Multi-state models for the analysis of Wheeze in a birth cohort of Western Cape children.

Patrick Luke Hannan (HNNPAT002)

BSc (UCT) BSc (Hons) (UCT)

Supervisor: Dr Maia Lesosky

Co-Supervisor: Professor Heather Zar

Dissertation presented for Degree of Master in Public Health (Epidemiology)

School of Public Health and Family Medicine

University of Cape Town

South Africa

September 2019

Declaration

I, Patrick Luke Hannan (HNNPAT002), 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: Patrick Luke Hannan

Signed by candidate

Date: September 2019

Abstract

Introduction

Wheezing is common in young children. By the age of six, approximately 50% of children in high-income countries will have experienced at least one episode of wheezing in their life. Furthermore, childhood wheezing may be associated with reduced lung function and increased risk of asthma in later life. Determining the epidemiology of wheeze is complex given that the risk factors vary based on the age of the child and the phenotype of wheeze. Little is known regarding the recurrent nature of childhood wheezing in low- and middle-income contexts. This study aimed to use multi-state models to estimate the rate of transition among various states of wheeze in children from birth to the age of three years. This study also aimed to investigate the association between possible risk factors for childhood wheezing and the estimated transition intensities.

Methods

The rationale for conducting the study, as well as the objectives of the study, methods and data analysis plan are outlined in the study protocol (Part A). A summary of what is currently known about childhood wheezing is presented as part of the literature review (Part B). The aim of the literature review was to identify known risk factors for childhood wheeze and the methods used to analyse recurrent childhood wheezing, as well as identify the limitations of the current methods used to analyse recurrent childhood wheezing. A manuscript presenting the results of the study is included as Part C.

This study was a secondary analysis of data from 1086 children from birth to three years, born to mothers in the Drakenstein area of the Western Cape, South Africa, enrolled at one of two pri-

mary care clinics. The data were collected as part of a prospective birth cohort, the Drakenstein Child Health Study. Cox proportional hazards models were used to investigate the association of risk factors with time to first wheezing event and time to recurrent wheezing. Two multi-state models investigating the progression of childhood wheezing were constructed. Multiple definitions of childhood wheeze as an outcome were investigated for all constructed models. A simple unidirectional multi-state model and a complex multi-state model with three states (never wheeze, wheeze not associated with lower respiratory tract infection (LRTI), and, lower respiratory tract infection associated wheeze) were constructed. The multi-state model allowed four possible transitions: 1) from "never wheeze" to "wheeze not associated with LRTI" or from 2) "never wheeze" to "LRTI-associated wheeze" or from 3) "wheeze not associated with LRTI" to "LRTI-associated wheeze" and from 4) "LRTI-associated wheeze" to "wheeze not associated with LRTI". Transition intensities between wheeze states were estimated using discrete time multi-state models. The association of risk factors with transition intensities were estimated using multivariable proportional hazards models.

Results

Of the 1086 children included in the study, 476 (44%) experienced at least one episode of wheezing, and 227 (21%) experienced more than one episode of wheezing in the first three years of life. A total of 951 episodes of wheezing were recorded in the 36 months of follow-up time. In the multi-state analysis, LRTI-associated wheeze and wheeze not associated with LRTI were equally likely to be the first wheeze event. However, recurrent wheezing events were more likely to follow LRTI-associated wheeze as the first event (0.0020033 vs 8.6683754×10^{-4}). Male children were at significantly higher risk of experiencing wheeze associated with an LRTI as the first wheezing event and at significantly higher risk of subsequent recurrent wheezing.

Children exposed to maternal smoking prenatally had a significantly higher risk of transition to the wheeze state compared to unexposed children.

Conclusion

Multi-state models provide a novel method for the analysis of wheezing and recurrent wheezing in a cohort of children in South Africa. Multi-state models successfully predicted the progression of children through discrete states of wheeze and produced results consistent with existing literature on childhood wheeze, while accounting for recurrent events and interval-censored data.

Acknowledgements

I would like to thank my supervisor, Dr Maia Lesosky, for the support and feedback provided throughout the MPH process. I would like to acknowledge the contribution of Professor Heather Zar, the co-supervisor of this project.

Contents

List of Figures

List of Tables

Part A: Protocol

Introduction

Wheeze is associated with much of the non-communicable childhood morbidity and mortality in many low- and middle-income countries (LMICs), as well as pneumonia. The primary aim of the study is to make use of multi-state models to estimate the transition probabilities and risk factors associated with the development of wheeze and recurrent wheeze in a cohort of Western Cape children followed for at least five years following birth. Secondary aims of the study are to compare results from traditional survival analysis of wheeze to the results of analysis done using multi-state models.

Background

Globally, respiratory diseases are a major cause of morbidity and mortality, particularly in childhood[1]. It is estimated that respiratory diseases affect more than 400 million people annually around the world[2]. However, although respiratory infections are common globally, severe morbidity and mortality are associated largely with LMICs. Lower respiratory tract infections are responsible for roughly 652 572 (586 475 – 720 612) deaths annually in children under the age of five around the globe[2–4]. It is estimated that 90% of the childhood deaths associated with lower respiratory tract infections occur in LMICs[4,5].

Of particular concern in children in LMICs are wheeze and asthma, given that treatment and care of asthma place a considerable drain on the resources of already strained or resource-poor healthcare systems. Wheeze is defined as a continuous, coarse, whistling sound produced during breathing[6]. Wheezing is caused by restricted respiratory passages, however, diagnosis is challenging given wheezing is a symptom of many related respiratory diseases, such as pneumonia. According to Pearce et al. [7] asthma symptoms, defined as wheezing in the last 12 months,

are prevalent in roughly 11.5% of children aged six to seven around the globe. However, in Africa, the overall prevalence of asthma symptoms in children aged six to seven is estimated to be 5.6%, and the prevalence of asthma symptoms in children aged thirteen to fourteen is estimated to be 13.4%. The estimated prevalence of asthma symptoms in children aged thirteen to fourteen in Cape Town is higher than both the global and African estimates, with 20.3% of children experiencing wheeze[8].

The high burden of wheeze has generated interest in the risk factors associated with wheeze. There is evidence that wheezing is associated with respiratory syncytial virus (RSV) and asthma in children[9,10]. Furthermore, wheezing has been shown to be associated with gender and race in children in the United States, with black male children being at the highest risk of wheezing[11]. However, much of the work on wheeze and its associations has been conducted in high-income countries, it is not known if these risk factors are the same in LMICs. Furthermore, much of the existing work on wheeze has focused on wheeze, or recurrent wheeze as a single outcome, and has not made use of methods of analysis that allow for the incorporation of repeated outcome measures or recurrent events within an individual.

One method of accounting for recurring events in the analysis of longitudinal data is the use of multi-state models. Multi-state models create defined states, such as ”healthy”, ”sick”, and ”recovered/dead” and allow for participants in the study to occupy a given state and the transition to a different state in discrete time intervals. By creating defined states for participants to enter, multi-state models allow for analysis of recurrent events and the risk factors associated with them. Multi-state models have been used successfully in the analysis of HIV in a cohort in which there was incomplete disease history as well as incomplete information with regards to transition times between states[12]. Similarly, multi-state models have been used in the analysis

of asthma control, in order to determine the risk factors associated with optimal asthma control and sub-optimal asthma control[13]. Furthermore, multi-state models are useful in the analysis of datasets containing repeated measures. Multi-state models have been used to deal with repeated measures in the analysis of risk factors associated with survival in cancer patients[14], as well as determining risk factors associated with possible recurrent episodes of colon cancer[15].

Given that wheeze can recur multiple times throughout childhood, as well as the age-related change in risk factors for wheeze, multi-state models are ideal for determining the factors associated with wheeze in a resource-poor health care context. Therefore, this cohort study seeks to estimate the transition probabilities and risk factors associated with the first wheeze event as well as recurrent wheezing events by making use of multi-state models in a resource-poor health care setting in the Western Cape, South Africa.

Methods

Study Design

The study will be a secondary data analysis of data generated by the Drakenstein Child Health Study (DCHS), a population based-birth cohort study investigating the epidemiology and aetiology of childhood respiratory illness and the determinants of child health in a peri-urban area in South Africa[16].

The study will include 1086 mother-child pairs enrolled in the Drakenstein Child Health Study. Wheezing symptoms are expected to occur in roughly 30% of the children participating in the study. The primary outcome of interest will be wheezing symptoms as indicated by case-report forms. Firstly, wheezing was identified by parental report as a positive answer to the ques-

tion “Has your child had a wheezing or whistling noise coming from his/her chest in the past 6 months?”. Secondly, trained healthcare workers recorded if a child had wheezing evident while present at a study or sick visit. Identification of wheezing could occur at planned study visits or at unscheduled visits for LRTI. Recurrent wheezing was defined as two or more episodes of wheezing, either parental report or healthcare worker observed, within a twelve month period.

Characteristics of the Study Population

The data for this study is provided by the parent study, as such data collection for this study has already been completed. The study population consisted of participants enrolled in the parent study (DCHS). The study participants were located in the Drakenstein area in Paarl, an area with a population of roughly 200,000. Members of the Drakenstein community are of low socio-economic status and live in informal housing or crowded conditions. There is a high prevalence of various poverty-related exposures, such as tobacco smoke exposure, alcohol misuse, and malnutrition. More than 90% of the population make use of the freely available public healthcare system. The public health system in the Drakenstein area is comprised of 23 primary care clinics and one hospital, Paarl Hospital, where all births and hospital care occur[16].

Participants in the parent study are children born to mothers in the Drakenstein area that made use of one of the two primary care clinics that serve as enrolment sites and who intended to remain in the area for at least a full year following enrolment. The two primary care clinics serve two different populations, TC Newman (serving a mixed-ancestry population) and Mbekweni (serving a black African population).

The inclusion of pregnant women and mother-child pairs constitutes the inclusion of vulnerable

groups. However, the inclusion of these vulnerable groups is strictly necessary given the high morbidity and mortality caused by respiratory diseases in children. Furthermore, wheezing is primarily diagnosed in young children, and thus can best be studied in young children. It is important to note that the Drakenstein community has both high-levels of poverty related exposures as well as high-levels of access to primary healthcare and might not be generalizable to low-income and middle-income countries with lower levels of access to primary healthcare. However, as countries improve access to health services these results are likely to become widely applicable in low-income and middle-income countries.

Recruitment and Enrolment

The data for this study will be provided by the parent study (DCHS). In the DCHS, pregnant women who had attended either the TC Newman or Mbekweni primary care clinics and who were 20 - 28 weeks gestation age were recruited for the study. Pregnant women who consented to participate in the study completed study questionnaires at scheduled study follow-up visits. Child clinical and respiratory symptoms were completed at each of the study visits, which occurred at birth, 6, 10, and 14 weeks as well as at 6, 12, 18, and 24 months post-delivery at the primary care clinics. All data collected in the DCHS were collected by trained study staff. Staff included research nurses and clinicians at the two primary care clinics. No new individuals will be recruited or enrolled for the current study.

Research Procedures and Data Collection Methods

Given that this study is an analysis of secondary data provided by the DCHS, no participants will be subject to any medical, behavioural or observational interventions. The DCHS staff measured childhood wheezing by parental report as well as active surveillance during lower

respiratory tract infections that required clinic or hospital care. Maternal health and associated risk factors were assessed by means of a standardized questionnaire of relevant demographic and clinical information administered by trained study staff. Lung function in the children was tested at six weeks of age and annually thereafter at Paarl Hospital. All data collection procedures in the parent study have been approved by the UCT Human Research Ethics Committee (HREC), University of Cape Town (401/2009) and the Western Cape Provincial Health Research committee.

Data Safety and Monitoring

Data collected for use in this study, by the parent study (DCHS), does not consist of any identifying information about the participants in the study. However, in order to further protect the information of the participants in the study, all study data will be password-protected.

Data Analysis

Data collected for use in this study, by the parent study (DCHS), consists of standardized questionnaires of demographic and clinical information as well as case-report forms completed by study personnel during planned routine study visits and after discharge from hospital in children admitted for pneumonia and four weeks after an inpatient or ambulatory pneumonia. The data provided for this study will contain no identifying information, all participants are provided with a unique participant-ID that will be used for the purposes of the study.

Initial descriptive analysis of the data will include estimation of medians and proportions as necessary. Cox proportional hazards models will be used to estimate the time to first wheezing event, as well as time to recurrent wheezing[17]. Multi-state Markov models will be used to estimate transition probabilities as well as factors associated with the transition between defined

states. The multi-state models will be created and transition probabilities, as well as risk factors associated with transition, will be determined using the *msm* package[18] in R. All analysis will be done using the open-source statistical programming language R[19].

Description of Risks and Benefits

There are no direct risks to the study participants as this study is an analysis of secondary data generated by the parent study (DCHS). Given the nature of the study, the potential risks for participants involve loss of confidentiality or stigmatization if sensitive information about the participants, provided for use in the study, is made public. Therefore, in order to minimise the potential risks to study participants, the data requested from the parent study will contain no identifying information. Participants will be identified only by the unique participant-ID assigned to them at enrolment into the study.

While this research is not expected to directly benefit the participants of this study, the study is expected to generate information that may contribute to the knowledge about the risk factors associated with wheezing episodes in young children, and may, therefore, allow study participants and the population, in general, to be more informed about possible risk factors that may lead to the development of wheeze in children. Given the minimal risks to the participants and the potential long-term benefits the results of the study may provide to participants, we believe that the study is justified.

Informed Consent Process

Informed consent was obtained in writing at the primary care clinic at the time of enrolment. Mothers were not coerced or influenced in any other way to participate in the study. Informed

consent forms were provided to the mothers in one of three languages of their choosing, either English, Afrikaans or isiXhosa. Informed consent was obtained by study personnel trained to assist the mothers in completing the informed consent forms in the chosen language, and is renewed annually. Consent from each mother in the study was renewed annually.

All the adult participants in the DCHS have the capacity to consent. Given that the infants participating in the study are aged 0 - 24 months, consent for infant participation was sought from the mother of the infant as well as the father if possible. The participants can remove themselves from the study at any time should they feel the need to. Furthermore, participants are informed that they may refuse to participate in any aspect of the study and sub-studies but remain in the study. It is also made clear that should participants refuse to participate in any part of the study it will in no way affect the treatment and care they receive at the primary care clinic that they visit.

Privacy and Confidentiality

The data obtained for this study from the DCHS parent study contains no identifying information, all participants are listed by their unique participant-ID only. Only the study coordinators of the DCHS have access to the full linked study data. In order to maintain privacy in the parent study, all interviews with participants were conducted in private, with only the study participant and the researcher present. All information gathered in the study is stored in a locked filing system at the two primary care clinics serving as study sites. Data is abstracted and stored in a password protected REDCap database, with access restricted to the primary study investigator and study personnel tasked with monitoring the data in the database.

Reimbursement for Participation

Not applicable as the study is an analysis of secondary data.

Emergency care and insurance for research-related injuries

Not applicable as the study is an analysis of secondary data.

Dissemination of Research Findings

The proposed study will be submitted as a mini-dissertation in partial fulfilment of the requirements for a Master of Public Health degree at the University of Cape Town. A manuscript describing the findings of this study will be prepared for submission to a relevant peer-reviewed journal.

References

1. Zar, H. J. and Ferkol, T. W. “The Global Burden of Respiratory Disease-Impact on Child Health”. In: *Pediatric Pulmonology* 49.5 (Mar. 2014), pp. 430–434. ISSN: 87556863. DOI: 10.1002/ppul.23030.
2. Mathers, C., Fat, D. M., Boerma, J. T., and World Health Organization, eds. *The Global Burden of Disease: 2004 Update*. OCLC: ocn264018380. Geneva, Switzerland: World Health Organization, 2008. 146 pp. ISBN: 978-92-4-156371-0.
3. Leowski, J. “Mortality from acute respiratory infections in children under 5 years of age: global estimates”. In: *World Health Statistics Quarterly* 39.2 (Apr. 1986), pp. 138–144. ISSN: 0379-8070.
4. Troeger, C. 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”. In: *The Lancet Infectious Diseases* 18.11 (Nov. 2018), pp. 1191–1210. ISSN: 1473-3099. DOI: 10.1016/S1473-3099(18)30310-4.
5. Nair, H. et al. “Global and Regional Burden of Hospital Admissions for Severe Acute Lower Respiratory Infections in Young Children in 2010: A Systematic Analysis”. In: *The Lancet* 381.9875 (Apr. 2013), pp. 1380–1390. ISSN: 1474-547X (Electronic)\r0140-6736 (Linking). DOI: 10.1016/S0140-6736(12)61901-1.
6. Kliegman, R. M., Behrman, R. E., Jenson, H. B., and Stanton, B. M. D. *Nelson Textbook of Pediatrics*. Elsevier Health Sciences, Aug. 15, 2007. 3200 pp. ISBN: 978-1-4377-2180-5.
7. Pearce, N. et al. “Worldwide Trends in the Prevalence of Asthma Symptoms: Phase III of the International Study of Asthma and Allergies in Childhood (ISAAC)”. In: *Thorax* 62.9 (May 2007), pp. 757–765. ISSN: 0040-6376. DOI: 10.1136/thx.2006.070169.
8. Zar, H. J., Ehrlich, R. I., Workman, L., and Weinberg, E. G. “The Changing Prevalence of Asthma, Allergic Rhinitis and Atopic Eczema in African Adolescents from 1995 to 2002”. In: *Pediatric Allergy and Immunology* 18.7 (Nov. 2007), pp. 560–565. ISSN: 0905-6157, 1399-3038. DOI: 10.1111/j.1399-3038.2007.00554.x.
9. Stein, R. T. et al. “Respiratory Syncytial Virus in Early Life and Risk of Wheeze and Allergy by Age 13 Years”. In: *The Lancet* 354.9178 (Aug. 1999), pp. 541–545. DOI: 10.1016/S0140-6736(98)10321-5.
10. Lee, D. A., Winslow, N. R., Speight, A. N., and Hey, E. N. “Prevalence and Spectrum of Asthma in Childhood.” In: *British Medical Journal* 286.6373 (Apr. 1983), pp. 1256–8. ISSN: 0267-0623 (Print)\r0267-0623 (Linking).

11. Gergen, P. J., Mullally, D. I., Evans, R., and Gergen, J. “National Survey of Prevalence of Asthma among Children in the United States, 1976 to 1980”. In: *Pediatrics* 81.1 (Jan. 1988), pp. 1–7.
12. Gentleman, R. C., Lawless, J. F., Lindsey, J. C., and Yan, P. “Multi-state Markov Models for Analysing Incomplete Disease History Data with Illustrations for Hiv Disease”. In: *Statistics in Medicine* 13.8 (Apr. 1994), pp. 805–821. ISSN: 0277-6715 (Print)\r0277-6715 (Linking). DOI: 10.1002/sim.4780130803.
13. Saint-Pierre, P., Combescure, C., Daurès, J., and Godard, P. “The Analysis of Asthma Control under a Markov Assumption with Use of Covariates”. In: *Statistics in Medicine* 22.24 (Dec. 2003), pp. 3755–3770. ISSN: 0277-6715\r1097-0258. DOI: 10.1002/sim.1680.
14. Kay, R. “A Markov Model for Analysing Cancer Markers and Disease States in Survival Studies”. In: *Biometrics* 42.4 (Dec. 1986), pp. 855–855. ISSN: 0006-341X (Print). DOI: 10.2307/2530699.
15. Conlon, A. S., Taylor, J. M., and Sargent, D. J. “Multi-State Models for Colon Cancer Recurrence and Death with a Cured Fraction”. In: *Statistics in Medicine* 33.10 (Dec. 2013), pp. 1750–1766. ISSN: 0277-6715. DOI: 10.1002/sim.6056.
16. Zar, H. J., Barnett, W., Myer, L., Stein, D. J., and Nicol, M. P. “Investigating the Early-Life Determinants of Illness in Africa: The Drakenstein Child Health Study”. In: *Thorax* 70.6 (June 2015), pp. 592–594. ISSN: 0040-6376, 1468-3296. DOI: 10.1136/thoraxjnl-2014-206242.
17. Cox, D. R. “Regression Models and Life-Tables”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2 (Mar. 1972), pp. 187–202. ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1972.tb00899.x.
18. Jackson, C. H. “Multi-State Models for Panel Data: The Msm Package for R”. In: *Journal of Statistical Software* 38.8 (Apr. 2011). ISSN: 1548-7660. DOI: 10.18637/jss.v038.i08.
19. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, July 2018.

Part B: Literature Review

Literature Review Objectives

The objectives of this structured literature review are as follows: to present literature on the known risk factors for childhood wheezing, recurrent childhood wheezing, and childhood wheezing associated with respiratory tract infections. Furthermore, the review aims to examine the statistical methods used for the analysis of longitudinal data and repeated measures with regards to childhood wheezing. Finally, this review will discuss the limitations of the current literature with respect to childhood wheezing and the need for further research.

Search Strategy

Initial searches were conducted using PubMed, Google Scholar and Scopus. Search terms used were: "(child OR childhood) AND (wheeze OR wheezing OR wheezing symptoms OR wheezing illness) AND (risk factor)". Further information on recurrent and severe childhood wheezing was searched for in the same databases using the following search terms: "(child OR childhood) AND (wheeze OR wheezing OR wheezing symptoms OR wheezing illness) AND (severe OR recurrent OR persistent)". Additional articles were identified from reference lists and bibliographies of articles extracted by the initial search.

Only studies conducted in English were considered for the structured literature review. The articles included in the review were not limited to a particular timeframe, although any articles that did not include childhood wheezing as an outcome were excluded. Furthermore, all study designs were considered, however, prospective longitudinal study designs were of particular importance. One study could not be obtained from the databases or contact with the authors and was therefore excluded, all other studies found were considered for the summary of the literature.

Summary of the Literature

The Epidemiology of Wheeze

Determining the epidemiology and risk factors for childhood wheezing is complex, given that wheezing is a symptom of many related respiratory diseases such as asthma, pneumonia and other respiratory infections[1]. Another challenge in determining the risk factors associated with wheezing is the lack of standardisation in the definition of wheezing as an outcome. There is currently no agreed upon definition of wheezing used for research, and definitions vary by study. The most common definition of wheezing is the definition used by the International Study of Asthma and Allergies in Childhood (ISAAC) which defines current wheezing as: "wheezing or whistling in the chest in the last 12 months?"[2]. Studies that use the ISAAC definition of current wheezing are known to report significantly higher prevalences for wheezing than non-ISAAC studies[3], this is likely due to the ISAAC definition of current wheezing including any wheezing recalled in the preceding 12 months of the study.

Wheezing is common in both adults and children, by six years of age approximately 50% of children in high-income countries will have experienced at least one episode of wheezing in their life[4]. The regional point prevalence of childhood wheezing varies greatly based on the country and age group that was surveyed, from 2.4% in Jodhpur, India to 37.6% in Costa Rica amongst six to seven-year-old children and from 0.8% in Tibet, China to 32.6% in Wellington, New Zealand in thirteen to fourteen-year-old children[5]. Although the prevalence of wheezing is similar in adults and children, the medical burden of wheeze disproportionately affects young children.

The overall global impact of wheeze is difficult to determine given that wheezing is associated

with respiratory morbidity in children but is also closely linked to reduced lung function and the development of asthma in later life[6,7]. Wheezing negatively impacts health in both the short and long-term, it is also a considerable financial burden. Not only do the families of children have to pay for care for each individual episode of wheezing, but additional costs such as time and wage loss when caring for the sick child are also significant contributors to the financial burden of wheezing. It is estimated that the economic cost of childhood wheezing for a single year in the United Kingdom was 53 million GBP in 1998, roughly 90 million GBP per year when adjusted for inflation[8].

Childhood Wheeze Phenotypes

Along with the difficulty in measuring and classifying wheezing as an outcome, determining the risk factors for wheezing is complicated due to the fact that wheezing in young children is not a single disorder but a heterogeneous condition with distinct phenotypes and risk factors that vary as children age. Although children may wheeze for a variety of different reasons at any given age, the majority of early childhood wheezing is associated with respiratory tract infections (RTIs), and late-onset wheezing and asthma associated with allergy and atopic disease. Both genetic and environmental risk factors play critical roles in the development of childhood wheeze.

Seminal early work on wheezing and asthma in young children in the Tuscon Children's Respiratory Study, a birth cohort in the United States, found that children with wheezing can be broadly classified into four heterogeneous wheeze phenotypes based on the age at which the child first wheezed as well as whether or not the wheezing persisted throughout early childhood[4]. In this description children are classified as those with no wheeze at three and six years of life; those that have at least one episode of wheezing in the first three years of life but not

at six years (transient wheeze); those who had no episodes of wheezing in the first three years of life but who had wheeze at by the age of six (late-onset wheeze) and those who had at least one episode of wheezing in the first three years of life and had wheezing at six years of age (persistent wheeze)[4].

The childhood wheeze phenotypes described by Martinez, Wright, et al. [4] were further expanded upon in a study by Lodge, Zaloumis, et al. [9]. This study followed 620 children at high-risk for allergy from birth to the age of seven, with a final follow-up at twelve years of age. The study used latent class analysis (LCA) to detect subgroups of wheezing in the children. The study found that the children could be classified into one of five wheeze phenotypes: never/infrequent wheeze, early transient wheeze, early persistent wheeze, intermediate-onset wheeze, and late-onset wheeze. Early transient wheeze appeared in the first twelve months of life and resolved permanently at around three years. Early persistent wheeze appeared in the first 6 months of life, intermediate-onset wheeze emerged at around eighteen months, and late-onset wheeze occurred after the age of four.

Given the above phenotypes for wheeze and the periods in a child's life when they are most prevalent, childhood wheezing can be broadly described as occurring in three distinct age ranges with risk factors that vary based on the age: the first three years of childhood; middle and late childhood, from the age of four to ten, and early adolescence.

Wheeze in the First Three years of Life

The majority of children who experience wheezing in the first three years of life are classified as transient wheezers, as they do not experience wheezing after the first three years of life[4]. The Tucson Children's Respiratory Study[10], a study of 826 new-born infants in the United States

found that one-third of all children under three years of age had wheezed in conjunction with a lower respiratory tract illness. However, 60% of those children had stopped wheezing by the age of six[4]. Similarly, in Australia, 73% of children who wheezed in the first six months of life did not wheeze three years later[11].

Children with transient wheeze are believed to have smaller airways than children who never wheeze, and it is the reduced airway calibre that predisposes children to obstructive respiratory tract infections and wheezing during early childhood[12]. As the children age, the predisposition for wheezing is mitigated by the increase in airway size of older children. The majority of children experience transient wheezing in early life and are then considered symptom-free in later life.

For example, young children with respiratory syncytial virus (RSV) were 3.2 times more likely to have zero to three episodes of wheezing in the past year and 4.3 times more likely to have more than three episodes of wheezing in the past year when compared to children with no RSV. However, as airway size increased with age the risk of wheezing decreased and by age thirteen there was no significant difference in the risk of wheezing between children with RSV and children without RSV in the first three years of life[7,13]. The association between wheezing and respiratory tract infections was not unique to RSV, although RSV is the most commonly diagnosed respiratory tract infection. The presence of any lower respiratory tract infection (LRTI) was associated with a three-fold increase in the risk of transient wheezing. Furthermore, the risk of wheezing was 3.7 times higher in children in the lowest tertile of total respiratory conductance (a measure of overall respiratory function) prior to the first episode of wheezing[14], indicating that diminished lung function may not be a consequence of respiratory infections but was present prior to any respiratory tract infections.

Given that transient wheezing is closely associated with respiratory tract infections, many of the risk factors for respiratory infection in young children are also risk factors for transient wheezing. For example, there is a well-documented association between respiratory tract infections and exposure to other children, either siblings or at a day-care, in early life[15–17]. Similarly, exposure to other children is also a risk factor for transient wheeze[17–19]. Other environmental risk factors for respiratory infections and transient wheezing include the season of birth[11,20,21] and length and type of breastfeeding[11,20,22]. Exposure to maternal smoking is also a critical risk factor in the development of transient wheezing and many studies of childhood wheeze that have included it as a covariate have found that it significantly increases the risk of developing wheezing[4,11,21,23–25].

There is also evidence that genetic factors may predispose young children to respiratory infections and transient wheeze. Initially, male children in the United States were believed to have a higher risk of wheezing in the first three years of life[21], however later work on children of the same age has found no increase in transient wheezing amongst males, sex only increased the risk of late-onset and persistent wheezing[4]. Furthermore, another study that investigated sex as a risk factor found no significant difference in the risk of wheezing between the sexes, however, the study included only 132 infants with wheeze and thus may be underpowered for the analysis that was conducted[26].

Wheeze in Middle and Late Childhood

Although the majority of children wheeze only during early childhood and never again, a small subset of children wheeze persistently throughout childhood. Similarly, there is a separate subset of children who do not experience wheezing in the first three years of life but develop wheezing after the age of three, this wheeze phenotype is known as late-onset wheeze. The majority of

children who wheeze after the first three years of life are classified as having either late-onset or persistent wheeze, by the age of ten persistent wheezing is the most common wheeze phenotype. Although the two wheeze phenotypes are distinct they share many of the same risk factors. A combination of genetic factors and exposure to adverse events postnatally predispose children to both persistent and late-onset wheezing in middle childhood and beyond[11].

The differences in lung function between transient wheeze and other wheeze phenotypes persist beyond early childhood. Lodge, Lowe, et al. [27] demonstrated that all childhood wheeze phenotypes, except transient wheeze, were associated with lower lung function growth throughout early childhood and continuing into late adolescence when compared to children who had never wheezed. Furthermore, all wheeze phenotypes, except transient wheezing, were associated with an increased risk of current wheezing at twelve years of age in children in the United States[9].

Many of the risk factors for late-onset and persistent wheezing are related to allergic sensitization and atopic disease. Asthma and other atopic diseases are known to commonly co-occur in both adults and children[28], and are also key risk factors for the development of persistent or late-onset wheezing in middle and late childhood[26,29–31]. For example, children born to mothers with atopy were almost twice as likely to suffer from persistent wheezing than those born to mothers without atopy, if the mother had both asthma and eczema the child was almost three times as likely to develop persistent wheezing before the age of seven[25].

Parental asthma is also a strong risk factor for the development of wheezing[32,33]. However, when the risk factors for each wheeze phenotype were considered independently maternal asthma was only found to significantly increase the risk of persistent and late-onset wheezing and not transient wheezing[4]. Work done on wheezing since has strengthened the evidence for a strong as-

sociation between maternal asthma and various forms of childhood wheeze, regardless of the cultural setting of the study or participant socio-economic status[11,20,23,25,31,34,35]. Only one of the studies classified wheezing into separate phenotypes. The study found that the proportion of children who wheeze persistently and had mothers with a history of asthma was double the proportion of those who wheeze persistently but did not have a maternal history of asthma. It is difficult to estimate the effect of maternal and paternal asthma independently, as much of the work on childhood wheezing either measures maternal asthma alone or creates a composite measure of "parental history of asthma". However, the study found that maternal, but not paternal asthma was associated with an increased risk of persistent wheezing. However, asthma in both parents has been associated with an increased risk of late-onset wheezing[11].

The association between childhood asthma and allergic sensitisation has not been as well studied as parental atopy, given the difficulty of assessing whether a young child is allergic or not. However, childhood asthma and atopy have been consistently implicated as important predictors of late-onset or persistent wheezing when they have been studied[4,32,36,37]. The prevalence of asthma in late childhood and adulthood is strongly associated with serum IgE levels[38]. However, one study found no association between serum IgE levels at birth and wheezing shortly after birth[39]. Although the results appear contradictory, Martinez, Wright, et al. [4] found that total IgE at birth was not predictive of wheeze phenotype, but by nine months, persistent wheezers had significantly higher IgE compared to the other wheeze phenotypes. This suggests that those with persistent wheezing may be genetically predisposed to developing a different immunologic response to exposures and events in the first few months of life.

Initially, it was believed that LRTIs early in childhood were associated with an increased risk of developing respiratory symptoms, including wheezing, in later childhood and early adoles-

cence[40]. It was believed that LRTIs predisposed children to asthma by damaging their airways in early life. However, the literature on the relationship between LRTIs in early life and development of respiratory symptoms in later life is contradictory[41]. Stein et al. [13] investigated the relationship between RSV infection in the first three years of life and wheezing at the age of thirteen and found that RSV infection by the age of three was associated with an increased risk of wheezing at age six. However, by the age of thirteen, there was no association between RSV infection in the first three years of life and wheezing. Furthermore, there was no association between RSV infection and atopy later in life. Later research suggests that RSV infection may be associated with recurrent wheezing in later life however more research is needed[42].

The epidemiology of late-onset wheeze is made more complicated by the bimodal effect of some of the risk factors for wheezing. For example, exposure to other children in early life is associated with an increased risk of respiratory infection and wheezing[17–19], however, exposure to other children in early life was also found to be associated with lower levels of asthma in thirteen-year-old children[18]. It is thought that exposure to microbes in early childhood may provide protection against atopic disease in later life.

Need for Further Research

Many studies have investigated the epidemiology of childhood wheezing, however, it is important to note the limitations of the current literature. Although some work on wheeze has been conducted in Brazil and Ecuador[43–45], the majority of the literature on childhood wheezing has been conducted in high-income countries (HIC), there is a lack of studies investigating childhood wheezing in low and middle-income countries (LMIC) particularly in an African context. Africa and South Africa are of interest given the ever-increasing prevalence of asthma and wheezing

in South African children[46].

Childhood wheezing is a fundamentally longitudinal process, with risk factors that can vary dramatically based on the age of the child, and the individual child's disease history. Although much of the literature on childhood wheezing utilises longitudinal data from prospective birth cohorts, the methods used to analyse the data generally only make use of a cross-sectional snapshot of the data. The most common method of analysis for childhood wheezing in the literature was a multivariable logistic regression model of the last year of the study only (current wheezing defined as any wheezing in the past twelve months), only two of the studies included in this review utilised methods that accounted for the repeated measures of longitudinal data[18,24]. By using only the final twelve months of any given study much of the data collected in the prospective birth cohort is ignored. Furthermore, by using only a snapshot of the data studies may not capture information about risk factors that have varied as the child ages. This effect is of particular concern when young children are the focus of the study, given the rapid growth and change in risk factors that occur in young children. This may lead to misclassification of the child's wheeze phenotype and their potential risk factors. Further research that takes into account the longitudinal nature of wheeze progression and the recurrent measures inherent in such data sources is necessary to provide an accurate assessment of the epidemiology of wheezing in early childhood.

References

1. Kliegman, R. M., Behrman, R. E., Jenson, H. B., and Stanton, B. M. D. *Nelson Textbook of Pediatrics*. Elsevier Health Sciences, Aug. 15, 2007. 3200 pp. ISBN: 978-1-4377-2180-5.
2. Asher, M. I. and Weiland, S. K. "The International Study of Asthma and Allergies in Childhood (ISAAC)". In: *Clinical and Experimental Allergy* 28 Suppl 5 (Nov. 1998), 52–66, discussion 90–1. ISSN: 0954-7894. DOI: 10.1046/j.1365-2222.1998.028s5052.x.
3. Patel, S. P., Järvelin, M.-R., and Little, M. P. "Systematic Review of Worldwide Variations of the Prevalence of Wheezing Symptoms in Children". In: *Environmental Health* 7.1 (Nov. 2008), p. 57. ISSN: 1476-069X. DOI: 10.1186/1476-069X-7-57.
4. Martinez, F. D., Wright, A. L., Taussig, L. M., Holberg, C. J., Halonen, M., and Morgan, W. J. "Asthma and Wheezing in the First Six Years of Life". In: *New England Journal of Medicine* 332.3 (Jan. 1995), pp. 133–138. DOI: 10.1056/NEJM199501193320301.
5. Lai, C. K. W., Beasley, R., Crane, J., Foliaki, S., Shah, J., and Weiland, S. "Global Variation in the Prevalence and Severity of Asthma Symptoms: Phase Three of the International Study of Asthma and Allergies in Childhood (ISAAC)". In: *Thorax* 64.6 (June 2009), pp. 476–483. ISSN: 3906770559. DOI: 10.1136/thx.2008.106609.
6. Hovland, V., Riiser, A., Mowinckel, P., Carlsen, K.-H., and Carlsen, K. C. L. "The Significance of Early Recurrent Wheeze for Asthma Outcomes in Late Childhood". In: *European Respiratory Journal* 41.4 (Apr. 2013), pp. 838–845. ISSN: 0903-1936, 1399-3003. DOI: 10.1183/09031936.00071512.
7. Zomer-Kooijker, K., van der Ent, C. K., Ermers, M. J. J., Uiterwaal, C. S. P. M., Rovers, M. M., and Bont, L. J. "Increased Risk of Wheeze and Decreased Lung Function after Respiratory Syncytial Virus Infection." In: *PLOS One* 9.1 (Jan. 2014), e87162. ISSN: 1932-6203 1932-6203. DOI: 10.1371/journal.pone.0087162.
8. Stevens, C. A., Turner, D., Kuehni, C. E., Couriel, J. M., and Silverman, M. "The Economic Impact of Preschool Asthma and Wheeze". In: *European Respiratory Journal* 21.6 (June 2003), pp. 1000–1006. ISSN: 0903-1936, 1399-3003. DOI: 10.1183/09031936.03.00057002.
9. Lodge, C. J., Zaloumis, S., et al. "Early-Life Risk Factors for Childhood Wheeze Phenotypes in a High-Risk Birth Cohort." In: *The Journal of Pediatrics* 164.2 (Feb. 2014), 289-294.e1–2. ISSN: 1097-6833 0022-3476. DOI: 10.1016/j.jpeds.2013.09.056.

10. Taussig, L. M., Wright, A. L., Morgan, W. J., Harrison, H. R., and Ray, C. G. “The Tucson Children’s Respiratory Study I. Design and Implementation of a Prospective Study of Acute and Chronic Respiratory Illness in Children”. In: *American Journal of Epidemiology* 129.6 (June 1989), pp. 1219–1231. ISSN: 0002-9262. DOI: 10.1093/oxfordjournals.aje.a115242.
11. Sherriff, A., Peters, T. J., Henderson, J., and Strachan, D. “Risk Factor Associations with Wheezing Patterns in Children Followed Longitudinally from Birth to 3(1/2) Years.” In: *International journal of epidemiology* 30.6 (Dec. 2001), pp. 1473–1484. ISSN: 0300-5771 0300-5771. DOI: 10.1093/ije/30.6.1473.
12. Martinez, F. D. “What Have We Learned from the Tucson Children’s Respiratory Study?” In: *Paediatric Respiratory Reviews* 3.3 (Sept. 2002), pp. 193–197. ISSN: 1526-0542. DOI: 10.1016/S1526-0542(02)00188-4.
13. Stein, R. T. et al. “Respiratory Syncytial Virus in Early Life and Risk of Wheeze and Allergy by Age 13 Years”. In: *The Lancet* 354.9178 (Aug. 1999), pp. 541–545. DOI: 10.1016/S0140-6736(98)10321-5.
14. Martinez, F. D., Morgan, W. J., Wright, A. L., Holberg, C. J., Taussig, L. M., and The Group Health Medical Associates’ Personnel*. “Diminished Lung Function as a Predisposing Factor for Wheezing Respiratory Illness in Infants”. In: *New England Journal of Medicine* 319.17 (Oct. 1988), pp. 1112–1117. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJM198810273191702.
15. Celedon, J. C., Litonjua, A. A., Weiss, S. T., and Gold, D. R. “Day Care Attendance in the First Year of Life and Illnesses of the Upper and Lower Respiratory Tract in Children With a Familial History of Atopy”. In: *Pediatrics* 104.3 (Sept. 1999), pp. 495–500. ISSN: 0031-4005, 1098-4275. DOI: 10.1542/peds.104.3.495.
16. Fleming, D. W., Cochi, S. L., Hightower, A. W., and Broome, C. V. “Childhood Upper Respiratory Tract Infections: To What Degree Is Incidence Affected by Day-Care Attendance?” In: *Pediatrics* 79.1 (Jan. 1987), pp. 55–60. ISSN: 0031-4005, 1098-4275.
17. Marbury, M. C., Maldonado, G., and Waller, L. “Lower Respiratory Illness, Recurrent Wheezing, and Day Care Attendance.” In: *American Journal of Respiratory and Critical Care Medicine* 155.1 (Jan. 1997), pp. 156–161. ISSN: 1073-449X 1073-449X. DOI: 10.1164/ajrccm.155.1.9001305.
18. Ball, T. M., Castro-Rodriguez, J. A., Griffith, K. A., Holberg, C. J., Martinez, F. D., and Wright, A. L. “Siblings, Day-Care Attendance, and the Risk of Asthma and Wheezing during Childhood.” In: *The New England Journal of Medicine* 343.8 (Aug. 2000), pp. 538–543. ISSN: 0028-4793 0028-4793. DOI: 10.1056/NEJM200008243430803.
19. Rusconi, F. et al. “Risk Factors for Early, Persistent, and Late-Onset Wheezing in Young Children.” In: *American Journal of Respiratory and Critical Care Medicine* 160 (5 Pt 1

Nov. 1999), pp. 1617–1622. ISSN: 1073-449X 1073-449X. DOI: 10.1164/ajrccm.160.5.9811002.

20. De Vries, A., Reynolds, R. M., Seckl, J. R., van der Wal, M., Bonsel, G. J., and Vrijkotte, T. G. M. "Increased Maternal BMI Is Associated with Infant Wheezing in Early Life: A Prospective Cohort Study." In: *Journal of Developmental Origins of Health and Disease* 5.5 (Oct. 2014), pp. 351–360. ISSN: 2040-1752 2040-1744. DOI: 10.1017/S2040174414000312.
21. Bisgaard, H., Dalgaard, P., and Nyboe, J. "Risk Factors for Wheezing during Infancy. A Study of 5,953 Infants." In: *Acta Paediatrica Scandinavica* 76.5 (Sept. 1987), pp. 719–726. ISSN: 0001-656X 0001-656X. DOI: 10.1111/j.1651-2227.1987.tb10555.x.
22. Silva, J. M. et al. "A Prospective Study of Wheezing in Young Children: The Independent Effects of Cockroach Exposure, Breast-Feeding and Allergic Sensitization." In: *Pediatric Allergy and Immunology* 16.5 (Aug. 2005), pp. 393–401. ISSN: 0905-6157 0905-6157. DOI: 10.1111/j.1399-3038.2005.00308.x.
23. Hallit, S. et al. "Wheezing Phenotypes and Risk Factors in Early Life: The ELFE Cohort." In: *PLOS One* 13.4 (May 2018). ISSN: 1932-6203 1932-6203. DOI: 10.1371/journal.pone.0196711.
24. Sahiner, U. M. et al. "Recurrent Wheezing in the First Three Years of Life: Short-Term Prognosis and Risk Factors." In: *The Journal of Asthma* 50.4 (May 2013), pp. 370–375. ISSN: 1532-4303 0277-0903. DOI: 10.3109/02770903.2013.770013.
25. Taylor-Robinson, D. C., Pearce, A., Whitehead, M., Smyth, R., and Law, C. "Social Inequalities in Wheezing in Children: Findings from the UK Millennium Cohort Study." In: *The European respiratory journal* 47.3 (Mar. 2016), pp. 818–828. ISSN: 1399-3003 0903-1936. DOI: 10.1183/13993003.01117-2015.
26. Camara, A. A. et al. "Risk Factors for Wheezing in a Subtropical Environment: Role of Respiratory Viruses and Allergen Sensitization." In: *The Journal of allergy and clinical immunology* 113.3 (Mar. 2004), pp. 551–557. ISSN: 0091-6749 0091-6749. DOI: 10.1016/j.jaci.2003.11.027.
27. Lodge, C. J., Lowe, A. J., et al. "Childhood Wheeze Phenotypes Show Less than Expected Growth in FEV1 across Adolescence." In: *American Journal of Respiratory and Critical Care Medicine* 189.11 (June 2014), pp. 1351–1358. ISSN: 1535-4970 1073-449X. DOI: 10.1164/rccm.201308-1487OC.
28. Lack, G. "Pediatric Allergic Rhinitis and Comorbid Disorders". In: *Journal of Allergy and Clinical Immunology* 108 (1, Supplement July 2001), S9–S15. ISSN: 0091-6749. DOI: 10.1067/mai.2001.115562.

29. Abramidze, T., Gotua, M., Rukhadze, M., and Gamkrelidze, A. “Risk Factors of Asthma in Georgian Schoolchildren.” In: *Georgian Medical News* 213 (Dec. 2012), pp. 48–51. ISSN: 1512-0112.
30. Bjerg, A., Sandstrom, T., Lundback, B., and Ronmark, E. “Time Trends in Asthma and Wheeze in Swedish Children 1996-2006: Prevalence and Risk Factors by Sex.” In: *Allergy* 65.1 (Jan. 2010), pp. 48–55. ISSN: 1398-9995 0105-4538. DOI: 10.1111/j.1398-9995.2009.02105.x.
31. Hryhorczuk, D. et al. “Risk Factors for Wheezing in Ukrainian Children: Ukraine European Longitudinal Study of Pregnancy and Childhood Group.” In: *Paediatric and Perinatal Epidemiology* 23.4 (July 2009), pp. 346–351. ISSN: 1365-3016 0269-5022. DOI: 10.1111/j.1365-3016.2009.01026.x.
32. Ehrlich, R. I. et al. “Risk Factors for Childhood Asthma and Wheezing. Importance of Maternal and Household Smoking.” In: *American Journal of Respiratory and Critical Care Medicine* 154 (3 Pt 1 Sept. 1996), pp. 681–688. ISSN: 1073-449X 1073-449X. DOI: 10.1164/ajrccm.154.3.8810605.
33. Xuan, W. et al. “Risk Factors for Onset and Remission of Atopy, Wheeze, and Airway Hyperresponsiveness.” In: *Thorax* 57.2 (Feb. 2002), pp. 104–109. ISSN: 0040-6376 0040-6376. DOI: 10.1136/thorax.57.2.104.
34. Quah, P. L., Huang, C.-H., Shek, L. P.-C., Chua, K. Y., Lee, B. W., and Kuo, I.-C. “Hyper-Responsive T-Cell Cytokine Profile in Association with Development of Early Childhood Wheeze but Not Eczema at 2 Years.” In: *Asian Pacific Journal of Allergy and Immunology* 32.1 (Mar. 2014), pp. 84–92. ISSN: 0125-877X 0125-877X. DOI: 10.12932/AP0355.32.1.2014.
35. Bernstein, D. I. “Traffic-Related Pollutants and Wheezing in Children.” In: *The Journal of Asthma* 49.1 (Feb. 2012), pp. 5–7. ISSN: 1532-4303 0277-0903. DOI: 10.3109/02770903.2011.641049.
36. De Sousa, R. B., Medeiros, D., Sarinho, E., Rizzo, J. A., Silva, A. R., and Bianca, A. C. D. “Risk factors for recurrent wheezing in infants: a case-control study.” In: *Revista de saude publica* 50 (May 2016), p. 15. ISSN: 1518-8787 0034-8910. DOI: 10.1590/S1518-8787.2016050005100.
37. Strachan, D. P., Butland, B. K., and Anderson, H. R. “Incidence and Prognosis of Asthma and Wheezing Illness from Early Childhood to Age 33 in a National British Cohort”. In: *British Medical Journal* 312.7040 (May 1996), pp. 1195–1199. ISSN: 0959-8138, 1468-5833. DOI: 10.1136/bmj.312.7040.1195.
38. Burrows, B., Martinez, F. D., Halonen, M., Barbee, R. A., and Cline, M. G. “Association of Asthma with Serum IgE Levels and Skin-Test Reactivity to Allergens”.

In: *New England Journal of Medicine* 320.5 (Feb. 1989), pp. 271–277. ISSN: 0028-4793. DOI: 10.1056/NEJM198902023200502.

39. Halonen, M., Stern, D., Taussig, L. M., Wright, A., Ray, C. G., and Martinez, F. D. "The Predictive Relationship Between Serum IgE Levels at Birth and Subsequent Incidences of Lower Respiratory Illnesses and Eczema in Infants". In: *American Review of Respiratory Disease* 146.4 (Oct. 1992), pp. 866–870. ISSN: 0003-0805. DOI: 10.1164/ajrccm/146.4.866.
40. Kattan, M., Keens, T. G., Lapierre, J.-G., Levison, H., Bryan, A. C., and Reilly, B. J. "Pulmonary Function Abnormalities in Symptom-Free Children After Bronchiolitis". In: *Pediatrics* 59.5 (May 1977), pp. 683–688. ISSN: 0031-4005, 1098-4275.
41. Samet, J. M., Tager, I. B., and Speizer, F. E. "The Relationship Between Respiratory Illness in Childhood and Chronic Air-Flow Obstruction in Adulthood". In: *American Review of Respiratory Disease* 127.4 (Apr. 1983), pp. 508–523. ISSN: 0003-0805. DOI: 10.1164/arrd.1983.127.4.508.
42. Jackson, D. J. et al. "Wheezing Rhinovirus Illnesses in Early Life Predict Asthma Development in High-Risk Children". In: *American Journal of Respiratory and Critical Care Medicine* 178.7 (Oct. 2008), pp. 667–672. ISSN: 1073-449X. DOI: 10.1164/rccm.200802-309OC.
43. Alcantara-Neves, N. M. et al. "The Effect of Single and Multiple Infections on Atopy and Wheezing in Children". In: *Journal of Allergy and Clinical Immunology* 129.2 (Feb. 1, 2012), 359–367.e3. ISSN: 0091-6749. DOI: 10.1016/j.jaci.2011.09.015.
44. Strina, A., Barreto, M. L., Cooper, P. J., and Rodrigues, L. C. "Risk Factors for Non-Atopic Asthma/Wheeze in Children and Adolescents: A Systematic Review". In: *Emerging Themes in Epidemiology* 11 (June 6, 2014), p. 5. ISSN: 1742-7622. DOI: 10.1186/1742-7622-11-5.
45. Rodriguez, A., Vaca, M. G., Chico, M. E., Rodrigues, L. C., Barreto, M. L., and Cooper, P. J. "Lifestyle Domains as Determinants of Wheeze Prevalence in Urban and Rural Schoolchildren in Ecuador: Cross Sectional Analysis". In: *Environmental Health* 14 (Feb. 4, 2015). ISSN: 1476-069X. DOI: 10.1186/1476-069X-14-15.
46. Pearce, N. et al. "Worldwide Trends in the Prevalence of Asthma Symptoms: Phase III of the International Study of Asthma and Allergies in Childhood (ISAAC)". In: *Thorax* 62.9 (May 2007), pp. 757–765. ISSN: 0040-6376. DOI: 10.1136/thx.2006.070169.

Part C: Manuscript

Title of article

Multi-state models for the analysis of wheeze in a birth cohort of Western Cape children.

Authorship*

Hannan, Patrick

Mailing Address: School of Public Health and Family Medicine, Faculty of Health Sciences, University of Cape Town, Private Bag X3, Observatory, 7925, Cape Town, South Africa

Email address: hnnpat002@myuct.ac.za

Corresponding author

Name: Hannan, Patrick

Mailing Address: School of Public Health and Family Medicine, Faculty of Health Sciences, University of Cape Town, Private Bag X3, Observatory, 7925, Cape Town, South Africa

Email address: hnnpat002@myuct.ac.za

*In accordance with the MPH dissertation guidelines, co-authors are not listed in the journal manuscript. The contribution of collaborators and supervisors are listed in the acknowledgments section of this dissertation. This manuscript is written according to the submission guidelines set out by the *International Journal of Epideiology* for submission as a regular research article. The guidelines are included in Appendix C. The following deviations from the style guide have been made to align with the requirements of the MPH dissertation: 1) Word count exceeds 3000 words; 2) Figures and tables have been inserted into the body of the text.

Abstract

Background: Childhood wheezing is associated with a range of morbidities including decreased lung function in later life. To date, the recurrent nature of childhood wheeze has not been examined in low- and middle-income countries. The objective of this study is to utilise multi-state models to estimate the rate of transition between various states of childhood wheeze and the effect of risk factors on those transitions in the context of a low- and middle-income country.

Methods: Longitudinal data are from 1086 children aged zero to three years, living in two peri-urban communities in South Africa collected as part of the Drakenstein Child Health Study. Transition intensities between wheeze states were estimated using discrete time multi-state models. The effects of risk factors on transition intensities were estimated using multivariable proportional hazards models.

Results: Transition from the No Wheeze state to the Wheeze state (0.00043) was equally likely as a transition to the LRTI-associated wheeze (0.0003). The transition from LRTI-associated Wheeze to Wheeze (0.002) was twice as likely as a transition from Wheeze to LRTI-associated Wheeze (0.00087). Male children were at significantly higher risk of having LRTI-associated wheeze as the first wheeze episode and at significantly higher risk of recurrent wheeze. Children exposed to prenatal maternal smoking had significantly increased risk of transition to the wheeze state compared to children not exposed to maternal smoking.

Conclusions: Multi-state models offer a novel way to study the progress of recurrent childhood wheeze with interval-censored data. Multi-state models for wheeze provide comparable results to traditional survival analysis while utilising data beyond the first occurrence of wheeze.

Key Messages: Multi-state models provide a novel framework for the analysis of childhood wheeze that deals with recurrent events and interval-censored data while providing estimates for risk factors comparable with traditional methods of analysis.

Introduction

Wheezing is defined as a continuous, coarse, whistling sound produced during breathing[1]. Wheezing is caused by obstructed respiratory passages, however, diagnosis is challenging given wheezing is a symptom of many related respiratory diseases, such as pneumonia. Childhood wheezing is a common occurrence, by the age of six approximately 50% of children in high-income countries will have experienced one or more episodes of wheezing in their life[2]. Furthermore, childhood wheeze is closely associated with reduced lung function and the development of asthma in later life[3,4]. Wheezing is not only a significant burden on the health of individual children but also places a significant burden on health systems, for example, it is estimated that the economic cost of childhood wheezing for a single year in the United Kingdom was 53 million GBP in 1998, approximately 90 million GBP per year when adjusted for inflation[5]. The prevalence of childhood wheezing and asthma has remained relatively constant in Europe, however, the prevalence of wheezing and asthma in Africa is increasing. The 12-month prevalence of wheezing in Africa in children aged six to seven is estimated to be 5.6%, and the 12-month prevalence of wheeze in children aged 13 to 14 is estimated to be 13.4%[6]. The estimated 12-month prevalence of wheezing in children aged thirteen to fourteen in Cape Town is higher than both the African and global estimates, with 20.3% of children experiencing wheezing[7].

The prevalence of wheeze and asthma are increasing in low- and middle-income countries (LMICs), although some work on wheeze has been conducted in Brazil and Ecuador[8–10], there is a paucity of research in these contexts, as most research has been conducted in high-income countries (HICs). Furthermore, although many wheeze studies are positioned in prospective birth cohort studies the most common method for the analysis of childhood wheeze is multivariable logistic regression investigating the outcome (typically "any episodes of wheeze in

the preceding 12 months") in a cross-sectional manner at a single time-point in the study. This study investigates the use of multi-state models to estimate the transition probability among various childhood wheeze states; no wheeze, wheeze, and lower respiratory tract (LRTI)-associated wheeze in a birth cohort of children followed up to three years of age. We also investigated the association between possible risk factors for childhood wheeze (sex, exposure to other children at home, exposure to maternal smoking, maternal history of asthma and season of birth), on the estimated transition intensities while accounting for the time-varying nature of covariates and the interval censored nature of the data.

Participants and Methods

Study Design

This study is a secondary analysis of data collected in a prospective birth cohort of mother-infant pairs enrolled in the Drakenstein Child Health Study (DCHS)[11]. The DCHS is a multi-disciplinary prospective birth cohort that aims to investigate the epidemiology, aetiology and long-term impact of early LRTI on child health in a LMIC. The protocol of the parent study has been described in detail elsewhere[11]. The study population for this analysis were children born to mothers in the Drakenstein area that made use of one of the two primary care clinics, TC Newman and Clinic and Mbekweni Clinic and who intended to remain in the area for at least a full year following enrolment. Study participants were located in the Drakenstein area in Paarl, the area has a population of approximately 200 000 people. The population is considered stable with little immigration or emigration[11].

The majority of the Drakenstein community members are of low socio-economic status and live in informal housing or crowded conditions. There is a high prevalence of various poverty-related

exposures, such as tobacco smoke exposure, alcohol misuse, and malnutrition[11]. More than 90% of the population made use of the freely available public healthcare system. The public health system in the Drakenstein area is comprised of 23 primary care clinics and one hospital, Paarl Hospital, where all births and hospital care occurred[11].

Data Collection

Socio-economic and medical maternal questionnaires were administered at antenatal visits. Child clinical and respiratory symptom questionnaires were completed at hospital visits and at routine study visits, which occurred at 6, 10 and 14 weeks, and 6, 9, 18, 30, 36, 42, 54 and 60 months as well as ad hoc visits for LRTI[11]. Active surveillance for LRTI was performed by the nurses at the study sites. Study nurses were trained to examine the respiratory health of children at sick visits. Measurements of LRTI included ambulatory and hospitalised pneumonia cases, as defined by World Health Organization (WHO) criteria. The current study reports on the first three years (36 months) of follow-up for each infant. Unattended visits were documented as missing but children were retained in intervals if they attended future visits.

Measures

Childhood wheezing was assessed in two distinct manners. Firstly, wheezing was identified by parental report as a positive answer to the question "Has your child had a wheezing or whistling noise coming from his/her chest in the past 6 months?". Secondly, trained healthcare workers recorded if a child had wheezing evident while present at a study or sick visit. Identification of wheezing could occur at planned study visits or at unscheduled visits for LRTI. Recurrent wheezing was defined as two or more episodes of wheezing, either parental report or healthcare worker observed, within a twelve month period.

Children who had no episodes of wheezing for the entire period of follow-up were classified as having never wheezed. Children who wheezed only once or had episodes of wheezing that were more than 12 months apart were classified as infrequent or transient wheezers. Similarly, children who had two or more episodes of wheezing in a 12-month period were classified as having recurrent wheezing. The time to first wheezing was the time in days from birth until the first episode of wheezing. The time to recurrent wheezing was the time in days from birth until the second episode of wheezing.

Statistical Analysis

All analyses were conducted in R (version 3.5.2)[12]. Descriptive statistics were presented as medians, interquartile ranges and frequencies (proportions), as appropriate. Survival curves for time to first wheeze and time to first recurrent wheezing were estimated using the product-limit estimator. Cox Proportional Hazards regression[13], including all potential confounding variables (infant sex, exposure to maternal smoke, exposure to other children and hospitalization), was used to estimate the association of potential risk factors with the time to first wheezing, as well as the time to first recurrent wheezing. The proportionality of hazards in the Cox model was investigated by plotting scaled Schoenfeld residuals of each covariate against survival time[14].

Multi-state Models

Multi-state models are a useful framework for modelling the progression of individuals through defined disease states, particularly when observations are captured at arbitrary timepoints relative to disease progression[15]. Therefore, a multi-state model was constructed to estimate the transition probabilities between pre-defined childhood wheeze states, as well as the effect of various risk factors (infant sex, exposure to maternal smoke, and exposure to other children) on

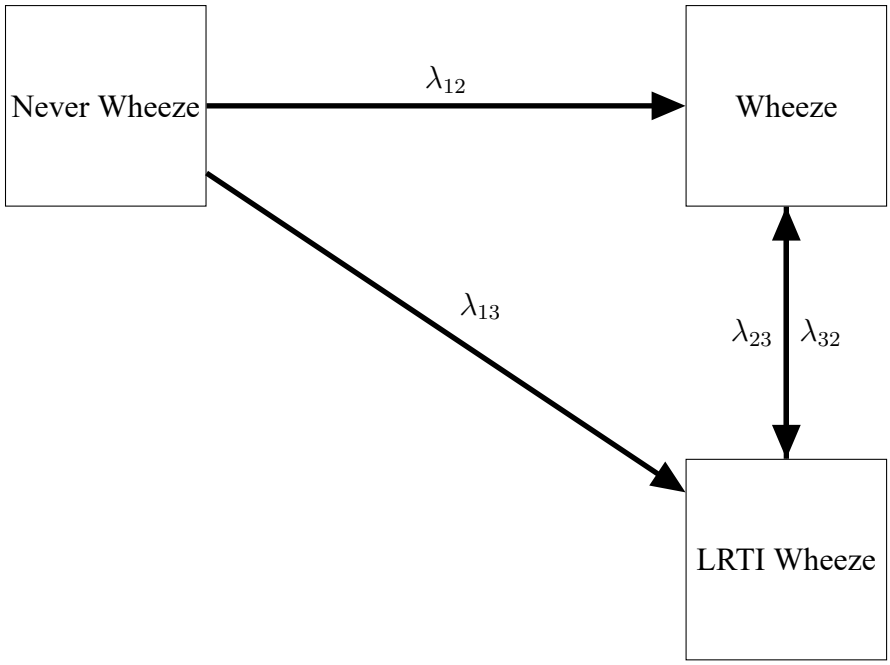


Figure 1: Diagram of the multi-state model used to estimate transition probabilities for childhood wheezing and LRTI-associated wheezing.

the transition probabilities. Given that follow-up in the DCHS begins at birth, all infants were assumed to be in the no wheeze state at $t = 0$.

The multi-state model allowed four possible transitions: 1) from "never wheeze" to "wheeze" or from 2) "never wheeze" to "LRTI-associated wheeze" or from 3) "wheeze" to "LRTI-associated wheeze" and from 4) "LRTI-associated wheeze" to "wheeze" (Fig. 1). The multi-state processes described above were all assumed to be *Markovian*, that is that future transition depended only on the currently occupied state. Although the multi-state processes were assumed to be Markovian, the risk of wheezing is known to decrease as infants age, with children younger than six months at greater risk of wheezing than older children[16]. Given that the risk of wheezing decreases with age, the transition intensities for the multi-state model in this study were assumed to have piecewise constant intensities. Namely, the transition intensities remained constant within each defined time period (0 - 180 days, 180 - 336 days, 336 - 728 days, and 728 - 1100 days), but could differ between time periods. Observation times of individuals were assumed to be non-

informative and interval censored with respect to disease progression. Covariate effects were estimated using a proportional hazards model for each transition[17]. All multi-state analyses were conducted using the **msm** package[18] for R[12]. Model assessment for panel-observed multi-state models is not straightforward due to transition times between adjacent states being unknown, it is further complicated by the interval censored nature of the data present in this study[19]. Therefore, in order to assess model fit in this study the expected prevalences in each state were compared to the empirical prevalences at all time points.

Ethical Approval

The DCHS received approval from the Ethics Committee of the Faculty of Health Sciences, University of Cape Town (HREC: 401/2009), as well as Stellenbosch University and the Western Cape Provincial Research committee. The current study was approved by the Ethics Committee of the Faculty of Health Sciences, University of Cape Town (HREC: 805/2018).

Results

In total 1086 infants were considered for this study (Fig. A1), the median follow-up time at the end of the study period was 913 days (784.25 - 923.75). Given the differing enrolment dates of infants into the DCHS, at the time of data extraction not all children in the DCHS had reached three years of age. Baseline demographic and clinical characteristics are presented in Table 1. An equal number of infants from both sexes were present in the study. The median birthweight for infants in the study was 3.0799999kg. A maternal history of asthma was rare in this cohort, only 1% of mothers in the study had a history of asthma. Approximately 23 of the mothers in the study self-reported smoking during pregnancy.

Table 1: Descriptive statistics [median (IQR) or *n* (%)] for children with transient or recurrent wheeze in the Drakenstein Child Health Study.

Characteristics	Overall $(n = 1086)$	Never/Infrequent Wheeze $(n = 859)$	Recurrent Wheeze $(n = 227)$
Age at First Wheeze (in days)[1]	180 (80.25 - 390.25)	255 (100 - 547)	130 (70 - 275.5)
Birth Weight (in kg)	3.08 (2.71 - 3.42)	3.12 (2.75 - 3.44)	3.04 (2.63 - 3.38)
Number of Children at Home	3 (3 - 4)	3 (3 - 4)	3 (3 - 4)
Sex			
Male	564 (51.93%)	420 (48.89%)	144 (63.44%)
Female	522 (48.07%)	439 (51.11%)	83 (36.56%)
Prenatal Maternal Smoking			
Yes	251 (23.11%)	181 (21.07%)	70 (30.84%)
No	822 (75.69%)	669 (77.88%)	153 (67.4%)
Unknown	13 (1.2%)	9 (1.05%)	4 (1.76%)
Maternal History of Asthma			
Yes	11 (1.01%)	7 (0.81%)	4 (1.76%)
No	1068 (98.34%)	847 (98.6%)	221 (97.36%)
Unknown	7 (0.64%)	5 (0.58%)	2 (0.88%)
Season of Birth			
Autumn (1 March – 31 May)	268 (24.68%)	189 (22%)	79 (34.8%)
Winter (1 June – 31 August)	288 (26.52%)	237 (27.59%)	51 (22.47%)
Spring (1 September – 30 November)	256 (23.57%)	214 (24.91%)	42 (18.5%)
Summer (1 December – 28 February)	274 (25.23%)	219 (25.49%)	55 (24.23%)

[1] **Amongst those who experienced wheeze.**

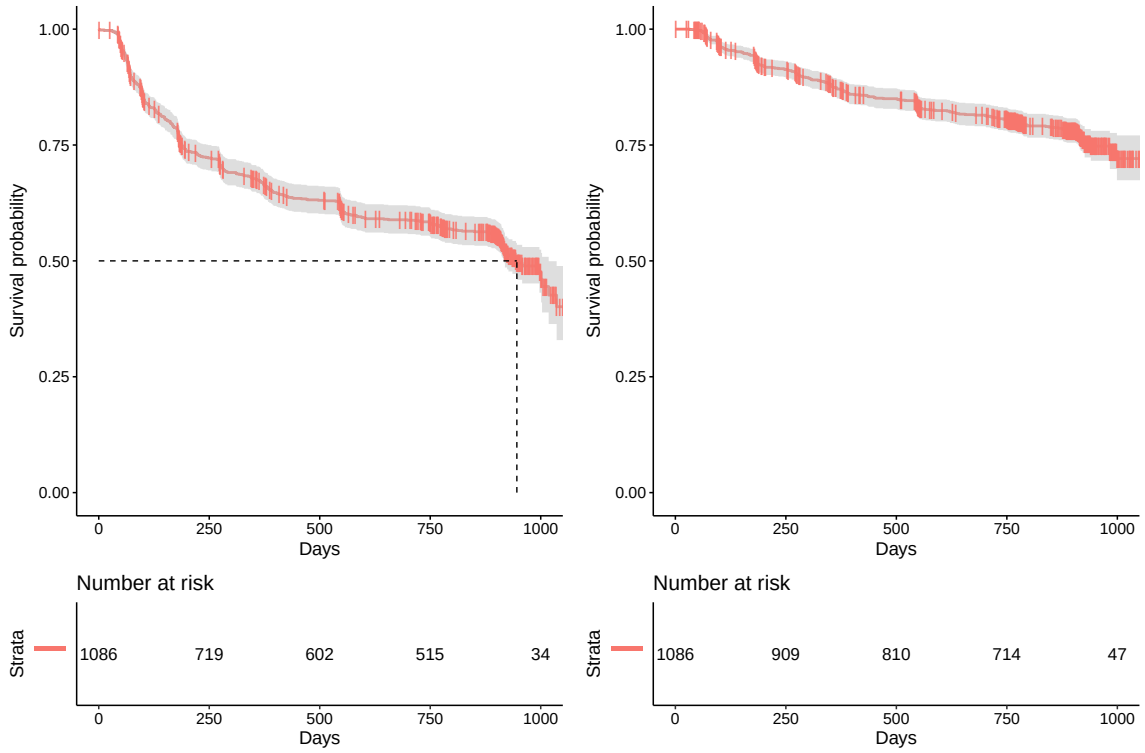
Composite wheezing was defined as either parentally reported or staff diagnosed wheezing at a given visit. A total of 951 episodes of wheezing were recorded in the 36 months of follow-up time reported on in this study, 610 (56%)children did not experience any episodes of wheezing. The incidence of wheezing in the cohort for the follow-up time reported on in this study was 0.114 episodes per child-year. Of the 951 episodes of wheezing, 650 (68.35%) were parentally reported, and 489 (51.42%) were diagnosed by study staff at a study visit. 188 (19.77%) episodes of wheezing were identified by both parental report and observed by study staff. The period prevalence of childhood wheezing in the first 36 months for the composite wheezing outcome was 43.8% (95% CI: 40.9% - 46.8%). The period prevalence of parentally-reported wheezing, for the same period, was 36.3% (95% CI: 33.4% - 39.1%), and the period prevalence of staff-diagnosed wheezing was 26.6% (95% CI: 24% - 29.2%).

The number of composite wheezing events in the first six months of follow-up time was 258, and the period prevalence of the composite wheezing outcome for the first six months only was 23.8% (95% CI: 21.2% - 26.3%). The period prevalence of parentally-reported and staff-reported wheezing was 16.7% (95% CI: 14.5% - 18.9%) and 13.5% (95% CI: 11.5% - 15.6%) respectively. Of the 1086 children in the study, 227 (20.9%) experienced more than one episode of wheezing in the three years of follow-up. The majority of children who experienced recurrent wheeze (63%) were male.

The median time to first wheezing episode was 946 days (95% CI: 916 - 1018 days), however, the median time to recurrent wheezing was not reached in the three years of follow-up (Fig. 2) as only 227 (20.9%) children had a recurrent wheezing episode. Male children had a significantly lower median time to first wheezing episode (916 days; 95% CI: 752 - 1001) than female children (1018 days; 95% CI: 924 - NA). However, given that so few individuals had recurrent episodes of wheezing the estimate for median survival time for recurrent wheeze is outside the follow-up window in this study.

In multivariable Cox regression analysis for time to first wheezing, no risk factors were associated with a significantly increased hazard of wheeze. While not significant, male children (HR = 1.16; 95% CI: 0.96 - 1.4), hospitalization for a pneumonia episode (HR = 1.22; 95% CI: 0.89 - 1.67), or exposure to other children at home (HR = 1.03; 95% CI: 0.95 - 1.11) were all associated with an increased hazard of wheezing. Furthermore, while not significant, the season of birth (Spring, Summer and Winter) was associated with a decreased hazard of wheezing relative to birth in Autumn. The results of the multivariable Cox regression for time to first recurrent wheezing were similar to the results for time to first wheeze episode in both magnitude and direction (Table A1). However, exposure to other children in the home was associated with a

significantly increased hazard of recurrent wheezing (HR = 1.19; 95% CI: 1.07 - 1.32). Diagnostic for both time to first wheeze and time to first recurrent wheeze Cox proportional hazards models indicate that the proportional hazards assumption had not been violated.



(a) First Wheezing Episode **(b)** First Recurrent Wheezing Episode

Figure 2: Kaplan-Meier Curve for time to first wheezing episode and first recurrent wheezing episode

Multi-state Model

The number of transitions between consecutive observed states for all states over the entire three year period of follow-up is presented in Table 2. There were a total of 672 transitions into adjacent states. The primary results of the multi-state analysis, using composite wheezing as the outcome, are displayed in Table 3, the maximum likelihood estimate (MLE) and corresponding 95% confidence intervals at the mean value for all covariates for each allowed transition are presented. For categorical variables, the mean of the 0/1 dummy variable for each factor level was used, representing an average over all values in the data, rather than a specific factor level.

The multi-state analysis was repeated for the two other definitions of wheeze (parental report and staff diagnosed; Appendix D), the covariate effects on transitions were similar to those reported for the composite wheezing endpoint. However, the confidence intervals were wider and transition intensities between states were significantly lower given the reduced number of transitions into each state.

Table 2: The number of transitions between each consecutive state in the LRTI-associated wheeze multi-state model for the full duration of follow-up.

	No Wheeze	**Wheeze**	**LRTI-Wheeze**
No Wheeze	4889	243	170
Wheeze	0	1237	110
LRTI-Wheeze	0	149	849

Transition from the No Wheeze (State 1)

Children who had never wheezed were equally likely to transition to having wheeze (0.0004) as they were to transition to LRTI-associated wheeze (0.0003). The probability that an individual who was in the no wheeze state at six-months would transition to the wheeze state by thirty-six months was 0.296 (95 % CI: 0.256 - 0.352), and the probability an individual would transition from No Wheeze to the LRTI-associated wheeze state by the end of the same time period was 0.101 (95 % CI: 0.077 - 0.14). The transition probabilities estimate the relative chance that an individual transitions from a given state to another state within the defined time period.

The hazard ratios for each transition of the multi-state model analysed in this study are presented in Table 4. Male children were more likely to transition to the Wheeze state (HR = 1.029; 95% CI: 0.775 - 1.365) and to the LRTI-associated Wheeze state (HR = 1.766; 95% CI: 1.265 - 2.465). Similarly, children exposed to prenatal maternal smoking were more likely to transition from the no wheeze state to the wheeze state (HR = 1.598; 95% CI: 1.173 - 2.179). The transition

intensity for all transitions out of the wheeze state decreased over time, after the first six months the baseline transition intensity for the no wheeze to wheeze transition was less than half the initial transition intensity (HR = 0.451; 95% CI: 0.31 - 0.656).

Table 3: Estimated baseline transition intensity matrix for the Wheeze and LRTI-associated Wheeze multi-state model in the DCHS.

State	Wheeze		LRTI-Wheeze	
	MLE	**95% CI**	**MLE**	**95% CI**
No Wheeze	0.00043	0.00036; 0.0005	0.0003	0.00025; 0.00037
Wheeze	0[a]		0.00087	0.00069; 0.00087
LRTI-Wheeze	0.002	0.0016; 0.0025	0[a]	

[a] Transition intensities were defined as zero in the multi-state model.

Transition from Wheeze (State 2)

The transition intensity for the transition from the wheeze state to the LRTI-associated wheeze state was 0.00087, approximately half the transition intensity for the LRTI-associated wheeze to wheeze transition (0.002). The six-month transition probability of moving from the Wheeze state to the LRTI-associated wheeze state was 0.181. Similarly, the probability an individual in the wheeze state at six months would be observed in the LRTI-associated wheeze state at 36 months was 0.227.

All transition intensities from the wheeze state decreased relative to the first time period (0 to 180 days). Male children were approximately twice as likely to transition from the wheeze state to the LRTI-associated wheeze state (HR = 1.79; 95% CI: 1.17 - 2.75) than female children. Children exposed to prenatal maternal smoking were more likely to transition from the wheeze state to the LRTI-associated wheeze state (HR = 1.31; 95% CI: 0.84 - 2.04), however, the increase was not statistically significant (Table 4).

Table 4: Hazard ratios from discrete-time event-history models of transitions between Wheeze states for children aged zero to three years in the DCHS.

Variable	Wheeze State Transition							
	No Wheeze - Wheeze		No Wheeze - LRTI Wheeze		Wheeze - LRTI Wheeze		LRTI Wheeze - Wheeze	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Number of Children at Home	0.925	(0.812 - 1.054)	1.097	(0.959 - 1.255)	1.005	(0.846 - 1.194)	0.894	(0.769 - 1.04)
Sex								
Male	1.029	(0.775 - 1.365)	1.766	(1.265 - 2.465)	1.794	(1.17 - 2.75)	1.3	(0.894 - 1.891)
Female	Reference							
Prenatal Maternal Smoking								
Yes	1.598	(1.173 - 2.179)	1.037	(0.698 - 1.539)	1.309	(0.84 - 2.039)	1.642	(1.112 - 2.425)
No	Reference							
Season of Birth								
Winter (1 June – 31 August)	0.842	(0.574 - 1.234)	0.465	(0.289 - 0.749)	0.48	(0.28 - 0.824)	0.76	(0.477 - 1.21)
Spring (1 September – 30 November)	0.847	(0.567 - 1.264)	0.665	(0.428 - 1.033)	0.454	(0.251 - 0.82)	0.54	(0.326 - 0.893)
Summer (1 December – 28 February)	0.681	(0.447 - 1.037)	0.755	(0.494 - 1.154)	0.643	(0.373 - 1.109)	0.622	(0.397 - 0.974)
Autumn (1 March – 31 May)	Reference							
Time Period								
0 to 180 days	Reference							
180 to 336 days	0.451	(0.31 - 0.656)	0.416	(0.276 - 0.627)	0.336	(0.173 - 0.653)	0.107	(0.053 - 0.215)
336 to 728 days	0.133	(0.082 - 0.215)	0.189	(0.123 - 0.292)	0.262	(0.15 - 0.459)	0.17	(0.109 - 0.263)
728 to 1100 days	0.354	(0.235 - 0.535)	0.082	(0.032 - 0.211)	0.213	(0.11 - 0.413)	0.191	(0.118 - 0.309)

Transition from LRTI-associated wheeze (State 3)

The transition intensity for the transition to the non-LRTI associated wheeze state was 0.002, approximately double the transition intensity for the transition from the wheeze state to the LRTI-associated wheeze state (0.00087). The transition probability for the 0 - 180 day time period for transition to the wheeze state was 0.642. Similarly, the transition probability for the 180 - 1100 day period for the same transition was 0.638. The transition intensity for all transitions out of the LRTI-associated wheeze state decreased significantly over the three year period of follow-up. Male children and children who were exposed to prenatal maternal smoking had a significantly increased hazard of transition from the LRTI-associated wheeze state (Table 4).

Model fit was assessed graphically by plotting the observed prevalence for each state against the expected prevalence for each state (Fig. A2). Model fit was generally acceptable, however, observed and expected prevalence deviated sharply for follow-up times greater than 1000 days.

Discussion

This study explored a novel application of multi-state Markov models for interval-censored observations to describe the childhood wheeze disease process from the ages of zero to three years, within the context of a low- and middle-income setting The period prevalence reported in this study is within the range of similar studies on childhood wheezing[2,20]. The period prevalence of healthcare worker ascertained wheezing (26.6%) was lower than both parentally-reported (36.3%) and composite wheezing (43.8%) likely due to the manner in which staff-reported wheezing was assessed. Study-staff only reported wheezing if it was present at a given study visit, whereas parentally-reported wheeze and composite wheeze reporting included the recol-

lection of the preceding months prior to a study or sick visit.

None of the known risk factors for childhood wheezing included in the multivariable Cox model for time to first wheezing were significantly associated with an increased risk of wheezing. For example, prenatal exposure to maternal smoking is known to be associated with an increased risk of childhood wheezing[2,16,21,22]. Although children exposed to prenatal maternal smoking did have a 1.08 (95% CI: 0.87 - 1.34) times higher hazard of wheezing it was not statistically significant. Similarly, the only risk factor associated with a significantly increased hazard of recurrent wheezing in these models was exposure to other children at home. One of the limitations of Cox models and other methods that do not account for the recurrent nature of wheezing is that data after the first event is not included in the analysis. This can be particularly problematic for childhood wheezing given that the risk factors for the first wheezing episode may be different from subsequent wheezing events[16].

The multi-state model in this study allowed for the inclusion of individuals with recurrent wheezing. Although, the majority of children had one or fewer episodes of wheezing in the three years of follow-up, by accounting for children with recurrent wheezing a further 259 transitions were included in the multi-state analysis. The risk of transition into any wheeze state decreased significantly as children aged, the first six months of life had the greatest transition intensity for all transitions in the multi-state model. These results support findings from several studies, that used cross-sectional methods of analysis, on the age-related risk of wheezing[3,16,22,23] and extend them into an LMIC context.

There was no significant difference between the sexes for transition into the wheeze state. However, male children were significantly more likely to transition into the LRTI-associated wheeze state from any state. Furthermore, male children were more likely to have more than one episode

of wheezing. The results support findings from similar studies in HICs[2,24,25]. It is believed that the increased risk of wheeze for male children is likely a combination of an inheritable predisposition to wheezing[26] and biomechanical differences in lung morphology between male and female children[27]. It is reasonable to believe that these risk factors would remain consistent regardless of setting.

There is a well-known association between maternal smoking and wheeze in children[2,16,21,28], however, the associations in this study were not consistent. Maternal smoking was associated with a significantly increased risk of transition into the wheeze state from any state. Maternal smoking was not associated with a significantly increased risk of transition into the LRTI-associated wheeze state. Given the strength of the evidence for an association between maternal smoking and wheezing, and maternal smoking and LRTI, it is likely that the discrepancy between the two transitions in this study is a result of the smaller number of transitions into the LRTI-associated wheeze state which may have led to reduced power. Another possible explanation for the discrepancy is that in this study maternal smoking was measured via maternal self-report and that this measure may under-represent true smoking exposure given the stigma associated with maternal smoking during pregnancy.

Although there were no significant associations detected by the Cox Proportional Hazards model, the direction of effect for all of the risk factors that were included in both the Cox regression model and the multi-state model was the same. However, the multi-state model made use of all of the wheezing events (944 events), whereas the Cox Proportional Hazards models for time to first wheeze and time to recurrent wheezing included only 450 and 254 events respectively. Furthermore, unlike Cox models, multi-state models provide an easily extensible framework, that can be used to deal with covariates that differ for different transitions or events.

This study has several strengths and limitations. A major strength of the analysis in this study was the use of prospectively collected longitudinal data and an appropriate analysis method that allowed for recurrent events and time-varying exposures, which to the author's knowledge is the first use of multi-state models for the analysis of childhood wheezing in the context of an LMIC. However, parametric multi-state Markov models make strong assumptions about the underlying disease process ("the memorylessness property" and constant baseline transitions) that may not be suitable for all disease processes. Multi-state models can be extended to deal with violations of these assumptions, for example, piecewise models, as used in this study, can be fitted if it is believed that baseline transition intensities differ over the entire study period but remain constant for shorter time periods. Another important limitation of this study is that the number of children that transitioned into some states was low which may lead to a lack of statistical power. A further limitation of this study is that a narrow set of prespecified covariates were included in the multi-state model. Furthermore, it is important to note that the parentally reported wheeze outcome is determined by parental recall, and parents may be more likely to recall wheezing symptoms if their children were ill at some point in the study. The composite wheezing outcome may be similarly biased given it is a composite of parental recall and healthcare worker observed wheeze.

Conclusion

This study has shown a novel application of multi-state Markov models for the analysis of childhood wheeze in a South African context. The multi-state model successfully predicted the progress of childhood wheezing and produced results consistent with existing literature on childhood wheezing while accounting for the recurrent nature of wheezing and interval-censored observations. However, further research should investigate the use of more complex multi-state

models to account for the differing wheeze phenotypes and their associated risk factors.

References

1. Kliegman, R. M., Behrman, R. E., Jenson, H. B., and Stanton, B. M. D. *Nelson Textbook of Pediatrics*. Elsevier Health Sciences, Aug. 15, 2007. 3200 pp. ISBN: 978-1-4377-2180-5.
2. Martinez, F. D., Wright, A. L., Taussig, L. M., Holberg, C. J., Halonen, M., and Morgan, W. J. "Asthma and Wheezing in the First Six Years of Life". In: *New England Journal of Medicine* 332.3 (Jan. 1995), pp. 133–138. DOI: 10.1056/NEJM199501193320301.
3. Hovland, V., Riiser, A., Mowinckel, P., Carlsen, K.-H., and Carlsen, K. C. L. "The Significance of Early Recurrent Wheeze for Asthma Outcomes in Late Childhood". In: *European Respiratory Journal* 41.4 (Apr. 2013), pp. 838–845. ISSN: 0903-1936, 1399-3003. DOI: 10.1183/09031936.00071512.
4. Zomer-Kooijker, K., van der Ent, C. K., Ermers, M. J. J., Uiterwaal, C. S. P. M., Rovers, M. M., and Bont, L. J. "Increased Risk of Wheeze and Decreased Lung Function after Respiratory Syncytial Virus Infection." In: *PLOS One* 9.1 (Jan. 2014), e87162. ISSN: 1932-6203 1932-6203. DOI: 10.1371/journal.pone.0087162.
5. Stevens, C. A., Turner, D., Kuehni, C. E., Couriel, J. M., and Silverman, M. "The Economic Impact of Preschool Asthma and Wheeze". In: *European Respiratory Journal* 21.6 (June 2003), pp. 1000–1006. ISSN: 0903-1936, 1399-3003. DOI: 10.1183/09031936.03.00057002.
6. Pearce, N. et al. "Worldwide Trends in the Prevalence of Asthma Symptoms: Phase III of the International Study of Asthma and Allergies in Childhood (ISAAC)". In: *Thorax* 62.9 (May 2007), pp. 757–765. ISSN: 0040-6376. DOI: 10.1136/thx.2006.070169.
7. Zar, H. J., Ehrlich, R. I., Workman, L., and Weinberg, E. G. "The Changing Prevalence of Asthma, Allergic Rhinitis and Atopic Eczema in African Adolescents from 1995 to 2002". In: *Pediatric Allergy and Immunology* 18.7 (Nov. 2007), pp. 560–565. ISSN: 0905-6157, 1399-3038. DOI: 10.1111/j.1399-3038.2007.00554.x.
8. Alcantara-Neves, N. M. et al. "The Effect of Single and Multiple Infections on Atopy and Wheezing in Children". In: *Journal of Allergy and Clinical Immunology* 129.2 (Feb. 1, 2012), 359–367.e3. ISSN: 0091-6749. DOI: 10.1016/j.jaci.2011.09.015.
9. Strina, A., Barreto, M. L., Cooper, P. J., and Rodrigues, L. C. "Risk Factors for Non-Atopic Asthma/Wheeze in Children and Adolescents: A Systematic Review". In: *Emerging Themes in Epidemiology* 11 (June 6, 2014), p. 5. ISSN: 1742-7622. DOI: 10.1186/1742-7622-11-5.
10. Rodriguez, A., Vaca, M. G., Chico, M. E., Rodrigues, L. C., Barreto, M. L., and Cooper, P. J. "Lifestyle Domains as Determinants of Wheeze Prevalence in Urban and

Rural Schoolchildren in Ecuador: Cross Sectional Analysis". In: *Environmental Health* 14 (Feb. 4, 2015). ISSN: 1476-069X. DOI: 10.1186/1476-069X-14-15.

11. Zar, H. J., Barnett, W., Myer, L., Stein, D. J., and Nicol, M. P. "Investigating the Early-Life Determinants of Illness in Africa: The Drakenstein Child Health Study". In: *Thorax* 70.6 (June 2015), pp. 592–594. ISSN: 0040-6376, 1468-3296. DOI: 10.1136/thoraxjnl-2014-206242.
12. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, July 2018.
13. Cox, D. R. "Regression Models and Life-Tables". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2 (Mar. 1972), pp. 187–202. ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1972.tb00899.x.
14. Grambsch, P. M. and Therneau, T. M. "Proportional Hazards Tests and Diagnostics Based on Weighted Residuals". In: *Biometrika* 81.3 (Sept. 1, 1994), pp. 515–526. ISSN: 0006-3444. DOI: 10.1093/biomet/81.3.515.
15. Cook, R. J. and Lawless, J. F. *Multistate Models for the Analysis of Life History Data*. Chapman and Hall/CRC, May 2018. ISBN: 978-1-4987-1561-4. DOI: 10.1201/9781315119731.
16. Sherriff, A., Peters, T. J., Henderson, J., and Strachan, D. "Risk Factor Associations with Wheezing Patterns in Children Followed Longitudinally from Birth to 3(1/2) Years." In: *International journal of epidemiology* 30.6 (Dec. 2001), pp. 1473–1484. ISSN: 0300-5771 0300-5771. DOI: 10.1093/ije/30.6.1473.
17. Marshall, G. and Jones, R. H. "Multi-State Models and Diabetic Retinopathy". In: *Statistics in Medicine* 14.18 (Sept. 1995), pp. 1975–1983. ISSN: 1097-0258. DOI: 10.1002/sim.4780141804.
18. Jackson, C. H. "Multi-State Models for Panel Data: The Msm Package for R". In: *Journal of Statistical Software* 38.8 (Apr. 2011). ISSN: 1548-7660. DOI: 10.18637/jss.v038.i08.
19. Titman, A. C. and Sharples, L. D. "Model Diagnostics for Multi-State Models". In: *Statistical Methods in Medical Research* 19.6 (Dec. 2010), pp. 621–651. ISSN: 1477-0334. DOI: 10.1177/0962280209105541.
20. Lai, C. K. W., Beasley, R., Crane, J., Foliaki, S., Shah, J., and Weiland, S. "Global Variation in the Prevalence and Severity of Asthma Symptoms: Phase Three of the International Study of Asthma and Allergies in Childhood (ISAAC)". In: *Thorax* 64.6 (June 2009), pp. 476–483. ISSN: 3906770559. DOI: 10.1136/thx.2008.106609.
21. Bisgaard, H., Dalgaard, P., and Nyboe, J. "Risk Factors for Wheezing during Infancy. A Study of 5,953 Infants." In: *Acta Paediatrica Scandinavica* 76.5 (Sept. 1987),

pp. 719–726. ISSN: 0001-656X 0001-656X. DOI: 10.1111/j.1651-2227.1987.tb10555.x.

22. Hallit, S. et al. “Wheezing Phenotypes and Risk Factors in Early Life: The ELFE Cohort.” In: *PLOS One* 13.4 (May 2018). ISSN: 1932-6203 1932-6203. DOI: 10.1371/journal.pone.0196711.
23. Stein, R. T. et al. “Respiratory Syncytial Virus in Early Life and Risk of Wheeze and Allergy by Age 13 Years”. In: *The Lancet* 354.9178 (Aug. 1999), pp. 541–545. DOI: 10.1016/S0140-6736(98)10321-5.
24. Taylor-Robinson, D. C., Pearce, A., Whitehead, M., Smyth, R., and Law, C. “Social Inequalities in Wheezing in Children: Findings from the UK Millennium Cohort Study.” In: *The European respiratory journal* 47.3 (Mar. 2016), pp. 818–828. ISSN: 1399-3003 0903-1936. DOI: 10.1183/13993003.01117-2015.
25. Bernstein, D. I. “Traffic-Related Pollutants and Wheezing in Children.” In: *The Journal of Asthma* 49.1 (Feb. 2012), pp. 5–7. ISSN: 1532-4303 0277-0903. DOI: 10.3109/02770903.2011.641049.
26. Tse, S. M., Coull, B. A., Sordillo, J. E., Datta, S., and Gold, D. R. “Gender- and Age-Specific Risk Factors for Wheeze from Birth through Adolescence”. In: *Pediatric Pulmonology* 50.10 (Oct. 2015), pp. 955–962. ISSN: 8755-6863. DOI: 10.1002/ppul.23113.
27. Liptzin, D. R., Landau, L. I., and Taussig, L. M. “Sex and the Lung: Observations, Hypotheses, and Future Directions”. In: *Pediatric Pulmonology* 50.12 (Dec. 2015), pp. 1159–1169. ISSN: 1099-0496. DOI: 10.1002/ppul.23178.
28. Sahiner, U. M. et al. “Recurrent Wheezing in the First Three Years of Life: Short-Term Prognosis and Risk Factors.” In: *The Journal of Asthma* 50.4 (May 2013), pp. 370–375. ISSN: 1532-4303 0277-0903. DOI: 10.3109/02770903.2013.770013.

Appendices

A Supplementary figures/tables included in Manuscript

Table A1: Relative risks of wheeze and recurrent wheeze amongst children aged zero to three years in the DCHS.

Variable	First Wheeze		First Recurrent Wheeze	
	HR	**95% CI**	**HR**	**95% CI**
Exposure to other children at home	1.03	0.95 - 1.11	1.19	1.07 - 1.32
Sex				
Male	1.16	0.96 - 1.4	1.23	0.95 - 1.6
Female		Reference		
Prenatal Maternal Smoking				
Yes	1.08	0.87 - 1.34	1.2	0.91 - 1.59
No		Reference		
Season of Birth				
Winter (1 June – 31 August)	0.83	0.64 - 1.08	0.75	0.53 - 1.06
Spring (1 September – 30 November)	0.75	0.58 - 0.98	0.69	0.49 - 0.99
Summer (1 December – 28 February)	0.9	0.7 - 1.16	0.7	0.49 - 0.98
Autumn (1 March – 31 May)		Reference		
Hospitalization for Pneumonia				
Hospitalized	1.22	0.89 - 1.67	1.19	0.76 - 1.85
Ambulatory		Reference		

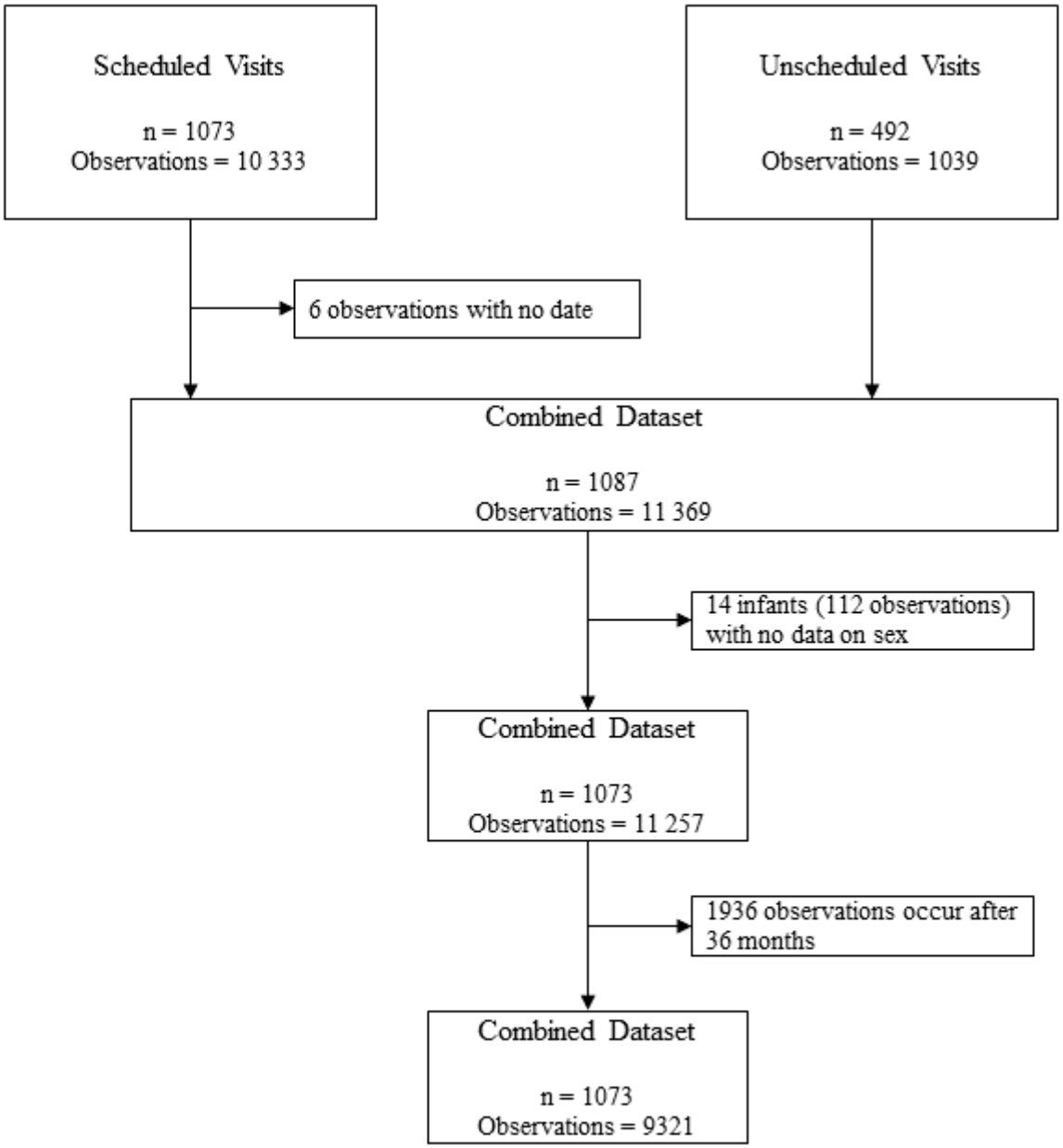


Figure A1: Consort diagram of enrolment, inclusion and exclusion criteria.

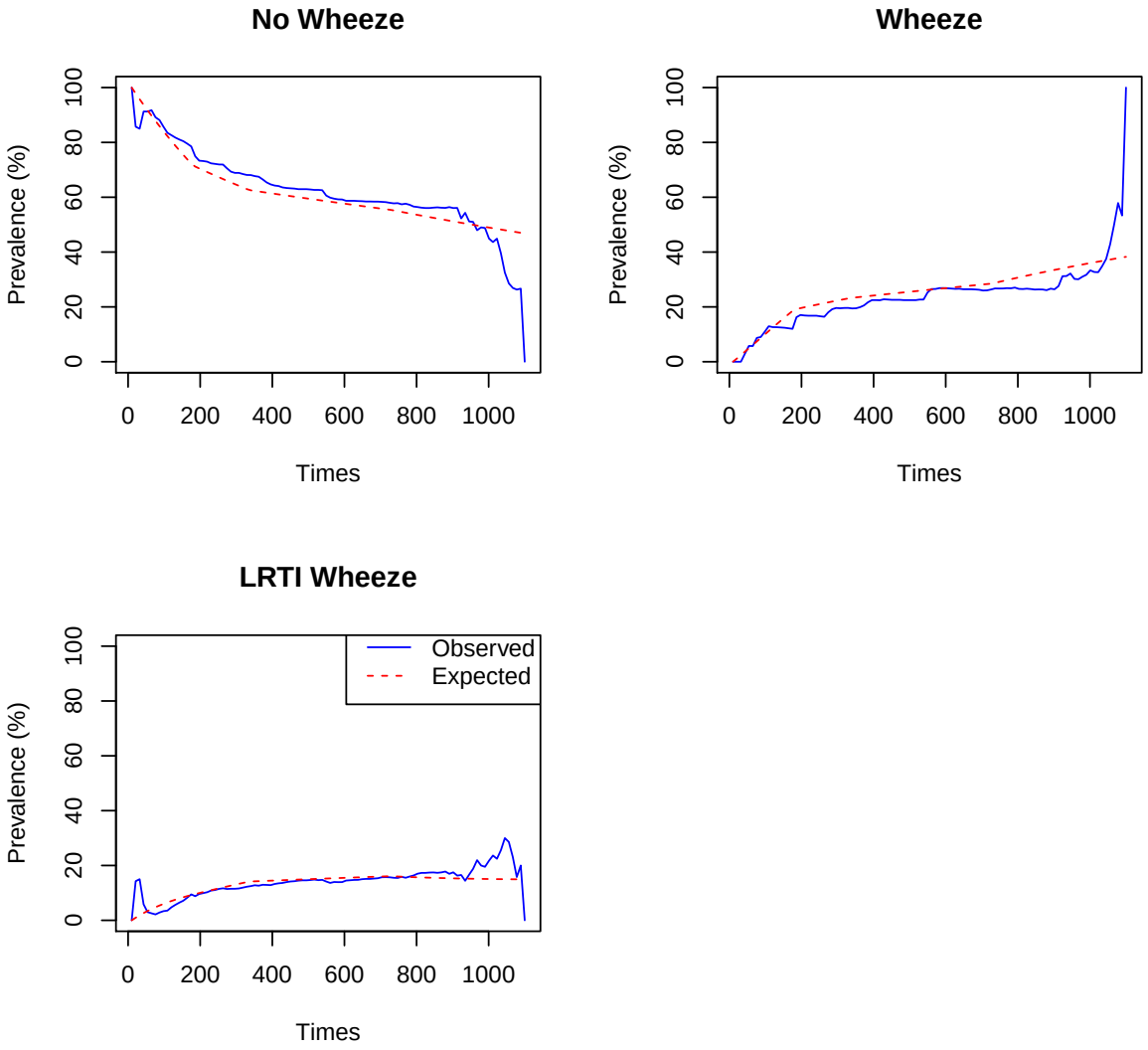


Figure A2: Observed vs expected prevalence for each state in the multi-state model at each time point.

B Human Research Ethics Committee (HREC) approval letter

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

Room E53-46 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6492
Email: sumayah.arlefdien@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

23 November 2018

HREC REF: 805/2018

Dr M Lesosky
School of Public Health & Family Medicine
Division of Epi Bios
Falmouth Building-FHS

Dear Dr Lesosky

PROJECT TITLE: MULTI-STATE MODELS FOR THE ANALYSIS OF WHEEZE IN A COHORT OF WESTERN CAPE CHILDREN (MMED Candidate - Mr P Hannan)

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.

Approval is granted for one year until the 30 November 2019.

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)

We acknowledge that the student: Mr Patrick Hannan will also be involved in this study.

Please quote the HREC REF 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.

Yours sincerely

Signature Removed

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

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 Convention on Harmonisation 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.

C Submission Guidelines for International Journal of Epidemiology

INSTRUCTIONS TO AUTHORS

All manuscripts must be submitted online. Once you have prepared your manuscript according to the instructions below please visit the online submission web site here . Further technical guidance on submitting your manuscript online via ScholarOne is available here .

CONDITIONS

All submissions must be in the English language. Articles are accepted for publication on 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.

Please note that 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 for the purpose of communicating 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.

Unfortunately the journal is not staffed to provide a pre-submission service. FORMS

Authorship Form : A scanned copy of the Authorship Form signed by the corresponding author must be submitted via the online manuscript submission system. Failure to submit the form will lead to delays in the processing of the manuscript.

Conflict of Interest Form : A Conflict of Interest (CoI) Form must be submitted via the online manuscript submission system when you first submit your article. The form must be completed by corresponding author on behalf of all authors. 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. It is the responsibility of the corresponding author to ensure that all authors adhere to this policy prior to submission. A CoI statement must also be included in the manuscript after any ”Acknowledgements” and ”Funding” sections and should summarize all aspects of any conflicts of interest included on the form. If there is no conflict of interest, authors must include ’Conflict of Interest: none declared’.

License to Publish Form: The License to Publish Form is only required when a paper is accepted for publication. It can be filled in online, and a link to the form will be sent to authors at the relevant time.

MANUSCRIPT PREPARATION

TITLES

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

AUTHOR LIST

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 (*). The editorial office retains the right to limit the number of authors appearing on the title page.

Writing groups may wish to list the group name instead of an author list. If individual members of such a group wish to be credited with authorship on PubMed, a full list of the authors (clearly identifying the PIs and corresponding author) must be uploaded as a 'support document (not for publication)' during the online submission process.

SUMMARY/ABSTRACT

The summary should be no more than 250 words and consist of four sections labelled Background, Methods, Results and Conclusions. They should briefly describe the problem addressed, how the study was performed, the salient results and what conclusions can be made from the results. Three to ten keywords should be added to the end of the Summary.

KEY MESSAGES

Please include a key messages box with the key messages of the paper in 3-5 succinct bullet points.

MAIN DOCUMENT

- The main document should be submitted as an editable word document (not as a pdf).
- Papers should not normally exceed 3000 words but review articles may be twice this

length.

- Manuscripts should be double spaced with margins of 2.5cm.

- All pages should be numbered.

- Numbers followed by a unit should be written as figures, as should all numbers above nine. Figures should not be used to start a sentence and those between 999 and 9999 should not be separated by spaces or commas while those over 10 000 should have a space after the thousand.

- Percent should be written as % throughout.

- Full points should not be used after initials or contractions: J Jones, FRCS, 17 g, dl, Dr, etc.

- All measures should be reported in SI units followed, in the text, by traditional units in parentheses. For general guidance on the International System of Units and useful conversion factors, see Conventional Units - International Units. There are two exceptions: blood pressure should be expressed in mm Hg and haemoglobin as g/dl.

- If the data are appropriate, age grouping should be mid-decade to mid-decade or in five-year age groups (e.g. 35-44 or 35-39, 40-44, etc, but not 20-29, 30-39 or other groupings.

- 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”. Where used, we ask authors to provide effect estimates with confidence intervals and exact P values, and to refrain from the use of the term ”significant” in ei-

ther the results or discussion section of their papers. Our justification of this position is given in the 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

FUNDING

The following rules should be followed:

- The sentence should begin: 'This work was supported by …'
- The full official funding agency name should be given, i.e. 'the National Cancer Institute at the National Institutes of Health' or simply 'National Institutes of Health' not 'NCI' (one of the 27 subinstitutions) or 'NCI at NIH' - see the full RIN-approved list of UK funding agencies for details
- 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]'
- Agencies should be separated by a semi-colon (plus 'and' before the last funding agency)
- Where individuals 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]'.

An example is given here: 'This work was supported by the National Institutes of Health [P50

CA098252 and CA118790 to R.B.S.R.] and the Alcohol & Education Research Council [HFY GR667789].

Crossref Funding Data Registry: In order to meet your 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 click here.

REFERENCES

- Footnotes are not permitted in the main body of papers (they may only be used in tables and figures). Instead, authors should provide a list of references.

- References in Vancouver Style should be listed in the order they appear in the text and numbered accordingly. These numbers should be inserted in the text in brackets whenever a reference is cited.

- The numbered list of references should appear at the end of the article and should consist of the surnames and initials of all authors when six or less (when seven or more list just three and add et al.), title of article, name of journal abbreviated according to Index Medicus style, year, volume, first and last page numbers Example: 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; 431:614-622.

- Titles of books should be followed by the place of publication, the publisher, and the year. 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 may be cited and should appear appropriately marked in the text, but not in the reference list. Manuscripts in press may be cited in the references and details added on proof if possible.

TABLES

- Tables should be submitted in editable Word or Excel format (not pdf).
- Tables should be numbered consecutively in arabic numerals and should be kept separate from the text.
- Particular care should be taken to make tables self-explanatory with adequate headings and footnotes.
- The position of each table in the text should be indicated (Table 1 here)

FIGURES

- Figures should be submitted in editable image formats (such as jpg or tiff, not pdf).
- Illustrations should be numbered and given suitable legends.
- They should be kept separate from the text.
- As standard figures appear in black and white in print, and in colour online. There is no charge for this.

- Authors will be expected to pay if they want their figures reproduced in colour in the print version of the Journal (£350/figure).

- Please state your preferred option (i.e. agreement to pay £350/figure for print and online colour or preference for online-only colour with no charge) upon submission via the online submission system.

- Please ensure that the prepared electronic image files print at a legible size and are of a high quality for publication (600dpi for line drawings; 300dpi for colour and half-tone artwork).

- For useful information on preparing your figures for publication, please see `http://cpc.cadmus.com/da`.

ABBREVIATIONS/ACRONYMS

The use of acronyms should be kept to a minimum. Words to be abbreviated (acronyms) should be spelled out in full the first time they appear in the text with the abbreviations in brackets. Thereafter the abbreviation should be used.

APPENDICES

As a general rule, material of this nature should be incorporated in the text but separate sections can be published after the main text.

SUPPLEMENTARY MATERIAL

- Supplementary material that is not essential for inclusion in the full text of the manuscript, but would nevertheless benefit the reader, can be made available by the publisher as

online-only content, linked to the online manuscript.

- The material should not be essential to understanding the conclusions of the paper.
- Such information might include more detailed methods, extended data sets/data analysis, or additional figures (including colour).
- All text and figures must be provided in suitable electronic formats. Permitted formats: PDF (.pdf), MS Word (.doc), HTML (.html), RTF (.rtf), MS Excel (.xls), CSV, mp3, mpeg.
- Please provide short (2-4 word) 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 Data 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 ensure that the Supplementary Material is referred to in the main manuscript where necessary.

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:

- capitals and small letters;
- ordinary and bold-faced letters;
- certain greek letters and similar roman letters;

- subscripts, superscripts and ’ordinary’ symbols.

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.

Figure D3: Model schematic for simple uni-directional three-state model of childhood wheeze

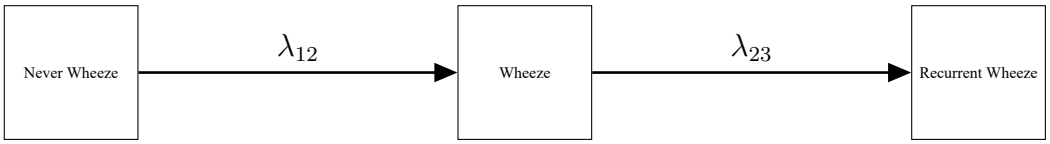


Table D2: The number of transitions between each consecutive state for simple mutli-state model for the full duration of follow-up.

Model 1: Unidirectional Three-State Model			
	Never Wheeze	**Ever Wheeze**	**Recurrent Wheeze**
Never Wheeze	4889	413	0
Ever Wheeze	0	1173	227
Recurrent Wheeze	0	0	945

D Appendices to the Thesis

All analysis that were completed as part of the MPH, but were not included in the manuscript portion are presented here for completeness.

Table D3: Estimated baseline transition intensity matrix, with covariates set at their mean values, for the simple unidirectional multi-state model.

Model 1: Unidirectional Three-State Model				
State	**Ever Wheeze**		**Recurrent Wheeze**	
	MLE	**95% CI**	**MLE**	**95% CI**
No Wheeze	0.00059	0.00052; 0.00067	0[a]	
Ever Wheeze	-0.0012	-0.0015; -0.001	0.0012	0.001; 0.0015
Recurrent Wheeze	0 [b]	-0	0; 0	

[a] Transition intensities were defined as zero in the multi-state model for non-adjacent states.
[b] Unidirectional models did not allow for transitions into previous states, and thus transition intensities were defined as zero.

Table D4: Estimated 6-month transition probabilities for the simple unidirectional multi-state model.

Model 1: Unidirectional Three-State Model						
State	**No Wheeze**		**Ever Wheeze**		**Recurrent Wheeze**	
	MLE	**95% CI**	**MLE**	**95% CI**	**MLE**	**95% CI**
No Wheeze	0.777	0.75; 0.802	0.171	0.149; 0.194	0.052	0.04; 0.068
Ever Wheeze	0	0; 0	0.585	0.492; 0.664	0.415	0.336; 0.508
Recurrent Wheeze	0	0; 0	0	0; 0	1	1; 1

Table D5: Relative risks of wheeze and recurrent wheeze amongst children aged zero to three years in the DCHS, using healthcare observed wheezing.

Variable	First Wheeze		First Recurrent Wheeze	
	HR	**95% CI**	**HR**	**95% CI**
Exposure to other children at home	1.04	0.95 - 1.15	1.16	1 - 1.35
Sex				
Male	1.4	1.09 - 1.81	1.81	1.17 - 2.8
Female		Reference		
Prenatal Maternal Smoking				
Yes	0.99	0.75 - 1.31	1.12	0.73 - 1.73
No		Reference		
Season of Birth				
Winter (1 June – 31 August)	0.9	0.64 - 1.28	0.6	0.34 - 1.07
Spring (1 September – 30 November)	0.66	0.46 - 0.93	0.63	0.38 - 1.06
Summer (1 December – 28 February)	1.01	0.73 - 1.39	0.59	0.35 - 1.01
Autumn (1 March – 31 May)		Reference		
Hospitalization for Pneumonia				
Hospitalized	1.82	1.34 - 2.48	2.39	1.48 - 3.87
Ambulatory		Reference		

Table D6: Relative risks of wheeze and recurrent wheeze amongst children aged zero to three years in the DCHS, using parentally reported wheezing.

Variable	First Wheeze		First Recurrent Wheeze	
	HR	**95% CI**	**HR**	**95% CI**
Exposure to other children at home	1.05	0.96 - 1.14	1.18	1.04 - 1.34
Sex				
Male	1.09	0.89 - 1.35	1.23	0.9 - 1.66
Female		Reference		
Prenatal Maternal Smoking				
Yes	1.1	0.87 - 1.4	1.25	0.9 - 1.76
No		Reference		
Season of Birth				
Winter (1 June – 31 August)	0.78	0.59 - 1.02	0.76	0.51 - 1.15
Spring (1 September – 30 November)	0.66	0.5 - 0.88	0.67	0.44 - 1.01
Summer (1 December – 28 February)	0.75	0.57 - 1	0.72	0.48 - 1.08
Autumn (1 March – 31 May)		Reference		
Hospitalization for Pneumonia				
Hospitalized	0.4	0.2 - 0.78	1.03	0.5 - 2.12
Ambulatory		Reference		

Table D7: Hazard ratios from discrete-time event-history models of transitions between Wheeze states for children aged zero to three years in the DCHS for parentally reported wheeze.

Variable	Wheeze State Transition							
	No Wheeze - Wheeze		No Wheeze - LRTI Wheeze		Wheeze - LRTI Wheeze		LRTI Wheeze - Wheeze	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Number of Children at Home	0.934	(0.84 - 1.039)	1.139	(0.915 - 1.417)	0.958	(0.72 - 1.273)	0.775	(0.562 - 1.069)
Sex								
Male	1.179	(0.93 - 1.495)	1.158	(0.689 - 1.949)	1.472	(0.77 - 2.813)	2.69	(1.367 - 5.294)
Female					Reference			
Prenatal Maternal Smoking								
Yes	1.681	(1.304 - 2.168)	0.542	(0.256 - 1.148)	0.715	(0.344 - 1.489)	0.544	(0.237 - 1.25)
No					Reference			
Season of Birth								
Winter (1 June – 31 August)	0.699	(0.505 - 0.966)	0.492	(0.252 - 0.96)	0.631	(0.273 - 1.459)	0.682	(0.318 - 1.463)
Spring (1 September – 30 November)	0.683	(0.486 - 0.959)	0.553	(0.28 - 1.094)	0.522	(0.198 - 1.378)	0.778	(0.34 - 1.78)
Summer (1 December – 28 February)	0.691	(0.5 - 0.956)	0.319	(0.142 - 0.717)	0.995	(0.44 - 2.254)	0.811	(0.314 - 2.096)
Autumn (1 March – 31 May)					Reference			
Time Period								
0 to 180 days					Reference			
180 to 336 days	0.415	(0.302 - 0.571)	1.135	(0.489 - 2.633)	0.716	(0.046 - 11.125)	0.008	(0 - $3.3699407 \times 10^{157}$)
336 to 728 days	0.137	(0.093 - 0.203)	1.139	(0.552 - 2.35)	4.286	(0.573 - 32.082)	1167.179	(0 - 1.8182306×10^{30})
728 to 1100 days	0.281	(0.192 - 0.412)	0.895	(0.344 - 2.332)	7.066	(0.934 - 53.449)	1695.34	(0 - 2.64169×10^{30})

Table D8: Hazard ratios from discrete-time event-history models of transitions between Wheeze states for children aged zero to three years in the DCHS for healthcare worker observed wheeze.

Variable	Wheeze State Transition							
	No Wheeze - Wheeze		No Wheeze - LRTI Wheeze		Wheeze - LRTI Wheeze		LRTI Wheeze - Wheeze	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Number of Children at Home	1.055	(0.794 - 1.401)	1.042	(0.92 - 1.181)	1.529	(1.104 - 2.117)	0.725	(0.454 - 1.157)
Sex								
Male	1.413	(0.736 - 2.714)	1.801	(1.339 - 2.423)	1.708	(0.742 - 3.931)	1.749	(0.631 - 4.851)
Female					Reference			
Prenatal Maternal Smoking								
Yes	2.379	(1.225 - 4.618)	1.046	(0.744 - 1.472)	2.283	(0.936 - 5.571)	1.032	(0.379 - 2.808)
No					Reference			
Season of Birth								
Winter (1 June – 31 August)	1.228	(0.47 - 3.203)	0.508	(0.34 - 0.759)	0.133	(0.043 - 0.407)	0.505	(0.133 - 1.917)
Spring (1 September – 30 November)	1.212	(0.435 - 3.374)	0.617	(0.413 - 0.919)	0.447	(0.15 - 1.328)	0.783	(0.2 - 3.063)
Summer (1 December – 28 February)	1.339	(0.498 - 3.6)	0.744	(0.512 - 1.081)	0.303	(0.107 - 0.861)	1.017	(0.364 - 2.84)
Autumn (1 March – 31 May)					Reference			
Time Period								
0 to 180 days					Reference			
180 to 336 days	0.487	(0.217 - 1.092)	0.504	(0.35 - 0.726)	0.188	(0.056 - 0.636)	0.208	(0.06 - 0.719)
336 to 728 days	0.103	(0.034 - 0.308)	0.217	(0.15 - 0.314)	0.101	(0.037 - 0.271)	0.122	(0.044 - 0.34)
728 to 1100 days	0.237	(0.088 - 0.642)	0.091	(0.042 - 0.194)	0.063	(0.018 - 0.223)	0	(0 - 3.2276553×10^{96})