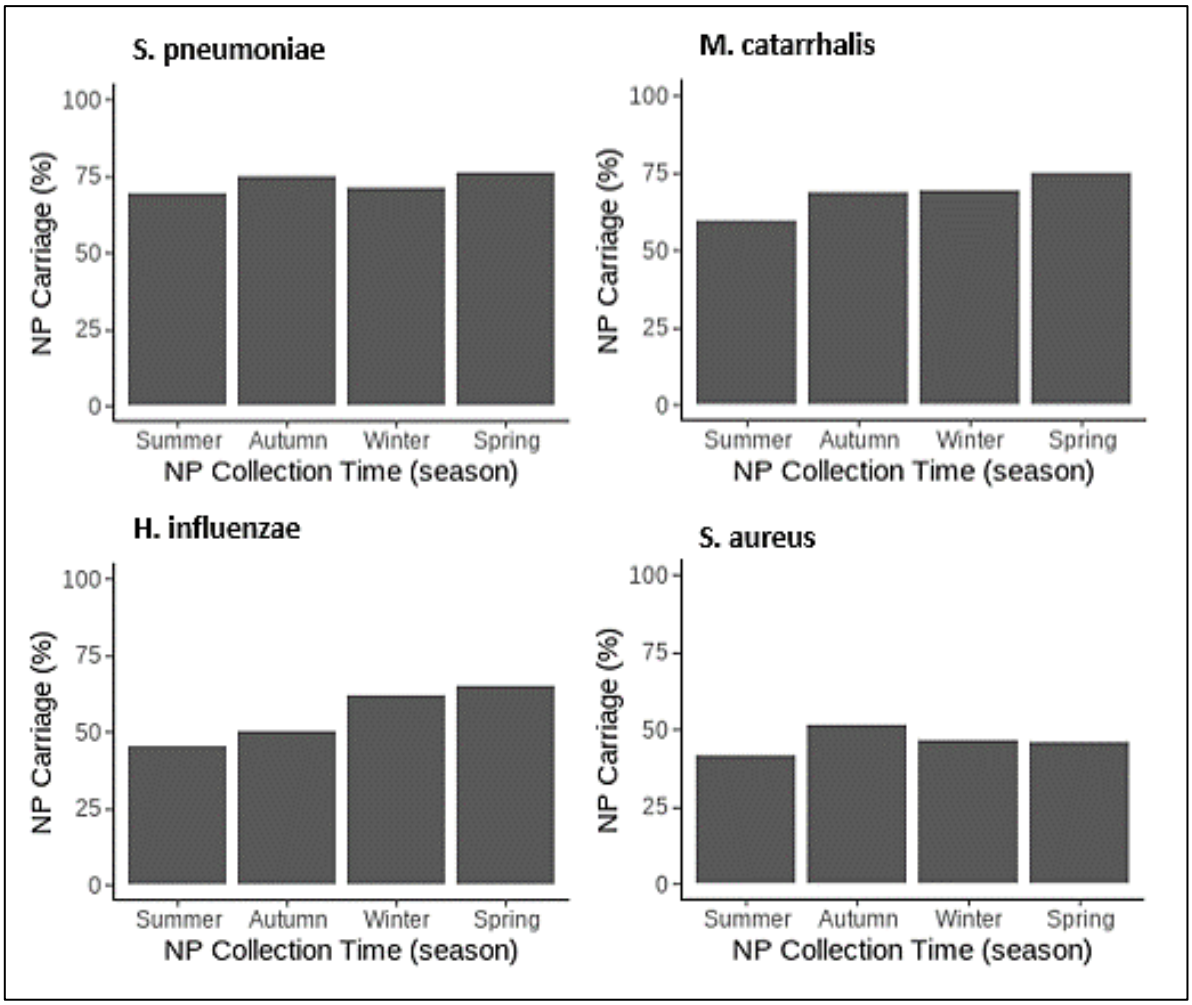


Figure S3 Prevalence of bacterial carriage in infants (*n*=851) at the time of collection according to seasons. Note: An infant can have nasopharyngeal carriage for more than one organism at any given time.

The carriage of *S. pneumoniae* was steady across seasons but slightly more common in autumn and spring. *H. influenzae* carriage appears to increase from summer to autumn, winter, and spring – when it is most prevalent. Similarly, the proportion of *M. catarrhalis* carriers was highest in spring and lowest in summer. Unlike the other bacterial pathogens, *S. aureus* carriage did not display a clear seasonal pattern, being most common in autumn and least prevalent in summer.

S_pneumo_bin ~ 1

	S_pneumo_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.20	1.16 – 1.24	**<0.001**
Observations	16860		
R^2 Tjur	0.000		

S_pneumo_bin ~ (1 | Infant_ID)

	S_pneumo_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.13	1.05 – 1.23	**0.002**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.03		
ICC	0.24		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.000 / 0.238		

Model A: **S_pneumo_bin ~ (1 | Infant_ID) + scale(age_collection_days)**

	S_pneumo_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.24	1.14 – 1.35	**<0.001**
age collection days	2.26	2.17 – 2.35	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.27		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.127 / 0.370		

Model B: **S_pneumo_bin ~ (1 | Infant_ID) + scale(age_collection_days) + season_collect**

	S_pneumo_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.30	1.17 – 1.45	**<0.001**
age collection days	2.27	2.18 – 2.36	**<0.001**
season collect [Autumn]	1.07	0.97 – 1.19	0.191
season collect [Winter]	0.80	0.72 – 0.89	**<0.001**
season collect [Spring]	0.97	0.88 – 1.08	0.590
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.28		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.130 / 0.374		

Model C: **S_pneumo_bin ~ (1 | Infant_ID) + scale(age_collection_days) + (1 | Infant_ID) + sinday + cosday + sinday_v2 + cosday_v2**

	S_pneumo_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.25	1.14 – 1.36	**<0.001**
age collection days	2.27	2.18 – 2.36	**<0.001**
sinday	1.08	1.02 – 1.13	**0.005**
cosday	1.12	1.06 – 1.18	**<0.001**
sinday v2	0.90	0.85 – 0.95	**<0.001**
cosday v2	0.94	0.89 – 0.99	**0.012**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.29		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.130 / 0.375		

Figure S4(a) Regression outputs for the computed models with *S. pneumoniae* carriage as the outcome.

Logistic mixed models (estimated using ML and BOBYQA optimizer) were conducted using statistical software R, with the lme4 (Bates et al., 2015) package.

Model D: **S_pneumo_bin ~ (1 | Infant_ID) + scaled_age * season_collect**

	S_pneumo_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.31	1.17 – 1.46	**<0.001**
scaled age	2.33	2.14 – 2.54	**<0.001**
season collect [Autumn]	1.08	0.97 – 1.20	0.172
season collect [Winter]	0.79	0.72 – 0.88	**<0.001**
season collect [Spring]	0.96	0.87 – 1.07	0.494
scaled age * season collect [Autumn]	1.11	0.98 – 1.25	0.099
scaled age * season collect [Winter]	0.94	0.84 – 1.06	0.314
scaled age * season collect [Spring]	0.88	0.79 – 0.99	**0.032**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.28		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.131 / 0.374		

Model E: **S_pneumo_bin ~ (1 | Infant_ID) + scaled_age * lrti**

	S_pneumo_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.23	1.13 – 1.35	**<0.001**
scaled age	2.29	2.20 – 2.38	**<0.001**
lrti [1]	1.23	0.98 – 1.54	0.072
scaled age * lrti [1]	0.63	0.51 – 0.79	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.27		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.128 / 0.372		

Model F: **S_pneumo_bin ~ (1 | Infant_ID) + scaled_age * season_collect * lrti**

	S_pneumo_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	1.32	1.18 – 1.47	**<0.001**
scaled age	2.37	2.17 – 2.58	**<0.001**
season collect [Autumn]	1.07	0.96 – 1.20	0.202
season collect [Winter]	0.78	0.70 – 0.87	**<0.001**
season collect [Spring]	0.95	0.85 – 1.05	0.323
lrti [1]	0.89	0.52 – 1.53	0.676
scaled age * season collect [Autumn]	1.11	0.98 – 1.26	0.095
scaled age * season collect [Winter]	0.93	0.83 – 1.05	0.256
scaled age * season collect [Spring]	0.88	0.78 – 0.99	**0.033**
scaled age * lrti [1]	0.56	0.32 – 0.98	**0.041**
season collect [Autumn] * lrti [1]	1.14	0.57 – 2.28	0.714
season collect [Winter] * lrti [1]	1.65	0.83 – 3.27	0.152
season collect [Spring] * lrti [1]	1.66	0.84 – 3.28	0.143
(scaled age * season collect [Autumn]) * lrti [1]	1.12	0.56 – 2.24	0.756
(scaled age * season collect [Winter]) * lrti [1]	1.37	0.68 – 2.77	0.380
(scaled age * season collect [Spring]) * lrti [1]	1.02	0.50 – 2.07	0.965
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.28		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.133 / 0.376		

Figure S4(b) Regression outputs for the computed models with *S. pneumoniae* carriage as the outcome.

Logistic mixed models (estimated using ML and BOBYQA optimizer) were conducted using statistical software R, with the lme4 (Bates et al., 2015) package.

H_influenzae_bin ~ 1

	H_influenzae_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.52	0.51 – 0.54	**<0.001**
Observations	16860		
R^2 Tjur	0.000		

H_influenzae_bin ~ (1 | Infant_ID)

	H_influenzae_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.41	0.37 – 0.44	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.11		
ICC	0.25		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.000 / 0.253		

Model A: **H_influenzae_bin ~ (1 | Infant_ID) + scale(age_collection_days)**

	H_influenzae_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.40	0.37 – 0.44	**<0.001**
age collection days	1.90	1.83 – 1.97	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.20		
ICC	0.27		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.084 / 0.328		

Model B: **H_influenzae _bin ~ (1 | Infant_ID) + scale(age_collection_days) + season_collect**

	H_influenzae_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.30	0.27 – 0.34	**<0.001**
age collection days	1.91	1.84 – 1.99	**<0.001**
season collect [Autumn]	0.99	0.89 – 1.11	0.849
season collect [Winter]	1.62	1.45 – 1.80	**<0.001**
season collect [Spring]	1.75	1.58 – 1.95	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.25		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.098 / 0.347		

Model C: **H_influenzae _bin ~ (1 | Infant_ID) + scale(age_collection_days) + (1 | Infant_ID) + sinday + cosday + sinday_v2 + cosday_v2**

	H_influenzae_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.38	0.35 – 0.42	**<0.001**
age collection days	1.92	1.85 – 2.00	**<0.001**
sinday	0.68	0.65 – 0.72	**<0.001**
cosday	0.82	0.77 – 0.86	**<0.001**
sinday v2	1.08	1.03 – 1.14	**0.003**
cosday v2	0.96	0.91 – 1.01	0.110
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.27		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.104 / 0.354		

Figure S5(a) Regression outputs for the computed models with *H. influenzae* carriage as the outcome.

Logistic mixed models (estimated using ML and BOBYQA optimizer) were conducted using statistical software R, with the lme4 (Bates et al., 2015) package.

Model D: **H_influenzae_bin ~ (1 | Infant_ID) + scaled_age * season_collect**

	H_influenzae_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.31	0.28 – 0.35	**<0.001**
scaled age	1.61	1.48 – 1.75	**<0.001**
season collect [Autumn]	0.94	0.84 – 1.06	0.315
season collect [Winter]	1.55	1.39 – 1.73	**<0.001**
season collect [Spring]	1.70	1.53 – 1.89	**<0.001**
scaled age * season collect [Autumn]	1.33	1.18 – 1.50	**<0.001**
scaled age * season collect [Winter]	1.29	1.15 – 1.45	**<0.001**
scaled age * season collect [Spring]	1.14	1.02 – 1.28	**0.021**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.25		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.100 / 0.348		

Model E: **H_influenzae_bin ~ (1 | Infant_ID) + scaled_age * lrti**

	H_influenzae_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.39	0.36 – 0.42	**<0.001**
scaled age	1.91	1.84 – 1.99	**<0.001**
lrti [1]	2.88	2.31 – 3.61	**<0.001**
scaled age * lrti [1]	0.69	0.56 – 0.86	**0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.22		
ICC	0.27		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.089 / 0.336		

Model F: **H_influenzae_bin ~ (1 | Infant_ID) + scaled_age * season_collect * lrti**

	H_influenzae_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.30	0.27 – 0.34	**<0.001**
scaled age	1.63	1.50 – 1.78	**<0.001**
season collect [Autumn]	0.93	0.83 – 1.04	0.200
season collect [Winter]	1.52	1.36 – 1.70	**<0.001**
season collect [Spring]	1.69	1.51 – 1.88	**<0.001**
lrti [1]	2.49	1.43 – 4.32	**0.001**
scaled age * season collect [Autumn]	1.31	1.16 – 1.48	**<0.001**
scaled age * season collect [Winter]	1.29	1.14 – 1.46	**<0.001**
scaled age * season collect [Spring]	1.13	1.01 – 1.27	**0.039**
scaled age * lrti [1]	0.46	0.26 – 0.82	**0.008**
season collect [Autumn] * lrti [1]	1.18	0.59 – 2.39	0.637
season collect [Winter] * lrti [1]	1.50	0.75 – 2.99	0.254
season collect [Spring] * lrti [1]	1.05	0.53 – 2.08	0.896
(scaled age * season collect [Autumn]) * lrti [1]	1.86	0.91 – 3.79	0.088
(scaled age * season collect [Winter]) * lrti [1]	1.13	0.55 – 2.31	0.734
(scaled age * season collect [Spring]) * lrti [1]	2.19	1.04 – 4.63	**0.039**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.28		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.106 / 0.356		

Figure S5(b) Regression outputs for the computed models with *H. influenzae* carriage as the outcome.

Logistic mixed models (estimated using ML and BOBYQA optimizer) were conducted using statistical software R, with the lme4 (Bates et al., 2015) package.

M_catarrhalis_bin ~ 1

	M_catarrhalis_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.67	0.65 – 0.69	**<0.001**
Observations	16860		
R^2 Tjur	0.000		

M_catarrhalis_bin ~ (1 | Infant_ID)

	M_catarrhalis_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.61	0.58 – 0.65	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	0.45		
ICC	0.12		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.000 / 0.121		

Model A: **M_catarrhalis_bin ~ (1 | Infant_ID) + scale(age_collection_days)**

	M_catarrhalis_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.62	0.58 – 0.65	**<0.001**
age collection days	1.79	1.73 – 1.85	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	0.49		
ICC	0.13		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.082 / 0.202		

Model B: **M_catarrhalis_bin ~ (1 | Infant_ID) + scale(age_collection_days) + season_collect**

	M_catarrhalis_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.53	0.48 – 0.58	**<0.001**
age collection days	1.79	1.73 – 1.86	**<0.001**
season collect [Autumn]	1.12	1.01 – 1.24	**0.025**
season collect [Winter]	1.15	1.05 – 1.27	**0.004**
season collect [Spring]	1.36	1.24 – 1.50	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	0.50		
ICC	0.13		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.085 / 0.206		

Model C: **M_catarrhalis_bin ~ (1 | Infant_ID) + scale(age_collection_days) + (1 | Infant_ID) + sinday + cosday + sinday_v2 + cosday_v2**

	M_catarrhalis_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.61	0.57 – 0.64	**<0.001**
age collection days	1.79	1.73 – 1.86	**<0.001**
sinday	0.89	0.84 – 0.93	**<0.001**
cosday	0.92	0.88 – 0.97	**0.002**
sinday v2	0.89	0.85 – 0.94	**<0.001**
cosday v2	0.93	0.89 – 0.98	**0.005**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	0.50		
ICC	0.13		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.087 / 0.208		

Figure S6(a) Regression outputs for the computed models with *M. catarrhalis* carriage as the outcome.

Logistic mixed models (estimated using ML and BOBYQA optimizer) were conducted using statistical software R, with the lme4 (Bates et al., 2015) package.

Model D: **M_catarrhalis_bin ~ (1 | Infant_ID) + scaled_age * season_collect**

	M_catarrhalis_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.53	0.49 – 0.58	**<0.001**
scaled age	1.73	1.60 – 1.87	**<0.001**
season collect [Autumn]	1.12	1.01 – 1.23	**0.032**
season collect [Winter]	1.14	1.03 – 1.26	**0.011**
season collect [Spring]	1.36	1.23 – 1.50	**<0.001**
scaled age * season collect [Autumn]	1.05	0.94 – 1.17	0.386
scaled age * season collect [Winter]	1.13	1.01 – 1.26	**0.029**
scaled age * season collect [Spring]	0.97	0.87 – 1.07	0.518
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	0.49		
ICC	0.13		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.087 / 0.205		

Model E: **M_catarrhalis_bin ~ (1 | Infant_ID) + scaled_age * lrti**

	M_catarrhalis_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.61	0.58 – 0.65	**<0.001**
scaled age	1.81	1.74 – 1.87	**<0.001**
lrti [1]	1.19	0.96 – 1.47	0.107
scaled age * lrti [1]	0.73	0.59 – 0.90	**0.003**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	0.49		
ICC	0.13		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.083 / 0.203		

Model F: **M_catarrhalis_bin ~ (1 | Infant_ID) + scaled_age * season_collect * lrti**

	M_catarrhalis_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.53	0.48 – 0.58	**<0.001**
scaled age	1.75	1.62 – 1.89	**<0.001**
season collect [Autumn]	1.13	1.02 – 1.25	**0.021**
season collect [Winter]	1.14	1.03 – 1.26	**0.009**
season collect [Spring]	1.36	1.24 – 1.50	**<0.001**
lrti [1]	1.65	0.99 – 2.74	0.053
scaled age * season collect [Autumn]	1.05	0.94 – 1.17	0.394
scaled age * season collect [Winter]	1.12	1.01 – 1.25	**0.037**
scaled age * season collect [Spring]	0.97	0.87 – 1.07	0.508
scaled age * lrti [1]	0.57	0.33 – 0.96	**0.036**
season collect [Autumn] * lrti [1]	0.59	0.31 – 1.13	0.113
season collect [Winter] * lrti [1]	0.69	0.36 – 1.32	0.260
season collect [Spring] * lrti [1]	0.72	0.38 – 1.36	0.311
(scaled age * season collect [Autumn]) * lrti [1]	1.36	0.71 – 2.61	0.354
(scaled age * season collect [Winter]) * lrti [1]	1.42	0.73 – 2.75	0.302
(scaled age * season collect [Spring]) * lrti [1]	1.26	0.65 – 2.47	0.496
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	0.49		
ICC	0.13		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.088 / 0.206		

Figure S6(b) Regression outputs for the computed models with *M. catarrhalis* carriage as the outcome.

Logistic mixed models (estimated using ML and BOBYQA optimizer) were conducted using statistical software R, with the lme4 (Bates et al., 2015) package.

S_aureus_bin ~ 1

	S_aureus_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.33	0.32 – 0.34	**<0.001**
Observations	16860		
R^2 Tjur	0.000		

S_aureus_bin ~ (1 | Infant_ID)

	S_aureus_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.26	0.24 – 0.29	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.28		
ICC	0.28		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.000 / 0.281		

Model A: **S_aureus_bin ~ (1 | Infant_ID) + scale(age_collection_days)**

	S_aureus_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.22	0.20 – 0.24	**<0.001**
age collection days	0.48	0.46 – 0.50	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.44		
ICC	0.30		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.101 / 0.375		

Model B: **S_aureus_bin ~ (1 | Infant_ID) + scale(age_collection_days) + season_collect**

	S_aureus_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.25	0.22 – 0.28	**<0.001**
age collection days	0.48	0.46 – 0.51	**<0.001**
season collect [Autumn]	1.01	0.90 – 1.14	0.833
season collect [Winter]	0.80	0.71 – 0.90	**<0.001**
season collect [Spring]	0.80	0.71 – 0.90	**<0.001**
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.44		
ICC	0.30		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.103 / 0.376		

Model C: **S_aureus_bin ~ (1 | Infant_ID) + scale(age_collection_days) + (1 | Infant_ID) + sinday + cosday + sinday_v2 + cosday_v2**

	S_aureus_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.22	0.20 – 0.24	**<0.001**
age collection days	0.48	0.46 – 0.51	**<0.001**
sinday	1.15	1.09 – 1.22	**<0.001**
cosday	1.09	1.02 – 1.15	**0.006**
sinday v2	1.04	0.99 – 1.10	0.126
cosday v2	0.98	0.93 – 1.04	0.478
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.44		
ICC	0.30		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.103 / 0.377		

Figure S7(a) Regression outputs for the computed models with *S. aureus* carriage as the outcome.

Logistic mixed models (estimated using ML and BOBYQA optimizer) were conducted using statistical software R, with the lme4 (Bates et al., 2015) package.

Model D: **S_aureus_bin ~ (1 | Infant_ID) + scaled_age * season_collect**

	S_aureus_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.25	0.22 – 0.28	**<0.001**
scaled age	0.50	0.45 – 0.55	**<0.001**
season collect [Autumn]	0.99	0.88 – 1.12	0.888
season collect [Winter]	0.80	0.71 – 0.91	**<0.001**
season collect [Spring]	0.80	0.71 – 0.89	**<0.001**
scaled age * season collect [Autumn]	0.90	0.79 – 1.02	0.108
scaled age * season collect [Winter]	1.01	0.89 – 1.16	0.828
scaled age * season collect [Spring]	0.98	0.86 – 1.12	0.758
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.44		
ICC	0.30		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.103 / 0.376		

Model E: **S_aureus_bin ~ (1 | Infant_ID) + scaled_age * lrti**

	S_aureus_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.22	0.20 – 0.24	**<0.001**
scaled age	0.48	0.46 – 0.51	**<0.001**
lrti [1]	1.05	0.81 – 1.37	0.703
scaled age * lrti [1]	0.83	0.62 – 1.12	0.218
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	1.44		
ICC	0.30		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.101 / 0.375		

Model F: **M_catarrhalis_bin ~ (1 | Infant_ID) + scaled_age * season_collect * lrti**

	M_catarrhalis_bin		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.53	0.48 – 0.58	**<0.001**
scaled age	1.75	1.62 – 1.89	**<0.001**
season collect [Autumn]	1.13	1.02 – 1.25	**0.021**
season collect [Winter]	1.14	1.03 – 1.26	**0.009**
season collect [Spring]	1.36	1.24 – 1.50	**<0.001**
lrti [1]	1.65	0.99 – 2.74	0.053
scaled age * season collect [Autumn]	1.05	0.94 – 1.17	0.394
scaled age * season collect [Winter]	1.12	1.01 – 1.25	**0.037**
scaled age * season collect [Spring]	0.97	0.87 – 1.07	0.508
scaled age * lrti [1]	0.57	0.33 – 0.96	**0.036**
season collect [Autumn] * lrti [1]	0.59	0.31 – 1.13	0.113
season collect [Winter] * lrti [1]	0.69	0.36 – 1.32	0.260
season collect [Spring] * lrti [1]	0.72	0.38 – 1.36	0.311
(scaled age * season collect [Autumn]) * lrti [1]	1.36	0.71 – 2.61	0.354
(scaled age * season collect [Winter]) * lrti [1]	1.42	0.73 – 2.75	0.302
(scaled age * season collect [Spring]) * lrti [1]	1.26	0.65 – 2.47	0.496
Random Effects			
σ^2	3.29		
$\tau_{00\ Infant_ID}$	0.49		
ICC	0.13		
N_{Infant_ID}	851		
Observations	16860		
Marginal R^2 / Conditional R^2	0.088 / 0.206		

Figure S7(b) Regression outputs for the computed models with *S. aureus* carriage as the outcome.

Logistic mixed models (estimated using ML and BOBYQA optimizer) were conducted using statistical software R, with the lme4 (Bates et al., 2015) package.

Appendix B: Study ethics approval letter

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room 45, E-52- Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6492
Email: hrec-enquiries@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

16 November 2021

HREC REF: 757/2021

A/Prof M Lesosky
School of Public Health & Family Medicine
Falmouth Building-FHS
Email: Maia.lesosky@uct.ac.za
Student: ackmar005@myuct.ac.za

Dear A/Prof Lesosky

PROJECT TITLE: AN INVESTIGATION IN THE SEASONAL PATTERNS OF BACTERIAL COLONISATION IN CHILDHOOD PNEUMONIA-MASTERS CANDIDATE-MS MARIE AUCKLOO-SUB-STUDY LINKED TO 485/2019

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee (HREC) for review.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

This approval is subject to strict adherence to the HREC recommendations regarding research involving human participants during COVID -19, dated 17 March 2020; 06 July 2020 & 01 July 2021.

Approval is granted for one year until the 30 November 2022.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

The HREC acknowledge that the student: Ms Marie Auckloo will also be involved in this study.

Please quote the HREC REF 757/2021 in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

HREC/REF 757/2021sa

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FACULTY OF HEALTH SCIENCES HUMAN RESEARCH ETHICS COMMITTEE
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
NHREC-registration number: REC-210208-007

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use: Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH 2006), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI), and Declaration of Helsinki (2013) guidelines. The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

HREC/REF 757/2021sa

Appendix C: Author Guidelines for the International Journal of Epidemiology

Information for Authors

The *International Journal of Epidemiology* is essential for readers wishing to keep up to date with epidemiological advances and developments throughout the world.

It encourages communication among those engaged in the research, teaching and application of epidemiology of both communicable and non-communicable disease, including research into health services and medical care.

The *International Journal of Epidemiology* also covers new methods – epidemiological and statistical – for the analysis of data used by those who practise social and preventive medicine.

Contents

- Editorial policies
  - Funders
  - Pre-submission service
  - Author self-archiving, pre-print, and public access policy
  - Open access and licensing
  - Offprints
- *IJE* forms
- Manuscript preparation
- Submitting an article to the *IJE*
- The publication process
- Supplements
- Types of articles published by the *IJE*

Editorial policies

- All submissions must be in the English language.
- Articles are accepted for publication on the condition that they are contributed solely to the *International Journal of Epidemiology*.
- The editors cannot enter into correspondence about papers considered unsuitable for publication and their decision is final.
- Neither the editors nor the publishers accept responsibility for the views and statements of authors expressed in their contributions.
- By submitting an article for publication, you confirm that you are the corresponding/submitting author and that Oxford University Press ('OUP') may retain your email address to communicate with you about the article. You agree to notify OUP immediately if your details change. If your article is accepted for publication, OUP will contact you using the email address you have used in the registration process.
- It is a condition of publication in the Journal that authors grant an exclusive licence to the International Epidemiological Association. This ensures that requests from third parties to reproduce articles are handled efficiently and consistently and will also allow the article to be as widely disseminated as possible. As part of the licence agreement, authors may use their own material in other publications provided that the Journal is acknowledged as the original place of publication and Oxford University Press as the Publisher.

Funders

In line with a growing number of health and scientific journals, the *International Journal of Epidemiology* will not consider for publication studies that are to any extent funded by the tobacco industry.

Pre-submission service

The *International Journal of Epidemiology* is not staffed to provide a pre-submission service. You will need to submit your manuscript to find out whether the editors are interested in publishing it.

Author self-archiving, pre-print, and public access policy

Authors retain the right to make an Author's Original Version (preprint) available through various channels, and the editors will consider papers that have been posted on bioRxiv, medRxiv and other preprint servers before submission. For further information see our Online Licensing, Copyright and Permissions policies.

If accepted, the authors are required to update the status of any preprint, including your published paper's DOI, as described on our Author Self-Archiving Policy page.

Open access and licensing

All articles published in the *International Journal of Epidemiology* are free to view from 12 months after they are published.

IJE offers the option of publishing under either a standard licence or an open access licence. Please note that some funders require open access publication as a condition of funding. If you are unsure whether you are required to publish open access, please do clarify any such requirements with your funder or institution.

Should you wish to publish your article open access, you should select your choice of open access licence in our online system after your article has been accepted for publication. You will need to pay an open access charge to publish under an open access licence.

Details of the open access licences and open access charges.

OUP has a growing number of Read and Publish agreements with institutions and consortia which provide funding for open access publishing. This means authors from participating institutions can publish open access, and the institution may pay the charge. Find out if your institution is participating.

Offprints

Offprints may be purchased using the Oxford Journals Author Services site by using the link sent to you after your paper is accepted. Orders from the UK will be subject to a 20% VAT charge. For orders from elsewhere in the EU, you or your institution should account for VAT by way of a reverse charge. Please provide Author Services with your or your institution's VAT number.

Back to Contents

IJE forms

Authorship Form

A scanned copy of the Authorship form (click to download) signed by the corresponding author on behalf of all authors must be submitted via the online manuscript submission system when you first submit your article.

It is the responsibility of the corresponding author to ensure that all authors listed on a paper meet the authorship criteria agreed by the International Committee of Medical Journal Editors (ICMJE), and that no one who meets these criteria is omitted. The ICMJE criteria are described in detail on the *IJE* Authorship Form.

You will be unable to complete your submission if the Authorship Form is not included. Failure to submit a properly completed Authorship Form will lead to delays in the processing of the manuscript.

Conflict of Interest Form

A Conflict of Interest (COI) form (click to download) must be submitted via the online manuscript submission system when you first submit your article. The form must be completed and signed by the corresponding author on behalf of all authors. On page 2 of the form, please complete *only one* of Section I or Section II, as appropriate.

It is the responsibility of the corresponding author to ensure that all authors have declared any potential conflicts of interest and that they adhere to the conflict of interest policy described on the COI form.

You will be unable to complete your submission if the COI form is not included. Failure to submit a properly completed COI form will lead to delays in the processing of the manuscript. A form must be submitted even if there is no conflict of interest.

A *conflict of interest statement* must also be included in the manuscript after any 'Funding' and 'Acknowledgements' sections and should summarize all aspects of any conflicts of interest declared by authors on the COI form. If there is no conflict of interest, authors must state 'Conflict of Interest: None declared'.

Licence to Publish Form

The Licence to Publish Form is only required when a paper is accepted for publication. Please note that the journal now encourages authors to complete their copyright licence to publish form online, and a link to the form will be sent to authors by Oxford Journals Author Services at the relevant time.

Back to Contents

Manuscript preparation

Papers must be submitted in good, grammatical English. If English is not your first language, you may wish to have your paper edited by an English-language editor before submitting.

Initial presentation

- Please check the specific instructions for the type of article you are submitting for full details of word limits and required article structure.
- The main document of an article should be submitted as an editable Word document (not as a pdf file*) and designated 'Main document' when uploaded.
- The main document may contain all tables and figures at submission, or these may be provided separately as 'Table' and 'Figure' files (note that you must be able to supply high-resolution versions of your figures if your paper is later accepted).
- Headings and titles of tables and figures should be presented in sentence case (i.e., capitalise only the first word and any proper nouns).
- Manuscripts should be double spaced with margins of 2.5cm.
- The main document should include a title page with author names and affiliations.
- The word count should be included on the title page.
- All pages should be numbered.
- Footnotes are not permissible in the text of a paper.
- UK, not US English, should be used, and jargon should be avoided.
- References should be formatted according to *IJE* style.

Supplementary material for online-only publication only should be submitted separately. It may comprise one or more files and each should be designated 'Supplementary file (for online publication only)' when uploaded.

* If your manuscript was created with LaTeX, please submit it as a PDF file initially. Either word processor or original LaTeX files will need to be supplied should your article proceed to acceptance.

Presentation for accepted papers

If accepted, papers must be presented as several files as follows:

- The main document of an article should be submitted as an editable word processor document (not as a pdf file) and should contain the text (including title page, abstract and references) only.* This file should be designated 'Main document' when uploaded.
- Tables should be provided as a separate editable word processor document (not as a pdf file) and designated 'Table' when uploaded.
- Each figure should be provided as a separate high-resolution file in an image format (such as tiff, eps or jpg) and designated 'Figure' when uploaded.
- Supplementary material for online publication only may comprise one or more files and each should be designated 'Supplementary file (for online publication only)' when uploaded.

* If your manuscript was created with LaTeX, you must provide either word processor or original LaTeX files if your article proceeds to acceptance.

General instructions

The following general instructions for manuscript preparation should be followed in conjunction with the specific requirements for the type of article being submitted.

Quick links to:
Title; Author list; How to list consortia and working groups; Availability of data and materials; Abbreviations/acronyms; Numbers and units; Describing significance; Special notes for statistical papers; Tables; Figures; Footnotes to tables and figures; Copyright permission; Declarations; Ethics approval; Author contributions; Data availability; Supplementary data; Funding; Acknowledgements; Conflict of interest; References; Citing preprints; Citing data sources; Appendices; Supplementary material

Title

Titles should be short and specific. Subtitles may be used to amplify the main title.

Author list

The *IJE* follows the recommendations of the International Committee of Medical Journal Editors (ICMJE). To comply with ICMJE recommendations, all the authors must meet *all* of the following four conditions:

- substantial contribution to conception and design, acquisition of data, or analysis and interpretation of data;
- drafting the article or revising it critically for important intellectual content;
- final approval of the version to be published; and
- agreement to be accountable for all aspects of the work thereby ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

All authors must be added to the author section within ScholarOne on submission of a paper. The corresponding author should ensure that the details added to ScholarOne for each author are correct. If a co-author is already listed in the system, but with incorrect or out-of-date details, the corresponding author should advise the relevant co-author(s) to correct these errors in their user accounts.

The contribution of each author and the identity of the author who will act as guarantor for the paper should be provided in the 'Author contributions' section (detailed instructions below).

The affiliations of each author must be given. If an author's present affiliation is different from that under which the work was done, both should be given.

The name of the corresponding author should be marked with an asterisk (*) and both a postal and an email address should be provided in the "*Corresponding author" under the affiliations list. Only one corresponding author is permitted.

The *IJE* allows authors to appear as joint first or senior authors or as equal contributors. Such authors should be marked with a superscript symbol (e.g. †) and a footnote describing their contribution (e.g. '†Joint first authors') should appear under the details for the corresponding author.

Note that the *IJE* will not publish studies that use data, infrastructure or personnel from any low- or middle-income country unless at least one researcher included in the author list is a national of that country.

Back to General instructions

How to list consortia and working groups

If a consortium or working group is involved, all members of the group who meet the full criteria and requirements for authorship, as described above, may be listed individually as authors. Alternatively, the official group or consortium name may appear at the end of the byline but must be accompanied by at least one named author who meets full authorship criteria. A single named author in the byline must be designated the Corresponding author and presented as described above. Affiliation details must be provided under the byline for named authors.

Remaining members of the group should be listed in a section headed "Notes" at the end of the text. For example:

> ### Notes
>
> The ABC Consortium: John B Smith (affiliation), Richard B Jones (affiliation), etc.

PubMed (but not PubMed Central) will list as collaborators the names of individual group members appearing in the Notes section when the group or consortium name is included in the author byline. PubMed will not list affiliations or other information from the Notes section. Please see the relevant PubMed rules for further information.

A consortium or working group may alternatively designate one or more individuals who meet full authorship criteria as authors writing "on behalf of" the group. In this case, the official group or consortium name should be included in the main author list after "on behalf of", and remaining group members may be listed at the end of the text under 'Notes" as shown above. Note that PubMed Central will include a group or consortium name following "on behalf of" in the author byline, but PubMed will omit it.

If named authors are writing "on behalf of" a group or consortium, the names of individual group members appearing in the Notes section will not be listed as collaborators in either PubMed or PubMed Central.

Back to General instructions

Availability of data and materials

Where ethically feasible, the *International Journal of Epidemiology* strongly encourages authors to make all data and software code on which the conclusions of the paper rely available to readers. Authors are required to include a Data Availability Statement in their article.

We suggest that data be presented in the main manuscript or additional supplementary material, or deposited in a public repository whenever possible. Information on general repositories for all data types, and a list of recommended repositories by subject area, is available on the Research Data Policy page.

Back to General instructions

Abbreviations/acronyms

- The use of abbreviations should be kept to a minimum.
- Non-standard abbreviations should be avoided.
- All abbreviations should be spelled out in full with the abbreviation in parentheses the first time they appear in:
    - the Abstract;
    - the Key Messages or Key Features;
    - the text;
    - tables and figures (usually as footnotes); and
    - any supplementary material.
- Thereafter the abbreviation should be used.
- Abbreviations should be avoided in titles, table and figure titles, headings and sub-headings.
- Any abbreviation that appears in a table or figure must be spelled out in full in a footnote to that table or figure. This applies also to Supplementary tables and figures.

Numbers and units

- In the text, numbers from one to nine should be spelled out and numbers above nine given as Arabic numerals.
- Numbers followed by a unit should always be written as numerals.
- Numerals should not be used to start a sentence.
- Numerals up to 9999 should not be separated by spaces or commas, while those from 10 000 on should use a space as a separator.
- Per cent should be written as % throughout.
- All measures should be reported in SI units, followed, in the text, by traditional units in parentheses. There are two exceptions: blood pressure should be expressed in mmHg and haemoglobin as g/dL.
- Where age grouping is appropriate, ages should be grouped as mid-decade to mid-decade (i.e. in five-year age groups such as 35–44 or 35–39, 40–44, etc, but not in full decades such as 20–29, 30–39 or other groupings).

Back to General instructions

Describing significance

In the *IJE*, we actively discourage the use of the term 'statistically significant' or just 'significant' and such statements in method sections as 'findings at P<0.05 were considered significant'. Please provide effect estimates with confidence intervals and exact P values, and refrain from using the term 'significant' in either the results or discussion sections of papers.

Our justification of this position is given in: Sterne J, Davey-Smith G. Sifting the evidence — What's wrong with significance tests? BMJ 2001; 322: 226-231. See also: Wasserstein RL, Lazar NA. The ASA's statement on P-values: context, process, and purpose. The American Statistician 2016: DOI: 10.1080/00031305.2016.1154108.

Special notes for statistical papers

The correct preparation of statistical manuscripts is particularly important, and the precise nature and position of each symbol must be clear. In general, distinction should be made between:

- upper case and lower case letters
- ordinary and bold-faced letters
- certain Greek letters and similar Roman letters
- subscripts, superscripts and 'ordinary' symbols.

Please use a P (upper case, italics) to indicate a P value. Please use n (lower case, italics) to indicate number of subjects.

Statistical symbols are automatically set in italics and need not be underlined except to prevent ambiguity (e.g. when an isolated letter, such as a, occurs in the text). Symbols should not be used to start a sentence.

Back to General instructions

Tables

Tables should be:

- able to be interpreted independently of the text, with meaningful titles, legends and adequate footnotes.
- submitted in editable Word or Excel* format (not pdf);
    - numbered consecutively in Arabic numerals;
    - presented in a separate file or files from that of the main text;
    - able to be interpreted independently of the text, with meaningful titles and adequate footnotes.
- All abbreviations appearing in tables should be spelled out in the footnote.
- Each table should be cited at an appropriate place in the text.
- To ensure accurate reproduction of tables, please use separate table rows for each line of data (including sub-headings). Do not use paragraph returns to separate data items in tables.
- Vertical rules should not be used in tables.

* If tables are submitted in Excel format, please carefully check the pdf proof of the article before submitting the paper to ensure that tables (particularly any in landscape format) are not split across pages, as these are difficult for the editors and reviewers to follow.

Back to General instructions

Figures

- Figures should be able to be interpreted independently of the text, with meaningful titles, legends and adequate footnotes.
- All abbreviations appearing in figures should be spelled out in the footnote.
- Each figure should be submitted separately from the text in an image format such as eps or high-resolution tiff or jpg files (a minimum resolution of 300dpi).
- As figures are published in black and white in the print version of the Journal, please use distinctly different shades, different line types (dashed etc), bar/column types (hatched etc) and different symbols to allow the figure to be interpreted when printed in black and white.
- Please ensure that your electronic image files print clearly at a legible size and are of high quality for publication (We recommend a minimum of 150mm wide at 300dpi for a single-column figure in the print Journal, 300mm wide at 300dpi for full page width figures).
- For graphs, please provide axis labels of a legible size; note that graphics programs often produce large figures with tiny labels which will become illegible when the entire figure is resized for publication.

- Figures and illustrations should be numbered consecutively in Arabic numerals.
- Each figure should be cited at an appropriate place in the text).
- Please state your preferred option (i.e. agreement to pay £350/figure for colour in both print and online, or black and white in print and colour online only at no charge) in your cover letter when submitting your article so the editors can determine whether the supplied images are satisfactory.

Footnotes to tables and figures

- Please use superscript a, b, c, d, etc. as footnote symbols.

Back to General instructions

Copyright permission

- Authors must obtain permission from the copyright holder to reproduce or adapt any previously published material (figures, tables, text, etc.) and be able to produce such permission on request. The permissions obtained should be detailed in your cover letter. Further advice is available on the Oxford Journals Permissions or RightsLink pages. For any questions about permissions, contact jnls.author.support@oup.com.
- If the publisher of previously published material for which you are seeking permission is a member of the International Association of Scientific, Technical and Medical Publishers (STM) — as is the case for most major publishers — permission can usually be granted free of charge via Rightslink. Please contact the editorial office (ije.editorial@sydney.edu.au) for advice on completing your Rightslink application. This is not intuitive and incorrect applications may cause delays and attract unnecessary fees.

Back to General instructions

Declarations

The following declarations, in the following order, should appear at the end of the text and before the References.

Ethics approval

Published research based on human subjects must include a declaration in relation to ethics approval.
Please place, under the subheading "Ethics approval" at the end of your main text immediately before the Funding section, a statement that:

- provides the name of the local ethics committee that approved the study and the approval number; OR
- confirms that such approval is not needed and includes the reason why this is the case; AND/OR
- states how the study conforms to recognised standards such as the Declaration of Helsinki.

Further information on publication ethics.

Author contributions

The contribution of each author and the identity of the author who will act as guarantor for the paper should be provided under the heading 'Author contributions'. The Author contributions statement should be added to the declarations for up to 25 authors.

If there are more than 25 authors, the Author contributions statement should be provided as a Supplementary file designated 'for online publication only' and the text under the Author contributions heading in the paper should read: "Author contributions are available as a Supplementary file at *IJE* online."

The Author contributions statement should be detailed. For example, “AB designed the Study. BC directed the study’s implementation. CD designed the analytical strategy and helped to interpret the findings. DE conducted the literature review and helped to prepare the Introduction and Methods sections of the text. EF drafted the Discussion.”

Data availability

The inclusion of a Data Availability statement is a requirement for articles published in the *International Journal of Epidemiology*. Data Availability statements provide a standardized format for readers to understand the availability of data underlying the research results described in the article. The statement may refer to original data generated in the course of the study or to third-party data analyzed in the article. The statement should describe and provide means of access, where possible, by linking to the data or providing the required unique identifier.

You can find more information and example Data Availability statements on the Research Data Policy page.

Supplementary data

Please add “Supplementary data are available at *IJE* online” under the heading “Supplementary data”.

Funding

The following rules should be followed:

- Any funding used while completing the submitted work should be detailed in a separate Funding section, placed before any Acknowledgements section at the end of the article. If no funding was received, add ‘Funding: None’.
- The sentence on funding should begin: ‘This work was supported by …’.
- The full official name of the funding agency should be given; e.g. ‘the National Cancer Institute at the National Institutes of Health’ or simply ‘National Institutes of Health’, not ‘NCI’ (one of the 27 sub-institutions) or ‘NCI at NIH’. See the full Research Information Network-approved list for details of UK funding agencies. It is important to have this information correct because institution names are used to tag NIH-funded articles so that they are deposited at PubMed Central (PMC). Note that papers are only deposited in PMC once they have been published in an issue of the journal and not after they been published on Advance Access.
- Grant numbers should be complete and accurate and provided in brackets as follows: ‘[grant number ABX CDXXXXXX]’.
- Multiple grant numbers should be separated by a comma as follows: ‘[grant numbers ABX CDXXXXXX, EFX GHXXXXXX]’.
- Funding agencies should be separated by semicolons (plus ‘and’ before the last funding agency).
- Where individual authors need to be specified for certain sources of funding, the following text should be added after the relevant agency or grant number: ‘to [author initials]’.
- Example:
- This work was supported by the National Institutes of Health [P50 CA098252 and CA118790 to R.B] and the Alcohol & Education Research Council [HFY GR667789].
- Crossref Funding Data Registry: To meet funding requirements, authors are required to name their funding sources, or state if there are none, during the submission process. For further information on this process or to find out more about the CHORUS initiative, please see the CHORUS website.

Acknowledgements

This section is optional. It should include consultants or minor contributors involved in the work. It is good practice to ensure that any person named in the Acknowledgements has given consent to be named.

Conflict of interest

A conflict of interest statement must be included at the end of the declarations, immediately before the references and should summarize all aspects of any conflicts of interest declared by authors on the required conflict of interest form. If there is no conflict of interest, the statement, under the sub-heading 'Conflict of Interest', should read 'None declared'.

Back to General instructions

References

- Footnotes are permitted only in tables and figures. They may not be used in the main body of papers. Instead, authors should provide a list of references.
- References should be in Vancouver Style and be listed in the order that they appear in the text and numbered accordingly. These numbers should be inserted in the text in brackets or as superscript numbers whenever a reference is cited.
- The numbered list of references should appear at the end of the article.
- Author details in references should consist of the surnames and initials of all authors when there are six or fewer authors, or the surnames and initials of the first three authors followed by 'et al.' when there are seven or more.
- The author list for each journal reference should be followed by the title of the article, the name of the journal abbreviated in the same way as in PubMed, year, volume, and first and last page numbers, and punctuated as shown below:
- Surname AB, Surname BC. Article title. A*bbreviated journal title* 2000;**2**:100–03.
  Examples:
  Steck N, Junker C, Maessen M, Reisch T, Zwahlen M, Egger M. Suicide assisted by right-to-die associations: a population based cohort study. *Int J Epidemiol* 2014;**431**:614–22.
  Wilbur J, Marquez D, Fogg L et al. The relationship between physical activity and cognition in older Latinos. *J Gerontol B Psychol Soc Sci* 2012;**67**:525–34.
- Titles of books should be followed by the place of publication, the publisher, and the year, as shown below.
  Example:
  McMichael AJ. Planetary overload: global environmental change and the health of the human species. Cambridge: Cambridge University Press, 1993.
- 'Unpublished Observations', 'Personal Communications' and submitted manuscripts should not appear in the reference list. They may be cited in parentheses in the text. Manuscripts accepted for publication may be cited in the reference list as 'In press'; publication details should added on the proof that the authors receive before publication if possible.
- References to websites or web pages should list author(s) (if there are identifiable individuals) or website title, web page title (if applicable), article title, year published, URL and date last accessed.
  Examples:
  Department for Communities and Local Government. The English Indices of Deprivation 2010. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010 (December 2019, date last accessed).
  Auguie B, Antonov A. gridExtra: Miscellaneous Functions for "Grid" Graphics. 2017. https://CRAN.R-project.org/package=gridExtra (12 December 2018, date last accessed).
  Cracknell R. The Ageing Population. House of Commons Library Research. 2010. http://www.parliament.uk/documents/commons/lib/research/key_issues/Key-Issues-The-ageing-population2007.pdf (5 February 2019, date last accessed).

Citing preprints

Preprints must be cited in reference lists and must be identified by using the following formatting:
Surname AB, Surname BC. Article title. Preprint service (e.g. bioXriv, medRxiv), doi:XXXXX, publication date (e.g. 2 July 2017), preprint: not peer reviewed.

Citing data sources

The I*nternational Journal of Epidemiology* supports the Force 11 Data Citation Principles and requires that all publicly available datasets be fully referenced in the reference list with an accession number or unique identifier such as a digital

object identifier (DOI). Data citations should include the minimum information recommended by DataCite: authors, title, publisher (repository name), identifier.

Example Data Citation

- [dataset]* de Cellès, Matthieu Domenech et al. (2018), Data from: Characterizing and comparing the seasonality of influenza-like illnesses and invasive pneumococcal diseases using seasonal waveforms, Dryad, Dataset, https://doi.org/10.5061/dryad.tg5qb

*The inclusion of the [dataset] tag at the beginning of the citation helps us to correctly identify and tag the citation. This tag will be removed from the citation published in the reference list.

Back to General instructions

Appendices

- Please do not submit appendices for publication in print.
- Material essential to the understanding of a paper should be incorporated in the text, or placed in a box that would occupy no more than 1 journal page.
- If placed in a box, please ensure that this box:
    - has a meaningful explanatory title; and
    - is referred to in the main manuscript.
- Additional non-essential material should instead be provided as Supplementary material, for online publication only.

Supplementary material

- Material that is not essential for inclusion in the full text of the manuscript but that would nevertheless benefit the reader can be made available by the publisher as supplementary online-only content, linked to the online manuscript.
- Supplementary material should not be essential to understanding the conclusions of the paper.
- Supplementary material might include more detailed methods, extended datasets/data analysis, additional tables and additional figures (which would be published as presented, including in colour).
- Please number Supplementary tables, figures and boxes with an S prefix (e.g. Supplementary Table S1, Supplementary Figure S1, Supplementary Box S1) and ensure that the Supplementary material is referred to in the main manuscript where necessary.
- All supplementary material must be provided in suitable electronic formats. Permitted formats are: pdf, MS Word (.doc/docx), html, rtf, MS Excel (.xls/xlsx), csv, mp3, mpeg).
- Supplementary material will be published as provided by authors and will not be edited.
- Please provide short titles for each individual file — these will be used to create links to the files from the index page.
- All material to be considered as Supplementary material must be submitted at the same time as the main manuscript for peer review. It cannot be altered or replaced after the paper has been accepted for publication.
- Please spell out at first mention and consistently in footnotes to tables and figures any abbreviations that appear in Supplementary material.

Back to General instructions

Submitting an article to the *IJE*

All manuscripts, including commissioned material and Letters to the Editor, must be submitted online via the ScholarOne manuscript submission website.

Once you have prepared your manuscript according to the instructions outlined here, please visit the submission website. You will need to create an account if you do not have one already.

Further technical guidance on submitting your manuscript online via ScholarOne is available at the Submission Online page.

Back to Contents

The publication process

Copy editing

All accepted manuscripts (excluding supplementary material) are subject to copy editing to apply journal style and correct minor errors. For papers that require extensive English-language editing, it is the author's responsibility to arrange this before the paper can be accepted.

Proofs

The corresponding author will receive a pdf proof of the accepted and copy edited article. Proof correction must not be used as an opportunity to revise the paper. Any essential changes should take up the same amount of space if possible. Alterations, other than corrections of printer's errors, are expensive and may be charged to authors. Corrections should be returned within 3 days to guarantee inclusion.

It is particularly important to check reference and author lists at the proof stage, as this is the final opportunity to catch errors before publication. It is the corresponding author's responsibility to ensure that all authors' names are spelled and presented correctly. Any errors that are not identified at the proof stage are very difficult to correct subsequently, particularly in third-party databases, such as PubMed.

The editors reserve the right to make grammatical and other changes at any stage before publication. These are sometimes necessary so that the paper conforms to the general style of the Journal.

Back to Contents

Supplements

The *IJE* publishes collections of articles on the same topic in sponsored supplements from time to time. Supplement articles must be of the same standard as articles published in the Journal and they are subject to the *IJE*'s usual editorial and peer-review processes.

To decide whether to approve a supplement, the Editor-in-Chief requires a list of planned manuscripts, including abstracts or detailed outlines, as well as the names and details of each manuscript author, including any competing interests. The submitting authors may also recommend a guest editor.

If approved by the Editor-in-Chief, Oxford University Press will provide a quote and a contract.

The *IJE* Editor-in-Chief may choose to have supplements edited in-house or appoint one or more guest editors. The guest editor(s) may assist in commissioning supplement material. Any competing interests of the guest editor(s) must be fully declared to the Editor-in-Chief.

It is possible that not all the papers submitted for a supplement will be accepted for publication. Decisions about which manuscripts to include in a supplement, after the peer-review process, will be made by the Editor-in-Chief in consultation with the guest editor(s).

For further information, please see OUP's Sponsored Supplements website and contact the editorial office if you wish to propose a supplement.

Back to Contents

Types of articles published by the *IJE*

The *IJE* publishes the following types of material. Detailed instructions for each article type can be found by clicking the links.

Please refer also to the general Manuscript Preparation instructions.

- Original Articles
- Cohort Profiles
- Cohort Profile Updates
- Data Resource Profiles
- Software Application Profiles
- Education Corner
- Opinion
- Letters to the Editor
- Editorials
- Commentaries
- Book Reviews
- Obituaries
- Poetry

Original Articles

Overall guidelines*	
Words:	Around 3000, excluding Abstract, Key Messages and References
Tables and figures:	Up to 8 in total
References:	Up to 50
Title:	Short and specific; subtitles may be used
Abstract:	250 words in four sections: Background, Methods (include study period and sample size), Results (include main results with estimates and confidence intervals) and Conclusions
Key Messages:	A summary of the main messages of the paper in 3–5 succinct bullet points, each a complete sentence

* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.

Original articles may be research papers, methods papers, systematic reviews or meta-analyses.

Format

Original articles should be around 3000 words (main text only; excluding Abstract, References, tables and figures).

Please refer also to the general Manuscript Preparation instructions.

Abstracts

Please ensure the article's abstract is included in the main document, as well as being entered into the Abstract field in the ScholarOne website during submission.

The abstract should be no more than 250 words and consist of the following four sections:

Background:

Briefly describe the problem addressed and why it is important.

Methods:

Describe how the study was performed and include:

- the study type
- the study period
- the study location or setting
- the sample size.

Results:

Report the salient results, including the response rate and the main findings. Please include the estimates and confidence intervals for the main results if appropriate.

Please state whether the results are for the main hypothesis or if there has been post hoc analysis or secondary use of data.

Conclusions:

Describe what conclusions can be drawn from the results.

Note that, in the case of methods papers, Methods and Results may be replaced with *Development* and *Application* if this is more appropriate.

Keywords

Three to 10 keywords should be provided at the end of the Abstract.

Key Messages

A Key Messages section should be added after the keywords and before the article's introduction, with the key messages of the paper made in 3-5 succinct bullet points, each a complete sentence.

Observational studies

For observational studies, including case–control, cohort and cross-sectional studies, the *IJE* requires authors to follow the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement and relevant checklist.

Where such studies include routinely collected health data, the *IJE* requires authors to follow the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement and relevant checklist.

Full details of all reporting guidelines are available at the EQUATOR web site.

Systematic reviews and meta-analyses

Note that publishing systematic reviews and meta-analyses is not a high priority for the *IJE*.

We encourage authors of systematic reviews or meta-analyses of randomised trials and other evaluation studies to follow the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement and relevant checklist.

For meta-analyses of observational studies, we encourage authors to follow the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) guidelines. The study protocol, if there is one, should be submitted as a supporting document, clearly labelled 'for reviewing purposes only'.

Full details of all reporting guidelines are available at the EQUATOR web site.

Types of articles published by the *IJE*

Cohort Profiles

Overall guidelines*	
Words:	Around 2500–3000, excluding the Key Features and References
Tables and figures:	Up to 5 in total
References:	Around 30
Title:	Cohort Profile: [your title]
Key Features:	Information as described below; 3–5 succinct bullet points, each a complete sentence

* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.

Purpose

- To make the research community aware of the existence of large cohorts around the world.
- To maximise the use of existing data by fostering collaboration, including sharing data, research questions and methods.

Note that:

- The IJE does not publish individual cohort studies in its Cohort Profile series. Please submit these as Original Articles.
- The aim of the Cohort Profile series is to provide up-to-date information about large population-based cohorts with planned long-term data collection. Cohorts with a narrow clinical focus will, in general, not be considered for this series.
- To warrant publication in the IJE, Cohort Profiles must add substantially to any information about cohorts that is already available on existing websites.
- Profiles that describe a cohort consortium should focus on the added value of the consortium rather than piecemeal descriptions of the constituent cohorts. Such profiles must include at least one publication based on the collated data from the constituent cohorts to illustrate the added value.
- At least one of the authors of a Cohort Profile must be a researcher involved with the data resource presented and be available to readers as a contact person. Please identify this author and provide contact details under the section 'Can I get hold of the data?'.

Checklist of essential criteria*

- ☐ Cohorts must include more than 5000 participants at baseline.
- ☐ Dates of the baseline collection (including start date and end date) must be stated.
- ☐ At least one wave of data must have been collected after baseline data collection.
- ☐ Dates of this and any further data collections must be stated.
- ☐ There must be plans and available funding for further waves of data collection for a period of at least 10 years.
- ☐ Preferably at least one or two papers based on the cohort data must have been published or submitted for publication.
- ☐ A contact person must be nominated (with contact details provided) to provide more information on request.

* Profiles that do not meet these criteria will be rejected.

Cohort Profiles will be published online only. They will still be assigned to an *IJE* issue and listed as ePages in the issue's Table of Contents.

Format

Please refer also to the general Manuscript Preparation instructions.

Title: Begin with 'Cohort Profile:' followed by your title (which is usually the name of the cohort) and then any acronym in parentheses.

Author List: (see General instructions).

Note: Cohort Profiles do not require an Abstract. Please enter the 'Key Features' (see below) into the Abstract field in the ScholarOne website during submission, and select Not Applicable when asked about Key Messages.

Cohort Profile sub-headings

Cohort Profiles MUST use the following sub-headings and address the criteria described under each sub-heading (Cohort Profiles that fail to meet these criteria will not be published):

Key Features

- Key features of the Cohort Profile should be provided as bullet points of no more than 200 words in total. Use complete sentences to provide the following information:
- why the cohort was established and unique feature(s) of the cohort;
- location, year(s) of baseline data collection;
- number of participants at baseline;
- composition of the study population including age range;
- frequency of follow-up;
- attrition;
- number of participants currently in the cohort;
- main categories of data collected; and

- collaboration and data access.

Example of Key Features for a Cohort Profile*

Key Features

- The KiGGS cohort was established in addition to periodically conducted nationwide representative health surveys of children and adolescents aged 0–17 years (KiGGS cross-section) to complement regularly reported trends in prevalence rates among children and adolescents with health development analysis over the life course.
- The first population-based nationwide sample of children and adolescents in Germany (KiGGS Baseline study; ages 0–17 years; n = 8656 girls and 8985 boys) was tracked for the first time using telephone health interviews (KiGGS Wave 1: 2009–2012; n = 6079 female and 5913 male participants; re-participation rate 68%). A total of 10 853 participants of the Baseline study (5790 female, 5063 male) completed questionnaires in the health interview of the second follow-up (KiGGS Wave 2: 2014–2017). Additional examination data are available for 6465 of these re-participants (3254 female, 3211 male).
- Data were collected using questionnaires, physician-administered personal interviews, health examinations and testing, and laboratory analysis. Topics of the KiGGS cohort include numerous physical and mental health indicators, health behaviours, and health care utilization and personal, familial, environmental and socio-economic health determinants.
- Cohort data are available via request with a description of planned projects at Research Data Centre, Robert Koch Institute, Berlin, Germany (e-mail: datennutzung@rki.de).

*Adapted from: Mauz E et al. Cohort Profile: KiGGS cohort longitudinal study on the health of children, adolescents and young adults in Germany. *Int J Epidemiol* epub 2019.

Please place your Key Features at the beginning of your article. (Note that Key Features replace the former Profile in a Nutshell.)

Why was the cohort set up?

Provide the rationale for setting up the cohort (including the original research questions it was set up to address), when it was established, where it is located and how it is funded.

Who is in the cohort?

Describe the:

- cohort design (e.g. random or cluster, population-based);
- how and when the participants were recruited;
- numbers invited and entering the cohort (give the participation rate or response rate); and
- differences between responders and non-responders at baseline (ideally in a table comparing the socio-demographic characteristics of responders and non-responders, or of responders and the general population).

How often have they been followed up?

This section must be sufficiently detailed to provide readers with a clear picture of the population represented by the cohort over time.

Describe:

- details of how often surveys or examinations have been conducted;
- loss to follow-up over time, with summary statistics presented in a table or figure;
- how those lost to follow-up differ from those remaining in the cohort; and
- the results of any work completed describing missing data.

What has been measured?

This section should describe any linkages to morbidity, mortality and other routine data sources.

Give broad categories for each follow-up phase, e.g.:

Phase	Measurements
Baseline 1935-40	Fasting blood samples taken, DNA extracted, fasting glucose, lipids, insulin assayed, serum aliquots stored at –80°C Self-reported socio-economic position Anthropometric measures: weight, height, waist & hip circumference Blood pressure Self-reported major behavioural CVD risk factors Dietary assessments (Questionnaire only) Self-reported major diseases and treatment
Follow-up 1950	Self-reported socio-economic position and behaviours
Ongoing	All participants are flagged with routine data sources providing deaths since baseline and cancer registry entries since 1980

- Provide descriptions of unusual measurements (e.g. specialised scans; unique assays) and/or measurements that have been undertaken in sub-groups of the cohort.

What has it found?

This should not be an exhaustive list, but an indicative summary.

Please:

- describe the most important findings and publications generated by the cohort (provide a web address if there is a complete list of publications online);
- describe what has been found rather than what has been examined; and

- illustrate one or two of the main findings with a table or figure.

What are the main strengths and weaknesses?

Please cover both strengths and weaknesses.

Can I get hold of the data? Where can I find out more?

Note that at least one of the authors of a Cohort Profile must be a researcher involved with the data resource presented and be available to readers as a contact person.

Please:

- provide a web address if the data are open access;
- if access to the data requires permission, indicate where and how to apply;
- if access to the data is more limited, describe opportunities for collaboration; and
- provide the name and contact details of a researcher to whom enquiries can be submitted.

Ethics approval

See the general Manuscript Preparation instructions.

Author contributions

These are now required. See the general Manuscript Preparation instructions.

Supplementary data

If your submission includes these, please add: “Supplementary data are available at IJE online”.

Funding

See the general Manuscript Preparation instructions.

Acknowledgements

This section is optional. See the general Manuscript Preparation instructions.

Conflict of interest

See the general Manuscript Preparation instructions.

References

See the general Manuscript Preparation instructions for details of formatting references.

Types of articles published by the *IJE*

Cohort Profile Updates

Overall guidelines*	
Words:	Around 2000–2500, excluding the Key Features and References
Tables and figures:	Up to 5 in total
References:	Around 30
Title:	Cohort Profile Update: [your title]
Key Features:	Information as described below; 3–5 succinct bullet points, each a complete sentence

* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.

Purpose

To provide readers with up-to-date information on cohorts already published in the *IJE*'s Cohort Profile series when the focus of the research, or the data collected, have changed *substantially*.

Essential criteria

- ☐ Analyses on the new research focus for the cohort must have been completed so that some results can be presented.

or

- ☐ At least one round of data collection that includes the new measures must have been completed so that descriptive statistics for the new data can be presented.

Cohort Profile Updates will be published online only. They will still be assigned to an *IJE* issue and listed as ePages in the issue's Table of Contents.

Format

Please refer also to the general Manuscript Preparation instructions.

Title: Begin with 'Cohort Profile Update:' followed by the cohort name as in the original Cohort Profile.

Author List: (see General instructions).

Note: Cohort Profile Updates do not require an Abstract. Please enter the 'Key Features' (see below) into the Abstract field in the ScholarOne website during submission, and select Not Applicable when asked about Key Messages.

Cohort Profile Update sub-headings

Cohort Profile Updates must use the sub-headings below and address the criteria described under each sub-heading (Cohort Profile Updates that fail to meet these criteria will not be published).

Key Features

Key Features of the Cohort Profile Update should be provided as bullet points of no more than 150 words in total. Points should be written as complete sentences that cover the following information as appropriate:

- a description of the original cohort;
- reason(s) for the change of focus or data collected and new areas of research or research questions;
- the number of participants remaining in the cohort at the last data collection, including their age or age range;
- main categories of new measures; and
- collaboration and data access.

Example of Key Features for a Cohort Profile Update*

Key Features

- The British Regional Heart Study, a cohort of 7765 men aged 40-59, was set up in 1978-80 to investigate regional variations in cardiovascular disease (CVD) mortality across Britain.
- 35 years of follow-up provide opportunities for new research on CVD and related outcomes that affect healthy ageing. New linkages enable examination of associations between early adult life measures and CVD risk, and vascular structure and function in old age.
- The most recent clinical examination in 2010–12 included 1722 men aged 71–92 years.
- In addition to a DNA databank, new measures include physical function, objective physical activity, oral health and quantitative assessments of vascular disease and vascular ageing. New data linkages provide information on comorbidities, area-level exposures and historical data on anthropometry and dental health.
- The British Regional Heart Study has several existing collaborations. For new collaborative projects and enquiries about data sharing please contact Lucy Lennon l.lennon@ucl.ac.uk

*Adapted from: Lennon LT et al. Cohort Profile Update: The British Regional Heart Study 1978-2014: 35 years follow-up of cardiovascular disease and ageing. Int J Epidemiol 2015.

Please place your Key Features at the beginning of your article. (Note that Key Features replace the former Profile in a Nutshell.)

The original cohort

Provide a very brief description of the original cohort, citing the original Cohort Profile.

What is the reason for the new focus (or new data collection)?

Use whichever version of this sub-heading is appropriate for your paper. Why has there been a change of focus? Or why is there a need for a significant change in the data collected?

What will be the new areas of research?

Provide a brief description of your new areas of research.

Who is in the cohort?

Provide the date of the most recent data collection and the current age of the participants. Include details of who remains in the cohort at this stage and how they differ from participants in the cohort at baseline.

What has been measured?

Provide a description or table of your new measures. Where new measures are unusual, please provide brief details of the individual measures. Include descriptions of measures undertaken in sub-groups of the cohort if applicable. This section should describe any new linkages to morbidity, mortality and other routine data sources.

What has it found? Key findings and publications

Describe findings that relate to the new research focus, or provide descriptive statistics and basic associations for the new data collected.

What are the main strengths and weaknesses?

Ensure both strengths and weaknesses are covered.

Can I get hold of the data? Where can I find out more?

The purpose of Cohort Profile Updates is to foster collaboration and maximise the use of existing data. If the data are open access, a web address must be provided. If an application is required to access the data, indicate where the application form can be found and the process for applying. If access to the data is more limited, please describe opportunities for collaboration. In all cases, the name and contact details of a researcher to whom enquiries can be submitted must be provided.

Ethics approval

See the general Manuscript Preparation instructions.

Author contributions

These are now required. See the general Manuscript Preparation instructions.

Supplementary data

If your submission includes these, please add: “Supplementary data are available at IJE online”.

Funding

See the general Manuscript Preparation instructions.

Acknowledgements

This section is optional. See the general Manuscript Preparation instructions.

Conflict of interest

See the general Manuscript Preparation instructions.

References

See the general Manuscript Preparation instructions for details of formatting references.

Overall guidelines*	
Words:	Around 2500–3000, excluding the Key Features and References
Tables and figures:	Up to 5 in total
References:	Around 30
Title:	Data Resource Profile: [your title]
Key Features:	Information as described below; 3–5 succinct bullet points, each a complete sentence

* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.

Purpose

- To make the research community aware of the existence of data resources across the world.
- To provide enough information to enable researchers to understand the scope of the data resource and how to access and make best use of the data.

Note that:

- Submissions for this category include large sets of data from a defined population that could be used for epidemiological research, are available to bona fide researchers and meet the criteria below.
- Preference will be given to datasets that we believe would be of interest to our readership.
- Data collected at global, regional, national or sub-national level may come from cross-sectional surveys, large randomized or quasi-randomized evaluation studies, or bio-resources. Examples include the European Union Statistics on Income and Living Conditions and the Human Mortality Database.
- This category may include cohorts and data resources that do not meet our criteria for Cohort Profiles.

Checklist of essential criteria

Data Resource Profiles **must**

- ☐ Describe a data resource to which data access is:
  - openly available on a website; or
  - able to be granted by the authors or their institutions by means of a straightforward and transparent application process (at least for collaborative access).
- ☐ Have completed baseline collection and have made available at least one additional wave of data (such as a survey or data linkage).
- ☐ Add substantially to any information about the data resource that is already available on existing websites.
- ☐ Include a large baseline sample size (preferably in the tens of thousands).

Data Resource Profiles will be published online only. They will still be assigned to an *IJE* issue and listed as ePages in the issue's Table of Contents.

Format

Please refer also to the general Manuscript Preparation instructions.

Title: Begin with 'Data Resource Profile:' followed by your title (which is usually the name of the database) and then any acronym in parentheses.

Author List: (see General instructions).

Note: Data Resource Profiles do not require an Abstract. Please enter the 'Key Features' (see below) into the Abstract field in the ScholarOne website during submission, and select Not Applicable when asked about Key Messages.

Data Resource Profile sub-headings

Data Resource Profiles must use the following sub-headings and address the criteria described under each sub-heading (Data Resource Profiles that fail to meet these criteria will not be published):

Key Features

Key Features about the Data Resource Profile should be provided as bullet points of no more than 200 words in total. Points should be written as complete sentences that cover the following information:

- why the dataset was established and unique feature(s);
- location, year(s) of baseline data collection, number of participants at baseline, the composition of the study population including age range;
- frequency of follow-up, response rates if repeat cross-sectional surveys, a brief description of any data linkage, the number of participants currently in the dataset;
- main categories of data collected; and
- collaboration and data access.

Example of Key Features for a Data Resource Profile*

Key Features

- The Committee of Nordic Assisted Reproductive Technology and Safety (CoNARTaS) cohort is a Nordic cohort set up from national health registry data to investigate health and safety for children born after assisted reproductive technology (ART) and for their mothers.
- The basis for the CoNARTaS cohort is that in the Nordic countries, individual-level data on all residents are routinely collected by the national health registries and may be linked using the unique national identity number of each resident.
- The CoNARTaS cohort consists of 7 853 958 children, including 172 161 children conceived by ART, as well as 4 130 772 mothers, including 127 317 mothers after ART, identified from the Medical Birth Registries in Denmark (1994–2014), Finland (1990–2014), Norway (1984–2015) and Sweden (1985–2015)
- Data on type of ART treatment, pregnancy and perinatal health, selected diagnosis from specialized health care including detailed information on cancer and causes of death are available for all individuals in the cohort.
- Interested research collaborators can contact the CoNARTaS researchers at www.conartas.com

*Adapted from: Opdahl S et al. Data resource profile: the Committee of Nordic Assisted Reproductive Technology and Safety (CoNARTaS) cohort. *Int J Epidemiol* epub 2019.

Data resource basics

- Provide the rationale for setting up the dataset (including the original research questions that it was set up to address), when it was established, where it is located and how it is funded.
- Define the area covered by the data resource and describe the units, groups and individuals from whom data were collected.
- If data from repeat surveys are available, dates and numbers surveyed, response rates and representativeness should be provided for each repeat survey. This information can be presented in tabular form.
- Provide a brief description of any linkages to other datasets (more detail can be provided in the next section).
- Indicate how long this data resource will be followed up and funded. We prefer data resources that will be followed up for at least 10 years.
- Provide details of ethical clearance (if appropriate) and funding sources.

Data collected

This section should describe, in detail:

- any linkages to morbidity, mortality and other routine data sources, including whether the data have been harmonised;
- measures to ensure data quality and anonymisation techniques.

Include brief details of:

- dataset production, including source data, sampling strategy, software and hardware, if appropriate;
- data type (e.g. secondary data) and survey type (e.g. repeat panel data);

Describe (in tabular form if possible):

- the measures collected at each level, e.g. individual, household, indicating the data collection method (e.g. questionnaire, clinic exam, biospecimen, data linkage),
- the type of data collected (e.g. demographic, anthropometric, genetic, hospital episode), and frequency.

Provide broad categories for the data collected. e.g.:

Phase	Measurements
National cross- sectional survey	Self-reported socio-economic position Self-reported major behavioural risk factors for cardiovascular disease Dietary assessments
Death register	Date of death, cause of death
Cancer registry	Details of cancer — type/grade

Provide descriptions of unusual measurements (e.g. specialised scans; unique assays) and/or measurements that have been undertaken in subgroups. Further examples of tables of data collected in Data Resource Profiles may be found in previously published articles here (Table 3), here (Box 1) and here (Table 1).

Data Resource Use

This should not be an exhaustive list, but an indicative summary.

Please:

- provide one or two brief illustrative examples of how the data resource can be used, or examples of published work which has used this data resource;
- indicate any ongoing specific data analysis plans here to avoid duplication of effort by others interested in the data resource;
- provide an up-to-date citation list for work published on the data resource.

Strengths and Weaknesses

Please cover both strengths and weaknesses.

Data Resource Access

Note that at least one of the authors of a Data Resource Profile must be a researcher involved with the data resource presented and be available to readers as a contact person.

Please:

- provide a web address if the data are open access;
- provide links to the relevant documentation such as variable lists or data dictionaries, where available

- describe variable formats and naming conventions;
- indicate any software required to access the data;
- indicate any digital object identifiers of the data resource here;
- indicate where and how to apply for access to the data if permission is required;
- describe opportunities for collaboration if access to the data is more limited; and
- provide the name and contact details of a researcher to whom enquiries can be submitted.

Ethics approval

See the general Manuscript Preparation instructions.

Author contributions

These are now required. See the general Manuscript Preparation instructions.

Supplementary data

If your submission includes these, please add: "Supplementary data are available at IJE online".

Funding

See the general Manuscript Preparation instructions.

Acknowledgements

This section is optional. See the general Manuscript Preparation instructions.

Conflict of interest

See the general Manuscript Preparation instructions.

Back to Types of articles published by the *IJE*

Software Application Profiles

Overall guidelines*	
Words:	1500-2000, excluding Abstract, Key Features and References
Tables and figures:	2 in total
References:	Up to 25
Title:	Software Application Profile: [your title]
Abstract:	150 words in four sections: Motivation, Implementation, General features and Availability (give URL if available)
General:	Please avoid overly technical jargon
* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.	

Purpose

- To present practical and current software applications for use by the epidemiological research community.

Each profile should adequately describe the motivation and use of the software application, thus enabling the reader to ascertain whether the application is suitable for their research goals.

IJE will consider the following software applications:

- desktop applications;
- web browser-based applications or services;
- server or command line applications; and
- plugins.

Checklist of essential criteria

The software **must**

- ☐ be non-trivial;
- ☐ require limited effort to install;
- ☐ be freely available to non-commercial users; and
- ☐ be made available for a full 3 years after publication.

Additionally:

- ☐ Browser-based applications or services must not require user registration.
- ☐ The software should run under a variety of conditions and, if applicable, must be compatible with multiple operating systems or hardware architectures.

Format

Each profile is required to follow a similar format, using the following sub-headings:

- *Title*: Should include the name of the software application and concisely describe its use.
- *Abstract*: Maximum of 150 words. Use the sub-headings: Motivation, Implementation, General features and Availability (include a URL and any software license requirements).
- *Introduction*: In addition to a concise introduction, please provide motivation for development of the software, as well as its context relative to previously existing software of similar use.
- *Implementation*: Concisely describe how the software is designed and implemented, such as its main modules and how they interact, as well as programming languages, application program interfaces or software libraries used.
- *Use*: A case example, showing substantial functionality of the software and how it can be used in a real-world scenario. If needed, figures and tables can be used to accompany the case example.
- *Conclusion or Discussion*: In addition to a general discussion or conclusions, discuss limitations or future development of the software application.

Back to Types of articles published by the *IJE*

Education Corner

Before preparing a full manuscript, we strongly recommend that you submit a proposal of no more than one page to the editorial office (ije.editorial@sydney.edu.au). Please download a template for an Education Corner article proposal.

Overall guidelines*	
Words:	Around 1500-2500 words (main text, excluding Summary, Key Messages and References)
Tables and figures:	No more than 5 in the main text
References:	No more than 30
Title:	Please include the name of the method(s) or topic(s) being discussed and your goal in choosing this topic (the editors might suggest modifications to titles)
Summary	Unstructured, no more than 250 words
Key Messages:	A summary of your main messages, set out in 3–5 succinct bullet points, each a complete sentence
Supplementary material:	Use supplementary material for example code and data, technical details, and/or other educational elements which cannot be included in the main text

* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.

Purpose

In general, Education Corner articles should be suitable for use in teaching courses in epidemiology, biostatistics and related fields, as well as being useful for continuing professional development. The articles should, therefore, be readily understandable, user-friendly and educationally sound, providing a "recipe book" for readers who wish to apply the concept or method.

The IJE welcomes submissions from non-English-speaking countries. If your mother-tongue is not English, please engage the services of a competent scientific or medical translator from the time you prepare your initial one-page proposal.

Note that:

- These articles will be published as 'Free to View' so that others can, with appropriate attribution, freely read and use them.
- Education Corner articles are indexed and count as regular publications.

Format

Please refer also to the general Manuscript Preparation instructions.

We encourage authors to use a tutorial style, step-by-step format. If you prefer an alternative style, please justify your choice in the initial proposal or, on submission, in a covering letter).

Education Corner papers should:

- provide easily accessible and concise updates and reviews of epidemiological concepts and methods;
- provide introductions to important and emerging methods being used in epidemiology; and
- discuss important insights about methods, including clarifications and examples of misuse.

Please observe that:

- When discussing advanced or complex models, you should provide some background and context. For example, an Education Corner submission on accelerated failure time models might briefly review survival methods and survival functions, in order to provide a basis for more advanced discussion. You may provide further background and context in online supplementary material.

- The methods under discussion should be presented within the context of related methods and empirical tools so that readers can assess alternative methods. A summary table might help to make such comparisons.
- The manuscript should include an illustration or motivating example, such as a recent analysis or setting which provides a real-world introduction to the topic(s). Real or simulated data should be provided, along with the statistical code. This data and code should be made available with the manuscript, either within the text or in the online-only supplementary material. Where applicable, manuscripts should include a practical example of the appropriate interpretation of the proposed concept or method described.
- Equations are welcome, but should be used strategically. If you include equations, you should provide intuitive and non-technical explanations alongside the technical sections.

Opinion

Overall guidelines*	
Words:	Up to 3000, excluding References
Tables and figures:	Up to 4 in total
References:	Up to 30
Title:	Short and specific; subtitles may be used
Abstract/Summary:	Unstructured summary or structured abstract of around 150 words for editorial use only; note that ***this abstract will not be published***
Key Messages:	None

* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.

Purpose

Articles that do not fall into the category of Original Articles, that have not been commissioned as an Editorial or Commentary and that are too long to be acceptable as a Letter to the Editor may be submitted in the Opinion category.

Opinion articles should be well researched and referenced and should be clearly of interest to an epidemiological readership. They should convey novel ideas or insights relating to epidemiology.

Please refer also to the general Manuscript Preparation instructions.

Back to Types of articles published by the *IJE*

Letters to the Editor

Overall guidelines*	
Words:	Up to 700; up to 500 for a letter in response
Tables and figures:	Up to 2 in total; tables should be concise, and any table or figure should illustrate your main point only
References:	Up to 10
Title:	Please use a brief, descriptive title. If your letter refers to an article published in the *IJE*, please include some part of that article's title in your letter title.
Authors:	Up to 3

* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.

The *IJE* welcomes Letters to the Editor. They should be submitted via ScholarOne.

Purpose

Letters to the Editor provide readers with a forum to:

- Comment on articles published in the Journal.
- Introduce preliminary findings and early collaborations.
- Offer opinions on other matters relevant to the Journal's readership.

Note that:

- Preference will be given to letters that are short and to the point.
- The language used must be civil, polite and not excessively emotive — negative personal comments will not be published.
- The Editor reserves the right to limit correspondence about articles published in the Journal to one letter (per correspondent) commenting on that article and one letter (per correspondent) in response from the original authors.
- Letters commenting on articles published in the *IJE must be submitted within 6 months of online publication of the article*.

Format

Letters should provide academic debate of epidemiological concepts and aim to move the field of epidemiology forward.

Critiques of published articles should be evidence-based and well argued.

Please refer also to the general Manuscript Preparation instructions.

Back to Types of articles published by the *IJE*

Editorials [invited only]

Overall guidelines*

Words:	Up to 1500
Tables and figures:	Up to 2
References:	Up to 20
Abstract:	None
Authors:	Up to 4

* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.

Purpose

- Editorials are commissioned to provide informed opinion on 'hot' topics and recent advances in epidemiology, health policy and methodology.
- Editorials do not have to be comprehensive or systematic in their coverage.
- They should give the opinions of the writer.

Note that:

- We commission editorials on specific topics by invitation. Please contact us (at ije.editorial@sydney.edu.au) about your ideas before offering us an unsolicited editorial.
- Unsolicited articles may instead be submitted as an Opinion or, if appropriate, an Original Article.
- Brief comments on published articles should be submitted as a Letter to the Editor.
- Rarely, we may find that a commissioned editorial is not suitable for publication and may reject commissioned work.

Please refer also to the general Manuscript Preparation instructions.

Back to Types of articles published by the *IJE*

Commentaries [invited only]

Overall guidelines*

Words:	Around 1000
Tables and figures:	Up to 2
References:	Up to 10
Abstract:	None
Authors:	Up to 4

* Aspects of your paper may exceed these guidelines, but if they do, please explain in your covering letter why this is necessary.

Purpose

- Commentaries present short perspectives on articles, either recent or historic, published in the *IJE*.
- Commentaries should describe the context within which published papers may be interpreted, and their significance to epidemiology.
- They may comment on aspects of methods used, the findings or their implications.

Note that:

- We commission commentaries on specific topics by invitation. Please contact us (at ije.editorial@sydney.edu.au) about your ideas before offering us an unsolicited commentary.
- The commentary should reference the article/s on which it is commenting.
- Authors may give their own title to their work.
- Unsolicited articles may instead be submitted as an Opinion or, if appropriate, an Original Article.
- Brief comments on published articles should be submitted as a Letter to the Editor.
- Rarely, we may find that a commissioned commentary is not suitable for publication and may reject commissioned work.

Please refer also to the general Manuscript Preparation instructions.

Back to Types of articles published by the *IJE*

Book Reviews

Purpose

- To review new books or substantially enhanced new editions of older books that may be of interest to the epidemiological community.

Note that:

- Book reviews are usually commissioned but we may consider uninvited submissions. Please contact us (at ije.editorial@sydney.edu.au) for advice before submitting an unsolicited book review.

Format

- Book reviews should generally be 500-1500 words long.
- Book reviews should be sensitive to the nature of the *IJE*'s readership and be pitched towards an epidemiological audience.
- Book reviews should go well beyond expanding on the table of contents and should offer a fresh perspective on the substance of the book and its place in the field of epidemiology.

Examples

- Braithwaite J. Growing Inequality: Bridging Complex Systems, Population Health and Health Disparities. *Int J Epidemiol* 2018;47(1):351–353. https://doi.org/10.1093/ije/dyy001
- Woodward A. Climate Change and the People's Health. Sharon *Friel, Int J Epidemiol*. https://doi.org/10.1093/ije/dyz167

Please refer also to the general Manuscript Preparation instructions.

Back to Types of articles published by the *IJE*

Obituaries

The *IJE* publishes obituaries about people of interest to the epidemiological community.

Obituaries are usually commissioned but we will consider uninvited submissions. Please contact us (at ije.editorial@sydney.edu.au) for advice before submitting an unsolicited obituary.

Please provide:

- 500-1000 words; and
- a high-quality image of the person (along with documented permission to use it) if possible.

Please refer also to the general Manuscript Preparation instructions.

Back to Types of articles published by the *IJE*

Poetry

Poetry is usually commissioned but we will consider uninvited submissions.

Please contact us (at ije.editorial@sydney.edu.au), if you wish to discuss whether your poem is appropriate for the Journal.

Please refer also to the general Manuscript Preparation instructions.