

https://dspace.org/

Hierarchical Bayesian estimation of the prevalence and incidence of viral acute respiratory tract infections in Kilifi, Kenya

Otieno, Grieven Paul

2021

https://hdl.handle.net/10539/32168

Downloaded from a DSpace repository. https://dspace.org/

Hierarchical Bayesian Estimation of the Prevalence and Incidence of Viral Acute Respiratory Tract Infections in Kilifi, Kenya.



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Grieven Paul Otieno

Student Number: 1846979

Supervisors:

1. **Dr. Innocent Maposa**- University of the Witwatersrand, South Africa

2. **Prof. James Nokes**- KEMRI Wellcome Trust Research Programme, Kenya.

3. **Prof. Ngianga-Bakwin Kandala**- University of the Witwatersrand, South Africa

A research report submitted to the School of Public Health, University of the Witwatersrand, in partial fulfillment of the requirements of the degree in Master of Science in Epidemiology (Biostatistics).

April, 2021

DECLARATION

I Grieven P. Otieno, student number 1846979, declare that this research report is my own work. It is being submitted for the degree of Master of Science in Epidemiology in the field of Biostatistics at the University of the Witwatersrand, Johannesburg. It has not been submitted before any degreee or examination or any other university.

Signature: 

April 2021

DEDICATION

To my wife and son.

ABSTRACT

Introduction: Acute respiratory infections (ARIs) are communicable diseases of public health concern due to high morbidity and mortality burden especially in low and middle income countries (LMICs). Accurate and precise estimates of disease burden inform policy and resource allocation hence contributing to the realisation of the sustainable development goals aimed at achieving universal health coverage and reducing preventable deaths. However, such estimates at sub-national levels are mostly imprecise due to small sample sizes or unavailable due to lack of sampling at such levels. Such limitations are overcome by methods such as small area estimation (SAE).

Methods: This study utilised data from a previously conducted study at 9 government owned outpatient healthcare facilities at the Kilifi health and demographic surveillance system (KHDSS) at the Kenyan coast. Each health facility was considered to represent a location out of the 15 locations in the KHDSS. ARI was defined as presence of at least one of cough, sneezing, nasal congestion, difficulty breathing or increased respiratory rate for age. Patients presenting to the clinic with any of the listed ARI symptoms were eligible for recruitment. A total of 15 samples per week per health facility were targeted for collection from the eligible patients over the entire study period. Samples were tested using multiplex polymerese chain reaction (PCR) assay for respiratory syncytial virus (RSV A and B), human rhinovirus (HRV), human coronaviruses (HCoVs-OC43, NL63 and E229), influenza (A, B and C), parainfluenza viruses (PIV 1-4), adenovirus (ADV) and human metapneumovirus (hMPV). Data analysis was conducted on the number of cases that turned positive for each of the 7 virus pathogens. Hierarchical Bayesian SAE models were used to estimate and predict the prevalence and incidence rates of the viral pathogens stratified by age group (all ages, 0-59 months and 60 months and above) and controlling for auxilliary variables (mean age of participants, number of health facilities in a location and mean monthly rainfall). For each virus, 5 models were fitted each assuming a different underlying probability distribution; Poisson, zero-inflated Poisson, binomial, zero-inflated binomial and logit-normal transformation of the observed proportions of positive cases. The model with best fit for each virus pathogen by age category was selected with low values of the deviance information criteria (DIC) and number of effective parameters (pD). Data analysis was conducted in R and Bayesian model specification done using R-INLA.

Results: For a period of 18 months starting January 2016, a total of 7985 ARI cases were recruited at 9 outpatient health facilities within the KHDSS. Generally, virus infections were observed through out the study period, though in relatively low numbers in the last 6 months of the study. Crude estimates showed that the highest prevalence and incidence were borne by children aged 0-59 months (p-values < 0.05) for all viral pathogens except FLU . In contrast, model-based estimates revealed variations of ARI prevalence across age groups while incidence rates were highest in the under 5s age band for all viruses. Model-based estimates over the 18 months of study yielded prevalence estimates for all ages at 4% for RSV, 14% HRV, 9% HCoV, 4% FLU, 4% PIV, 8% ADV and <1% hMPV. Among the under 5s the most prevalent infections were HRV (12%) and FLU (5.6%). In this age band, incidence rates per 100000 children were 502 for RSV, 3249 HRV, 330 HCoV, 1006 FLU, 910 ADV and 477 hMPV. Spatio-temporal variations of both prevalence and incidence rates were observed across all age bands. In general, the lowest prevalence (<5%) and incidence (<100 cases per 100000 population) rates for RSV, PIV and ADV were observed in the second half of 2016 while the first halves of both years (2016 and 2017) had relatively higher estimates across the 15 locations. The remaining viral pathogens displayed variations of the time interval with the lowest or highest burden of disease by location while for pathogens such as FLU these variations could not be differentially identified. The logit normal model generally gave the best fit for most virus targets when estimating the prevalence while for estimation of incidence rates the Poisson and binomial models were mostly the best performing across age categories.

Conclusion:

Our results have shown that estimates of ARI prevalence and incidence obtained by crude/ unadjusted methods differed from those obtained via SAE model-based techniques. SAE methods provided the advantage of quantifying disease burden within the KHDSS while adjusting for space and time variations as well as other auxilliary variables such as number of health facilities in a location rainfall patterns. This provided better measures, within a Bayesian framework, that were otherwise either over or under estimated by direct methods. We observe that there exists a considerable burden of viral ARI in Kilifi with a substantial part of the incidental burden borne by children aged under 5 years. This has implications on future interventions such as vaccination, localised therapies, surveillance and planning.

ACKNOWLEDGEMENTS

I am grateful to my supervisors Dr. Innocent Maposa, Professors James Nokes and Ngianga-Bakwin Kandala for the mentorship and guidance.

Special thanks to the Sub-Saharan Consortium of Advanced Biostatistics (SSACAB) for funding my Masters studies at the University of the Witwatersrand, the teaching and non-teaching staff at the School of Public Health who in so many ways have added value to my life. My classmates and friends (Morelearnings Sibanda and Zisiwe Mahlati); from whom I have learnt so much.

I wish to thank the KEMRI-Wellcome Trust Research Programme in Kilifi, Kenya for granting me access to the data used in this project through the data governance commitee. Specifically, many thanks to Prof. James Nokes and the entire Viral Epidemiology Control group (http://virec-group.org/). In addition, gratitude to the IDEAL team of the KEMRI-Wellcome for the support offered throughout my fellowship at KWTRP.

My family has been so supportive through out my studies; special gratitude to my parents, wife and the entire family whose prayers and support kept me going.

Glory and honour to God almighty. Erokamano Nyasaye

Contents

DECLARATION ii

DEDICATION iii

ABSTRACT iv

ACKNOWLEDGEMENTS vi

NOMENCLATURE ix

CHAPTER1: INTRODUCTION 1

1.1 Background . . . 1

1.1.1 Epidemiology of Acute Respiratory Infections . . . 1

1.1.2 Small Area Estimation . . . 2

1.2 Literature Review . . . 2

1.2.1 SAE in Epidemiology . . . 2

1.2.2 Respiratory Infections in Kilifi . . . 3

1.3 Problem Statement . . . 4

1.4 Justification of the study . . . 5

1.5 Research question . . . 5

1.6 Study aim and Objectives . . . 5

1.6.1 Study aim . . . 5

1.6.2 Study Objectives . . . 6

References 7

CHAPTER 2: MANUSCRIPT 11

2.1 SUBMISSIBLE MANUSCRIPT . . . 11

2.1.2 Manuscript Figures . . . 30

2.1.3 Supplementary File 1 . . . 38

CHAPTER 3: EXTENDED METHODS 42

CHAPTER 4: EXTENDED RESULTS **47**

4.1 Temporal Component 47

4.1.1 Exploratory Data Analysis 47

4.1.2 Non-linear effects of time 47

4.2 Spatial Component 50

4.2.1 Spatial Autocorrelation 50

4.2.2 Spatial random effects 50

4.3 Models comparisson 55

4.3.1 Model fit 55

4.3.2 Model estimates 55

4.3.2 Model fit diagnostics 58

CHAPTER 5: OVERALL DICUSSIONS **61**

APPENDIX **66**

A1. Plagiarism Statement 66

A2. Turnitin Report 67

A3. Ethics clearence 68

A4. Declaration: Student's contribution to article and agreement of co-author 69

A5. Hierarchical Bayes SAE specification 70

A6: Autocorrelation Plots 72

A7. Prevalence WAIC values 73

A8. Incidence WAIC values 74

A9. R packages 75

A10. INLA models R Codes 77

A11. Instruction for Authors 79

NOMENCLATURE

ADV	Adenovirus
ARI	Acute Respiratory Infections
ARI burden	Prevalence and Incidence of ARI
FLU	Influenza virus
HCoV	Human coronavirus
hMPV	Human metapneumovirus
HRV	Human rhinovirus
KCH	Kilifi county hospital
KHDSS	Kilifi health and demographic surveillance system
KWTRP	KEMRI Wellcome Trust Research Programme
iid	independent identically distributed
IG	Inverse Gamma distribution
LMIC	Low and Middle Income Countries
MCMC	Markov Chain Monte Carlo
PIV	Parainfluenza virus
RSV	Respiratory Syncitial Virus
RW	Random walk process
SAE	Small Area Estimation
SDG	Sustainable development goals

List of Tables

2.1 Crude prevalence and incidence estimates for each virus pathogen stratified by age category and overall. 25

2.2 Estimates of ARI prevalence and 95% credible intervals (CIs) for each virus pathogen tested by age category 26

2.3 Incidence rate estimates per 100000 mid-year population and 95% credible intervals (CIs) for each virus pathogen tested by age category 27

S.1 Deviance information criteria (DIC) and effective number of parameters (pD) used for comparisson of models used estimate the prevalence of viral ARI in Kilifi. 28

S.2 Deviance information criteria (DIC) and effective number of parameters (pD) used for comparisson of models used estimate incidence rates of viral ARI per 100000 population in Kilifi. 29

A.1 Watanabe–Akaike information criterion (WAIC) used for comparisson of models used estimate the prevalence of viral ARI in Kilifi. 73

A.2 Watanabe–Akaike information criterion (WAIC) used for comparisson of models used estimate the incidence rates of viral ARI in Kilifi. 74

List of Figures

2.1 Map showing administrative locations within the KHDSS. The grey fill highlights locations where the study was conducted and the red dots show the participating health facilities within the locations. 30

2.2 Panel of graphs showing the trends of each virus pathogen. The primary y-axis denotes the monthly numbers positive cases for each virus pathogen shown by the red solid line. The secondary y-axis denotes the number of samples tested shown by the black dashed line. 31

2.3 Panel of maps showing the estimated prevalence by virus pathogen for all ages within the KHDSS. 32

2.4 Panel of maps showing the estimated prevalence by virus pathogen children aged 0-59 months within the KHDSS. 33

2.5 Panel of maps showing the estimated incidence rates by virus pathogen for all ages within the KHDSS. 34

2.6 Panel of maps showing the estimated incidence rates per 100000 population by virus pathogen for children aged 0-59 months within the KHDSS. 35

2.7a Panel of estimated prevalence for all ages (i), under 5s (ii) and over 5s (iii). 36

2.7b Panel of estimated incidence rates for all ages (i), under 5s (ii) and over 5s (iii). . . . 37

4.1 Panel of graphs showing non-linear effects of time in months on the prevalence of viral ARI for all ages (a), under 5 year olds (b) and over 5 year olds (c). Shown are the posterior means together with 95intervals. 48

4.2 Panel of graphs showing non-linear effects of time in months on the incidence of viral ARI for all ages (a), under 5 year olds (b) and over 5 year olds (c). Shown are the posterior means together with 95intervals. 49

4.3a Maps showing the residual spatial locations effects on the prevalence of viral ARI in Kilifi for all ages 51

4.3b Panel of maps showing the residual spatial locations effects on the prevalence of viral ARI in Kilifi for under 5 year olds (i) and over 5 year olds (ii). 52

4.4a Maps showing the residual spatial locations effects on the incidence rates of viral ARI in Kilifi for all ages 53
4.4b Panel of maps showing the residual spatial locations effects on the incidence rates of viral ARI in Kilifi for under 5 year olds (i) and over 5 year olds (ii). 54
4.5 Panel of violin plots showing the estimated prevalence by virus pathogen tested from each model fitted on the data. The panel plots are presented for all ages (a), under 5 year olds (b) and over 5 year olds (c). 56
4.6 Panel of violin plots showing the estimated incidence rates by virus pathogen tested from each model fitted on the data. The panel plots are presented for all ages (a), under 5 year olds (b) and over 5 year olds (c). 57
4.7 Panel of posterior density plots from models used to estimate ARI prevalence by virus pathogen tested. The panel plots are presented for all ages (a), under 5 year olds (b) and over 5 year olds (c). 59
4.8 Panel of posterior density plots from models used to estimate incidence rates by virus pathogen tested. The panel plots are presented for all ages (a), under 5 year olds (b) and over 5 year olds (c). 60
4.9 Autocorrelation plots for each virus pathogen tested in Kilifi. 72

CHAPTER1: INTRODUCTION

1.1 Background

1.1.1 Epidemiology of Acute Respiratory Infections

Acute Respiratory Infections (ARIs) are among the leading causes of morbidity and mortality especially in children aged <5 years and older people aged >70 years (1,2). It is estimated that ARIs cause more than 1 million deaths in the world annually (1). In recent years, the global burden of ARIs has reduced although some regions, especially in countries with low Socio-demographic Index (SDI), continue to bear a substantial burden (1–3). In addition, Sub-Saharan Africa (SSA) reported the slowest decreases in ARI associated under-5 mortality (1).

In 2016, ARIs accounted for 650,639 deaths in SSA of which 48% were children aged <5 years and 21% were adults aged >70years (1). In addition, over 250,000 ARI associated deaths were estimated to have occured in eastern Sub-saharan Africa of which 48.2% and 22% were children aged <5 years and older adults aged >70 years respectively (1). Respiratory syncytial virus (RSV) and influenza contribute the biggest burden of viral ARI related hospitalisations in children (1,4). Intervention measures such as vaccines against influenza have been implemented and development of a vaccine against RSV is at an advanced stage (5). However, the epidemiology and aetiology of ARI remains largely unknown in many African countries (2).

Measuring, monitoring and evaluating the progress towards attaining Sustainable Development Goals (SGDs) as well as evaluating impact of vaccine interventions require accurate estimates of health indicators (6,7). In most cases, these indicators are reported at national or higher regional levels such as SSA, Asia, Europe etc (1). Due to limited resources and geographical variation of disease burden within countries, utilising sub-national health indicators in low and middle income countries (LMICs) informs policy and interventions for improved health care and progression towards realising SDGs starting at the local levels (8–10).

Subnational estimates of disease burden are affected by inadequate sample sizes leading to estimates of low precision or unavailable due to lack of sampling in some areas (11). These shortcomings, facing conventional methods of estimating disease prevalence, can be addressed by using advanced

statistical techniques called Small Area Estimation (SAE) to account for small or lack of samples at sub-national levels (12).

1.1.2 Small Area Estimation

SAE is a statistical technique used to estimate population parameters in domains (areas) where direct estimates cannot be precisely obtained due to small sample sizes or no sample is available in a given domain by borrowing strength through space and/or time (12,13). Recently, there has been increased interest in the field of SAE with significant advances made in both theoretical and application of components of SAE in disciplines such as employment, administration, education, economics and healthcare (12–14).

SAE methods are broadly classified as design-based or model-based (12,13). Design-based approaches use models to construct model assisted estimators whose properties such as bias and variance are evaluated under randomisation which is introduced by the sampling design. Model-based methods use either frequentist, empirical or full Bayesian approaches to obtain indirect estimates for small areas. Models used in SAE methods can be specified at unit (individual) or aggregate (area) level based on the type of auxilliary variables available (12,13).

In public health, accurate and precise indicators are useful for estimating disease burden, trends of disease over time and tracking progress of disease eradication both at the global and national levels (15). Governments in developed countries increasingly use SAE methods in public health assessment within small domains for improved policy and governance (16). For instance, in the United States of America (USA), it is a requirement of law for federal health system agencies at state and local governments to collect and analyse health data of citizens within their jurisdictions (17). Evidence of SAE use by governments in LMICs is lacking although several research studies on conditions such as Tuberculosis, HIV and child mortality have utilised SAE methods in their analysis (18–20).

1.2 Literature Review

1.2.1 SAE in Epidemiology

SAE micro-simulation technique has traditionally been used for poverty mapping in econometrics (21). An extension of this method was used to show spatial variation of ARI prevalence among

children(<5 years) in Bangladesh using nested error logistic regression extended to higher levels allowing for within cluster correlations (22). In Kenya, a Gaussian process regression was used to show that ARI associated rates of years of life lost due to premature death and years of life lived with disability among children aged <5 years consistently reduced over time for every county within 3 decades (23). A different study used Bayesian hierarchical spatial modelling approach to show spatial variability of ARI prevalence among children aged <5 years by borrowing strength from other sampled regions in the study (24).

The same Bayesian hierarchical technique showed spatial variation and better prediction of undernutrition among children in Ethiopia (25). A generalised linear mixed model with a Bayesian Hierarchical approach while accounting for unstructured spatial and intrinsic temporal effects was used to show that tuberculosis had space-time variations in Kenya (20).

Using the ability of SAE methods to borrow strength over space and time, a Bayesian spatio-temporal Gaussian process regression model with heteroskedastic error component was used to show varied spatial and declining temporal pattern of under 5 mortality in Kenya (18). Similary, hierarchical SAE modeling was used to show childhood mortality varied in 35 African countries by both space (sub national levels) and time (26). In several developed countries, SAE methods have also been used to estimate prevalence of non-communicable diseases / conditions (27,28).

1.2.2 Respiratory Infections in Kilifi

1.2.2.1 Inpatient surveillance

Previous studies on viral respiratory infections in the KHDSS have mainly focused on long-term surveillance of hospital admissions at Kilifi county hospital (KCH) among children aged 0-59 months associated with syndromic clinically defined severe or very severe pneumonia (29,30). From this surveillance, RSV was estimated to have a prevalence ranging from 15% to 34% and incidence rates of 293 to 535 per 100000 children (29,30). In the period January to December 2007, RSV was estimated to have the highest incidence rate at 535 per 100000 compared to the incidence rates of other viral respiratory infections; Human coronavirus 229E (HCoV-229E) at 105, influenza type A (FLU A) at 82, parainfluenza virus 3 (PIV3) at 57, adenovirus at 55 and human metapnemovirus (hMPV) at 44 hospitalisations per 100000 (30). The prevalence of other viral respiratory pathogens

at KCH have also been reported elsewhere. These were estimated at 4.1% for hMPV (31) and for FLU prevalence and incidence rates of 4.9% and 106 per 100000 respectively (32). A prevalence of 1.3% was reported on HCoV-NL63 over a period of 7 years (33). A separate study covering 13 years of surveillance reported an overall prevalence of 4% for all HCoVs (1% for NL63, 0.7% for OC43 and 1.4% for 229E) (34).

1.2.2.2 Outpatient surveillance

Fewer outpatient studies have been reported from Kilifi. At KCH outpatient department, it was estimated that RSV had a prevalence and incidence of 7.7% and 767 per 100000 children respectively over a period of 2 years (35). In a three year period, rhinovirus was reported to have a prevalence of 14% (36) while a lower prevalence of 3.9% at KCH was reported on FLU over a 4 year period (32).

1.2.2.3 Community surveillance

A study conducted at 47 households showed that the most prevalent viruses, in descending order, were rhinovirus (10.5%), HCoV (7.5%), AdV (7.3%) and RSV (3.2%) (37). In a primary school setup, the prevalence of viral ARIs among children and adolescents aged 2-19 years varied by virus pathogen as follows; 16.7% for rhinovirus, 2.7% parainfluenza virus, 2.0% HCoV, 0.9% adenovirus, 0.6% RSV and 0.3% FLU (38).

1.3 Problem Statement

Despite ARI's declining global prevalence, it remains a disease of huge concern that causes over a million mortality annually of which over 40% occur in Africa (2,3). Specifically, a huge burden of the ARI related morbidity and mortality in SSA are borne by children under the age of 5 years (1). In addition, ARI has been associated with increased hospitalisations among children who are at times hospitalised in high dependecy or intensive care units (ICUs) with very severe disease (39,40). Intervention strategies such as treatment and vaccination aimed at reducing ARI mortality and morbidity especially among children aged <5 years require a clear understanding of disease burden especially in LMICs (2). In order to achieve this reduction, in line with the Sustainable Development Goal (41) number 3, it is important to obtain accurate estimates of the burden of ARI which would inform global, national and small area specific intervention policies. However, shortage

of ARI surveillance data, especially in LMICs, poses a huge challenge on efforts meant to obtain accurate measures of ARI disease burden health indicators such as disease prevalence and incidence (42). This is evidenced by the low number of available literature on ARI within outpatient settings in Kenya (43,44) and specifically Kilifi (32,35,36) hence limited estimates of either crude or model based measures of ARI disease burden associated with viral pathogens currently circulating in the population.

1.4 Justification of the study

Reviewed literature has shown that ARI, especially among children, is still a major contributor to the global burden of diseases (2,4). In addition, SAE methods have been suggested to offer a superior approach of estimation and prediction especially in domains of very small sample sizes or no sample at all (12). SAE methodologies also extend to accommodate sampling methodology and borrow information through space and/or time (12,13). Therefore, fitting SAE models on ARI data in a small area, such as Kilifi within Kenya, will contribute to the body of knowledge and provide data on estimates of the burden of ARI which is currently scarce. Such estimates can be used to facilitate achievement of SDG 3 targets namely: (*i*) research and development of therapeutic interventions and (*ii*) strengthened capacity for early warning , risk reduction and management of health risks (45). SAE methods will therefore provide a better understanding of the space-time variation of age-specific ARI prevalence, especially among children aged <5 years who are most affected by ARI morbidity and mortality.

1.5 Research question

Is there space and time variation of the prevalence and incidence of viral ARI in Kilifi, Kenya?

1.6 Study aim and Objectives

1.6.1 Study aim

To investigate and estimate the space-time variation of the prevalence and incidence of viral ARI morbidity in Kilifi.

1.6.2 Study Objectives

1. To describe the temporal patterns of ARI cases in Kilifi.
2. To estimate and predict crude and adjusted prevalence and incidence of Acute Respiratory Infections for sampled and unsampled locations respectively in Kilifi between 2015 and 2017.
3. To investigate the space time variations of the prevalence and incidence of viral Acute Respiratory infections in Kilifi.
4. To determine locations within Kilifi that are at high risk of ARI infections

References

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

2. Madhi SA, Klugman KP. Acute Respiratory Infections. In: Jamison DT, Feachem RG, Makgoba MW, Bos ER, Baingana FK, Hofman KJ, et al., editors. Disease and Mortality in Sub-Saharan Africa [Internet]. 2nd ed. Washington (DC): World Bank; 2006 [cited 2019 Oct 21]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK2283/

3. Williams BG, Gouws E, Boschi-Pinto C, Bryce J, Dye C. Estimates of world-wide distribution of child deaths from acute respiratory infections. The Lancet Infectious Diseases. 2002 Jan;2(1):25–32.

4. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Swartz S, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: A systematic analysis for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases. 2017 Nov;17(11):1133–61.

5. Beeler JA, Eichelberger MC. Influenza and respiratory syncytial virus (RSV) vaccines for infants: Safety, immunogenicity, and efficacy. Microbial Pathogenesis. 2013 Feb;55:9–15.

6. Murray CJL. Choosing indicators for the health-related SDG targets. The Lancet. 2015 Oct;386(10001):1314–7.

7. Feikin DR, Flannery B, Hamel MJ, Stack M, Hansen PM. Vaccines for Children in Low- and Middle-Income Countries. In: Black RE, Laxminarayan R, Temmerman M, Walker N, editors. Reproductive, Maternal, Newborn, and Child Health: Disease Control Priorities, Third Edition (Volume 2). Washington (DC): The International Bank for Reconstruction; Development / The World Bank; 2016.

8. Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the Sustainable Development Goals era: Time for a revolution. The Lancet Global Health. 2018 Nov;6(11):e1196–252.

9. Ghebreyesus TA. How could health care be anything other than high quality? The Lancet Global Health. 2018 Nov;6(11):e1140–1.

10. Ndabarora E, Chipps JA, Uys L. Systematic review of health data quality management and best practices at community and district levels in LMIC. Information Development. 2014 May;30(2):103–20.

11. Chris Skinner. The use of synthetic estimation techniques to produce small area estimates [Internet]. London, GB: Office of Population Censuses; Surveys. Social Survey Division; 1993 [cited 2019 Oct 21]. (New Methodology Series; vol. 1). Available from: https://eprints.soton.ac.uk/34702/

12. J. N. K. Rao, Isabel Molina. Small Area Estimation. Second. John Wiley & Sons, Inc; 2015.

13. Pfeffermann D. New Important Developments in Small Area Estimation. Statistical Science. 2013 Feb;28(1):40–68.

14. Rao J. Some methods for Small Area Estimation. Rivista Internazionale di Scienze Sociali. 2008;116(4):387–406.

15. AbouZahr C, Boerma T, Hogan D. Global estimates of country health indicators: Useful, unnecessary, inevitable? Global Health Action. 2017 May;10(sup1).

16. Rahman A. Estimating small area health-related characteristics of populations: A methodological review. Geospatial Health. 2017 May;

17. Nandram B. An Empirical Bayes Prediction interval for the finite population Small Area. Statistica Sinica. 1999;9(2):325–43.

18. Macharia PM, Giorgi E, Thuranira PN, Joseph NK, Sartorius B, Snow RW, et al. Sub national variation and inequalities in under-five mortality in Kenya since 1965. BMC Public Health. 2019 Feb;19(1):146.

19. Gutreuter S, Igumbor E, Wabiri N, Desai M, Durand L. Improving estimates of district HIV prevalence and burden in South Africa using small area estimation techniques. PLOS ONE. 2019 Feb;14(2):e0212445.

20. Kipruto H, Mung'atu J, Ogila K, Adem A, Mwalili S, Ong'ang'o J, et al. Spatial Temporal Modelling of Tuberculosis in Kenya Using Small Area Estimation. International Journal of Science and Research (IJSR). 2015;

21. Elbers C, Lanjouw JO, Lanjouw P. Micro–Level Estimation of Poverty and Inequality. Econometrica. 2003;71(1):355–64.

22. Bappi Kumar, Sumonkanti Das, Abdullah Al Islam, Luthful Alahi Kawsar. Prevalence of Diarrhoea and Acute Respiratory Infection at District Level in Bangladesh using Small Area Estimation Method. In Asian University for Women, Chittagong, Bangladesh; 2018. pp. 64–9. Available from: https://www.researchgate.net/publication/327656052_Prevalence_of_Diarrhoea_and_Acute_Respiratory_Infection_at_District_Level_in_Bangladesh_using_Small_Area_Estimation_Method

23. Achoki T, Miller-Petrie MK, Glenn SD, Kalra N, Lesego A, Gathecha GK, et al. Health disparities across the counties of Kenya and implications for policy makers, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. The Lancet Global Health. 2019 Jan;7(1):e81–95.

24. Muthoni KA, Ngesa OO. Determinants and Spatial Modeling of Acute Respiratory Infections (ARI) Among Children Less Than Five Years in Kenya. American Journal of Theoretical and Applied Statistics. 2017 Mar;6(2):123.

25. Sohnesen TP, Ambel AA, Fisker P, Andrews C, Khan Q. Small area estimation of child undernutrition in Ethiopian woredas. PLoS ONE. 2017 Apr;12(4).

26. Li Z, Hsiao Y, Godwin J, Martin BD, Wakefield J, Clark SJ. Changes in the spatial distribution of the under-five mortality rate: Small-area analysis of 122 DHS surveys in 262 subregions of 35 countries in Africa. PLoS ONE. 2019 Jan;14(1).

27. Berkowitz Z, Zhang X, Richards TB, Nadel M, Peipins LA, Holt J. Multilevel small-area estimation of colorectal cancer screening in the United States. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2018 Mar;27(3):245–53.

28. Dwyer-Lindgren L, Squires ER, Teeple S, Ikilezi G, Allen Roberts D, Colombara DV, et al. Small area estimation of under-5 mortality in Bangladesh, Cameroon, Chad, Mozambique, Uganda, and Zambia using spatially misaligned data. Population Health Metrics. 2018 Aug;16(1):13.

29. Nokes DJ, Ngama M, Bett A, Abwao J, Munywoki P, English M, et al. Incidence and Severity of Respiratory Syncytial Virus Pneumonia in Rural Kenyan Children Identified through Hospital Surveillance. Clinical Infectious Diseases. 2009 Nov;49(9):1341–9.

30. Berkley JA, Munywoki P, Ngama M, Kazungu S, Abwao J, Bett A, et al. Viral Etiology of Severe Pneumonia Among Kenyan Infants and Children. JAMA. 2010 May;303(20):2051–7.

31. Oketch JW, Kamau E, Otieno GP, Otieno JR, Agoti CN, Nokes DJ. Human metapneumovirus prevalence and patterns of subgroup persistence identified through surveillance of pediatric pneumonia hospital admissions in coastal Kenya, 2007-2016. BMC infectious diseases. 2019 Aug;19(1):757.

32. Onyango CO, Njeru R, Kazungu S, Achilla R, Bulimo W, Welch SR, et al. Influenza surveillance among children with pneumonia admitted to a district hospital in coastal Kenya, 2007-2010. The Journal of Infectious Diseases. 2012 Dec;206 Suppl 1:S61–67.

33. Kiyuka PK, Agoti CN, Munywoki PK, Njeru R, Bett A, Otieno JR, et al. Human Coronavirus NL63 Molecular Epidemiology and Evolutionary Patterns in Rural Coastal Kenya. The Journal of Infectious Diseases [Internet]. 2018 May [cited 2020 Jun 11];217(11):1728–39. Available from: https://academic.oup.com/jid/article/217/11/1728/4948258

34. Otieno GP, Murunga N, Agoti CN, Gallagher KE, Awori JO, Nokes DJ. Surveillance of endemic human coronaviruses (HCoV-NL63, OC43 and 229E) associated with pneumonia in Kilifi, Kenya. Wellcome Open Research. 2020 Jun;5:150.

35. Okiro EA, Ngama M, Bett A, Nokes DJ. The Incidence and Clinical Burden of Respiratory Syncytial Virus Disease Identified through Hospital Outpatient Presentations in Kenyan Children. PLOS ONE. 2012 Dec;7(12):e52520.

36. Onyango CO, Welch SR, Munywoki PK, Agoti CN, Bett A, Ngama M, et al. Molecular Epidemiology of Human Rhinovirus Infections in Kilifi, Coastal Kenya. Journal of Medical Virology. 2012 May;84:823–31.

37. Munywoki PK, Koech DC, Agoti CN, Cane PA, Medley GF, Nokes DJ. Continuous Invasion by Respiratory Viruses Observed in Rural Households During a Respiratory Syncytial Virus Seasonal Outbreak in Coastal Kenya. Clinical Infectious Diseases. 2018 Oct;67(10):1559–67.

38. Adema IW, Kamau E, Uchi Nyiro J, Otieno GP, Lewa C, Munywoki PK, et al. Surveillance of respiratory viruses among children attending a primary school in rural coastal Kenya. Wellcome Open Research. 2020 Apr;5:63.

39. Stein RT, Bont LJ, Zar H, Polack FP, Park C, Claxton A, et al. Respiratory syncytial virus hospitalization and mortality: Systematic review and meta-analysis. Pediatric Pulmonology. 2017 Apr;52(4):556–69.

40. Shi T, McAllister DA, O'Brien KL, Simoes EAF, Madhi SA, Gessner BD, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: A systematic review and modelling study. The Lancet. 2017 Sep;390(10098):946–58.

41. Acharya S, Lin V, Dhingra N. The role of health in achieving the sustainable development goals. Bulletin of the World Health Organization. 2018 Sep;96(9):591–591A.

42. Rudan I, Lawn J, Cousens S, Rowe AK, Boschi-Pinto C, Tomasković L, et al. Gaps in policy-relevant information on burden of disease in children: A systematic review. Lancet (London,

England). 2005 Jun;365(9476):2031–40.

43. Dawa JA, Chaves SS, Nyawanda B, Njuguna HN, Makokha C, Otieno NA, et al. National burden of hospitalized and non-hospitalized influenza-associated severe acute respiratory illness in Kenya, 2012-2014. Influenza and Other Respiratory Viruses. 2018;12(1):30–7.

44. Emukule GO, Khagayi S, McMorrow ML, Ochola R, Otieno N, Widdowson M-A, et al. The Burden of Influenza and RSV among Inpatients and Outpatients in Rural Western Kenya, 2009–2012. PLOS ONE [Internet]. 2014 Aug [cited 2020 May 10];9(8):e105543. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0105543

45. United Nations. Goal 3: Ensuring healthy lives and promoting well-being at all ages [Internet]. United Nations Sustainable Development. [cited 2020 Jul 21]. Available from: https://www.un.org/sustainabledevelopment/health/

CHAPTER 2: MANUSCRIPT

2.1 SUBMISSIBLE MANUSCRIPT

In this chapter data analysis results from models used to estimate the prevalence rates of viral ARIs in Kilifi are presented in form of a submissible manuscript. The manuscript is presented in the following format: Introducion, Methods, Results, Discussion, References and Supplementary Files. In addition to estimating the prevalence rates of viral ARI in Kilifi,the manusctipt further presents and discusses estimates of the incidence rates. The target journal for this manuscript is the Journal of Global Health Reports (https://www.joghr.org/) whose instruction for authors is in Appendix A11. Additional methods and results not covered in the draft manuscript are presented in the subsequent chapters.

Estimating the Burden of Viral Acute Respiratory Infections (ARI) in Kilifi, Kenya.

Grieven P. Otieno[1,2*], D. James Nokes[2,3], Ngianga-Bakwin Kandala[1], Joyce Nyiro[2], Innocent Maposa[1]

Authors' affiliations:

1. School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg South Africa
2. Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, Kenya.
3. School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, United Kingdom

∗ Corresponding author:
KEMRI Wellcome Trust Research Programme
P.O Box 230- 80108
Kilifi, Kenya.
E-Mail: grievenoti@gmail.com

Abstract

Introduction

Acute respiratory infections (ARIs) are major contributors to the overall disease burden especially among children in the developing world. We aimed to quantify the burden of ARI in a rural setting in Kenya using small area estimation (SAE) procedures.

Methods

We used data from a 2016-17 surveillance of ARI in 9 out-patient health facilities located within the Kilifi health and demographic surveillance system (KHDSS). Samples collected from participants of all ages were tested for respiratory syncytial virus (RSV A and B), human rhinovirus (HRV), human coronaviruses (HCoVs- OC43, NL63 and E229), influenza (A, B and C), parainfluenza viruses (PIV 1-4), adenovirus (ADV) and human metapneumovirus (hMPV), using a multiplex real-time polymerase chain reaction (PCR) assay. Crude and model-based SAE methods were used to estimate the prevalence and incidence of ARI in Kilifi using hierarchical Bayesian models, assuming various probability distributions, that accounted for variations through space and time. The deviance information criterion (DIC) was used to select models that offered best fit to the data.

Results

In period January 2016 to June 2017, crude estimtes showed that the highest prevalence and incidence rates were observed the under 5s (p-values <0.05) with the xception of FLU which was more prevalent among the over 5 year olds. The overall model-based prevalence of viral ARIs ranged from $<1\%$ for hMPV to 13.5% for HRV while the overall incidence rates ranged from 36 to 651 per 100000 population for hMPV and HRV respectively. Model-based prevalence estimates varied between pathogens and age category compared to crude estimates which showed highest prevalence among the under 5s for all pathogens except FLU. In addition, the under 5s ageband recorded the highest crude and model-based incidence rates of viral ARI. Model estimates showed that disease incidence for this age band ranged from 330 to 3249 per 100000 for HCoVs and HRV respectively. Spatio-temporal heterogeneity of both prevalence and incidence estimates was observed for all viruses by age categories.

Conclusion

Our models showed that a minimum of 36 and 330 cases per 100000 population for all ages and under 5s respectively were observed across 15 locations in a population-based surveillance

system over a period of 18 months. This gives an indication of considerable spatio-temporal heterogeneous burden of disease due to viral ARI, hence pointing to the need of potential future interventions at local levels. Targeted interventions among the under 5s, who had the highest disease incidence, such as incorporating existing and future vaccines into the Kenyan national immunisation programme could contribute to substantial reduction of ARI in this age band.

Introduction

Acute respiratory infections (ARIs) play an important role in morbidity and mortality especially among children aged <5 year (1,2). The bulk of this burden, mainly among under 5s, is borne by less developed countries (3). Several influenza vaccines are currently available (4) and development of a vaccine against respiratory syncytial virus (RSV) is ongoing (5). However, Africa has limited data on uptake and impact of the influenza vaccine (6) while most countries, such as Kenya, have not introduced the vaccine into their national immunisation programmes (7,8). Models on ARI's transmission patterns (9,10) and estimates of disease burden (11,12) are used to inform intervention measures such as vaccines.

In Kenya, the 2014 demographic health survey estimated a national prevalence of ARI symptoms (cough with shortness of breath, rapid and difficulty breathing) at 8.5%, which ranged between 1.5%-16.6% at county level (13). However, disaggregating such survey-based estimates may mask heterogeneity within lower sub-national levels and provide very small sample sizes or none for certain regions (14). A possible solution to the limitation of failing to obtain estimates at lower sub-national levels would be small area estimation (SAE) techniques. SAE is anchored on the premise of borrowing strength for areas with small sample sizes from other areas and/or time in order to produce reliable estimates (15). In this study, we aim to estimate the incidence and prevalence of ARI associated with viral pathogens in Kilifi, Kenya within a population-based surveillance system.

Methods

SPReD-KHDSS study

Data used in this analysis were adopted from a previous study which was aimed at studying the pathways of respiratory diseases (SPReD) within the Kilifi health and demographic surveillance system (KHDSS). SPReD-KHDSS was conducted at 9 government owned out-patient health facilities namely Matsangoni, Ngerenya, Mtondia, Sokoke, Mavueni, Jaribuni, Chasimba, Pingilikani and Junju (Figure 2.1). Patients aged above 7 days were recruited in this study if they presented with at least one ARI symptom of cough, sneezing, nasal congestion, difficulty breathing or increased respiratory rate for age that had lasted for less than a week (16). At each health facility, 15 samples per week were collected and tested for RSV (A and B), rhinovirus (HRV), human coronaviruses (HCoV- OC43, NL63 and E229), influenza (FLU A, B and C), parainfluenza viruses (PIV 1-4), adenovirus (ADV) and human metapneumovirus (hMPV), using a multiplex real-time polymerase chain reaction (PCR) assay (16,17). Data collection and sample processing was conducted from December 2015

to June 2017. However, our study utilises data from January 2016 to June 2017 because December 2015 was considered a pilot phase in the SPReD KHDSS study.

Statistical Methods

In this study, each of the 9 health facilities where the study took place is used to represent a location within the KHDSS out of a total of 15 locations (Figure 2.1). Our analysis estimates both crude and adjusted prevalence and incidence of ARI in Kilifi. The adjusted (model-based) estimates were obtained from hierarchical Bayes spatio-temporal SAE models. This analysis, anchored on SAE methods' abilities to borrow strength through space and time (15,18), was used to predict the prevalence and incidence of viral ARI for both sampled and unsampled locations within the KHDSS. A binomial specification was assumed for the observed counts of the positive cases for each virus pathogen. We obtained the mean spatial proportions (P_i) for each location ($i = 1, \ldots, 15$ locations) as estimates of the overall prevalence and incidence rates of viral ARI for each location. Prevalence estimates for each virus pathogen were expressed as percentages while incidence rates were expressed per 100,000 population. We present results for two age bands (0-59 months and over 5 year olds) and overall (all ages). In addition, results are presented at three timepoints namely; January to June 2016, July to December 2016 and January to June 2017. Incident rates were estimated using data on individuals who were KHDSS residents (19) at recruitment. In the KHDSS, a population of over 260000, of which about 18% are children aged under 5 years, has been under surveillance for the last 2 decades (19). The mean spatial prevalence and incidence rates, over the entire study period, are presented in maps of the KHDSS by location; the darker the colour the higher the rate.

Five models were fitted for each virus pathogen assuming Poisson, zero-inflated Poisson, binomial, zero-inflated binomial and logistic-normal specifications. Zero inflated Poisson and binomial models were used to cater for the possible imbalance due to many zeros. Hierarchical Bayesian statistical models were fitted to accommodate multiple levels of modelling. A spatio-temporal structure assuming the Besag-York-Mollie (BYM) conditional autoregressive (CAR) process was included to account for spatial correlation of the spatial random effects. A random walk process of lag 1 (RW1) was assumed for the area-time random effects and a space-time interaction term included in the models. In general, the geo-additive models, described above, used to estimate the proportions were specified as;

$$\hat{p_{it}} = g(p_{it}) = X_{it}^{'}\beta + u_i + B_t + v_{it} + e_{it} \quad (1)$$

where X_{it} is a vector of k auxilliary variables, β is a vector of coefficients, $u_i's$ are the spatial random effects, B_t is the time component and $v_{it}'s$ are the space-time random effects. Further,we split the spatial random effect into spatially correlated (structured) and uncorrelated (unstructured) effects. All models were fitted using the monthly mean age of study participants in years, mean monthly rainfall (millimetres) and number of health facilities in a location as auxiliary variables. These variables, included in the models to enhance small area predictions (15), were selected to represent ecological relationships rather than risk factors. A statistical description of the fitted models and methods is contained in supplementary file 1. The deviance information criteria (DIC) was used for model comparison whereby a model with the lowest DIC value was considered best fit. In addition to the DIC, the effective number of parameters (pD) was used as a measure of model complexity. For competing models with a difference of less than 2 DIC values, the model with the lowest DIC and pD values was selected as the one of best fit.

Data analysis was conducted in R version 3.6.1 (https://cran.r-project.org/) and models were fitted using the integrated nested Laplace approximation (INLA) (20) approach due to its advantage of computational efficiency using the R-INLA package.

Results

In the period January 2016 to June 2017, a total of 7985 patients who were symptomatic for ARI were recruited into the study of which 4295 (53.8%) were aged 0-59 months. There were 5249 (65.7%) patients who were residents of the KHDSS at recruitment of which 3066 (58.4%) were aged under 5 years. There were variations in the observed monthly number of positive cases and the peak months for each virus pathogen. In the first half of 2017, the number of positive cases faded off for each virus pathogen except RSV. The highest number of RSV cases were observed within the first quarter of the year reflecting the peak of the RSV epidemic. The remaining 6 virus pathogens were observed fairly throughout the study period, though in relatively low numbers in the first half of 2017 (Figure 2.2). Unadjusted estimates showed that children aged under 5 years had the highest prevalences for all viral pathogens (p-values <0.05) except for FLU which was highest in the over 5 years ageband. Similarly, the under 5s had the highest disease incidence compared to those aged above 5 years or overall (p-vaues <0.001). HRV and hMPV had the highest and lowest estimates respectively of disease burden (both incidence and prevalence) compared to the other pathogens (Table 2.1).

Model-based estimates of ARI prevalence,by age category, are presented in table 2.1. Across the two agebands and overall, HRV was the most prevalent viral infection while hMPV was

the least prevalent pathogen. When stratified by age, the most prevalent pathogens were HRV (13.5%) and HCoV (8.9%) for all ages, HRV(12.1%) and FLU (5.6%) for under 5s and HRV (14.1%) and PIV (6.7%) for over 5s. Contrary to crude estimates, the age band with the highest prevalence estimates varied across the different infections. In addition, there were fluctuations on the method (crude or model-based) that had the highest/lowest prevalence across virus pathogen and age.

Spatial variations of the disease prevalence for each viral pathogen were observed across the KHDSS. In the entire age spectrum, the mean spatial prevalence ranged from 4.5% to 11.5% for RSV, 13.3% to 22.7% for HRV,5.1% to 11.0% for HCoV, 4.0% to 9.2% for FLU, 3.5% to 5.7% for PIV, 0.9% to 6.8% for ADV and 0.6 to 5.2% for hMPV. The most prevalent infection ranged from one location to the other; for instance HRV, PIV and hMPV were most prevalent infections in Mtwapa while FLU was most prevalent in Sokoke (Figure 2.3). Similarly, among children aged <5 years there were spatial variations of the prevalence of the infections (Figure 2.4). These ranged from 4.7% to 11.9% for RSV, 13.2% to 21.4% for HRV, 1.3% to 6.4% for HCoVs, 5.4% to 8.8% for FLU, 3.6% to 6.6% for PIV, 0.4% to 2.9% for ADV and 0.5% to 36.2% for hMPV. Spatial variations of ARI prevalence were also observed among those aged above 5 years (results not shown).

The model-based (adjusted) incidence rates for each viral pathogen per 100000 population, presented in table 2.2, were varied by virus pathogen. The estimated incidence rates for all virus pathogens were higher in the under 5s age band compared to either the entire population or those aged above 5 years. In descending order, the infections with the highest incidence rates per 100000 children were HRV (3249), FLU (1006) and ADV (910) while the rest had less than 600 infections. In the entire population, highest incidence rates were predicted on HRV (651 cases) followed by ADV (267 cases) and HCoV (161 case). It was observed that the adjusted estimates of disease incidence were higher across age categories and virus type compared to crude estimates.

There was spatial variations in the estimated incidence rates across the 15 locations of the KHDSS. In the entire population, these rates ranged for each virus pathogen (Figure 2.5); RSV (12 to 375), HRV (28 to 812), HCoVs (3.4 to 278), FLU (1 to 277), PIV (13 to 176), ADV (1 to 288) and hMPV (11 to 78). Children aged under 5 years had higher mean spatial incidence rates across all virus pathogens compared to the general population. For these rates per 100000, the under 5s displayed spatial variations for each virus pathogen (Figure 2.6); RSV (172 to 2527), HRV (162 to 3554), HCoVs (10 to 791), FLU (9 to 1129), PIV (117 to 844), ADV (10 to 853) and hMPV (75 to 609). Similarly, spatial variations of incidence rates for over 5 year olds was observed for all virus pathogens (results not shown).

Predicted prevalence and incidence of RSV, PIV and ADV were higher in the first half of both years compared to the second half of the year 2016 for both age bands and overall. The lowest prevalence of HRV and HCoV were predicted to occur between January and June 2017 across all locations and ages. In general, both prevalence and incidence estimates varied by time across locations and age (figure 2.7).

A comparison of models by DIC and pD values for prevalence and incidence rates is provided in tables S1 and S2 respectively. Model estimates from Poisson and binomial models gave nearly equal point estimates of prevalence/incidence when these models offered the best fit compared to the rest and there was a small difference (less than 2 DIC values) between these two models. Sensitivity analysis was further conducted on both models to identify the one whose estimates were least affected by change of priors. This led to a choice of either the two models, when they provided the best fit, as identified in tables 2.2 and 2.3 and whose results were subsequently reported. Generally, the logit normal model performed best in estimating prevalence of most pathogens across age categories while the Poisson and binomial models provided the best fit for estimating incidence rates for majority of the viral pathogens across the 2 age bands and overall.

Discussion

The burden of ARI within communities is essential to understand the epidemiology of ARI, identify and implement required interventions. This study has presented an ecological comparisson of crude and model-based (adjusted) estimates of prevalence and incidence rates of viral ARI from a rural coastal region in Kenya using hierarchical Bayesian spatio-temporal SAE models. We observed that prevalence and incidence rates estimates varied across several respiratory viral pathogens. Crude estimates of disease prevalence were highest among under 5s compared to the over 5 age band and overall. However, this pattern was not observed on model-based estimated prevalence and variations across age and virus were observed. Both crude and adjusted incidence rates were highest in the under 5s age band compared to the entire population and those aged above 5 years. Model-based incidence rates per 100000 KHDSS population were higher than crude estimates for all viruses by age. Our study findings present the potential benefit of SAE methods on estimating the burden of SAE by accounting for other ecological factors otherwise not considered when calculating crude estimates. We have shown instances whereby crude prevalence estimates were either under or over estimated compared to model-based estimates while crude incidence rates gave an under estimate of disease burden. In addition, using model-based SAE techiques we predicted estimates of disease prevalence and incidence for all viruses across the 2 age bands and overall

for both sampled and unsampled locations of the KHDSS as well as their temporal variations.

Our model-based estimates of the incidence rate of RSV among under 5s are lower than those of a previous study conducted at the Kilifi county hospital (767 per 100000) (21) and other studies in western Kenya (22.3 per 1000 children) and Egypt (2300 per 100000) for RSV associated influenza like illness (22,23). Consistent with our estimates, the same studies in western Kenya and Egypt reported higher incidence rates for RSV among children under the age of 5 years compared to other age bands (22,23). The high incidence rate estimates, especially among children, highlights the important contribution made by RSV to disease burden in children as previously reported (24). FLU incidence rates that we reported in Kilifi (90 for the entire population and 1006 in the under 5s population) fell within the range of estimates from western Kenya (7.2 and 21.8 per 1000 for the entire population and under 5s respectively) (22) and the Democratic republic of Congo (1205 and 1464 for the entire population and under 5s respectively) (25).

Visual inspections of the maps of the estimated prevalence and incidence rates for each virus pathogen reveal their geographical heterogeneity that is otherwise concealed by the overall rates. The spatial variations for all virus pathogens, observed in the entire population (overall) and separately for both age bands (0-59 months and those aged 5 years), may be attributed to different population structures, contact mixing and major economic activities in each location. Previous studies in the under 5s populaton in Kenya and Ethiopia have also reported geographical variations of the prevalence of ARI (26,27). We did not observe a concentration of high prevalence or incidence rates in the central, northern or southern parts of the KHDSS hence no indications of higher risk of ARI in concentrated in a particular region of the KHDSS.

Our study findings are limited to medically attended ARI infections which may be affected by health seeking behaviour of people living in the KHDSS (28). However, this was not accounted for in our model-based estimates due to unavailable data on this indicator across age and location. In addition, sampling weights could have been used to adjust for study design (29) and therefore potentially influence our model estimates but were not calculated due to limited data on the study design. To mitigate against this limitation, we included the number of health facilities per location which was the only available design variable as an auxiliary variable in the models. This study has potentially shown the possibility of drawing inaccurate inferences from crude estimates of disease burden especially under complex surveys which could be mitigated through SAE methods. However, SAE is an advanced statistical methodology which requires high computational capability hence may limit replicating our study methods to other settings.

We have presented evidence of considerable burden of ARI, especially among children aged 0-59 months, with varying prevalence and incidence rates for different viral ARI pathogens. Moreover, we have observed spatio-temporal heterogeneity for each viral ARI pathogen across two age bands and overall. Our results can be used to inform localised targeted interventions or surveillance by public health policy makers for sustained monitoring and evaluation of emerging patterns and trends of ARI that may impact future interventions such as vaccinations and therapy.

Conflict of Interest:

The authors report no potential conflicts of interest.

Grant information

The SPReD-KHDSS study was supported by the Wellcome Trust (102975,203077). GPO was supported through the DELTAS Africa Initiative (107754/Z/15/Z-DELTAS Africa SSACAB) and additional support from DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [203077/Z/16/Z] and the UK government.The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Acknowledgment

We thank the study participants and parents/ guardians of minors who participated, clinical and SPReD-KHDSS study team, led by Joyce Nyiro, who were involved at various stages of this study including sample collection and processing.

References

1. United Nations Children's Fund. The state of the world's children 2016: A fair chance for every child. [Internet]. New York: UNICEF; 2016 [cited 2019 Feb 20]. Available from: https://www.unicef.org/publications/files/UNICEF_SOWC_2016.pdf

2. Pinzón-Rondón ÁM, Aguilera-Otalvaro P, Zárate-Ardila C, Hoyos-Martínez A. Acute respiratory infection in children from developing nations: A multi-level study. Paediatrics and International Child Health. 2016 Apr;36(2):84–90.

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

4. Sautto GA, Kirchenbaum GA, Ross TM. Towards a universal influenza vaccine: Different approaches for one goal. Virology Journal. 2018 Dec;15(1):1–12.

5. Neuzil KM. Progress toward a Respiratory Syncytial Virus Vaccine. Clinical and Vaccine Immunology. 2016 Mar;23(3):186–8.

6. Madhi SA, Klugman KP. Acute Respiratory Infections. In: Jamison DT, Feachem RG, Makgoba MW, Bos ER, Baingana FK, Hofman KJ, et al., editors. Disease and Mortality in Sub-Saharan Africa. 2nd ed. Washington (DC): World Bank; 2006.

7. Emukule GO, Ndegwa LK, Washington ML, Paget JW, Duque J, Chaves SS, et al. The cost of influenza-associated hospitalizations and outpatient visits in Kenya. BMC Public Health. 2019 May;19(3):471.

8. Adetifa IMO, Karia B, Mutuku A, Bwanaali T, Makumi A, Wafula J, et al. Coverage and timeliness of vaccination and the validity of routine estimates: Insights from a vaccine registry in Kenya. Vaccine. 2018 Dec;36(52):7965–74.

9. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. American Journal of Epidemiology. 2006 Nov;164(10):936–44.

10. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLOS Medicine. 2008 Mar;5(3):e74.

11. Shi T, Arnott A, Semogas I, Falsey AR, Openshaw P, Wedzicha JA, et al. The Etiological Role of Common Respiratory Viruses in Acute Respiratory Infections in Older Adults: A Systematic Review and Meta-analysis. The Journal of Infectious Diseases. 2019 Mar;

12. Bellos A, Mulholland K, O'Brien KL, Qazi SA, Gayer M, Checchi F. The burden of acute respiratory infections in crisis-affected populations: A systematic review. Conflict and Health. 2010 Feb;4:3.

13. Kenya National Bureau of Statistics, Ministry of Health/Kenya, National AIDS Control Council/Kenya, Kenya Medical Research Institute, National Council for Population and

Development/Kenya. Kenya Demographic and Health Survey 2014 [Internet]. Rockville, MD, USA; 2015 [cited 2020 Feb 22]. Available from: http://dhsprogram.com/pubs/pdf/FR308/FR308.pdf

14. Chris Skinner. The use of synthetic estimation techniques to produce small area estimates. London, GB: Office of Population Censuses; Surveys. Social Survey Division; 1993. (New Methodology Series).

15. J. N. K. Rao, Isabel Molina. Small Area Estimation. Second. John Wiley & Sons, Inc; 2015.

16. Nyiro JU, Munywoki P, Kamau E, Agoti C, Gichuki A, Etyang T, et al. Surveillance of respiratory viruses in the outpatient setting in rural coastal Kenya: Baseline epidemiological observations. Wellcome Open Research. 2018 Jul;3:89.

17. Kamau E, Agoti CN, Lewa CS, Oketch J, Owor BE, Otieno GP, et al. Recent sequence variation in probe binding site affected detection of respiratory syncytial virus group B by real-time RT-PCR. Journal of Clinical Virology. 2017 Mar;88:21–5.

18. Pfeffermann D. New Important Developments in Small Area Estimation. Statistical Science. 2013 Feb;28(1):40–68.

19. Scott JAG, Bauni E, Moisi JC, Ojal J, Gatakaa H, Nyundo C, et al. Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). International Journal of Epidemiology. 2012 Jun;41(3):650–7.

20. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2009;71(2):319–92.

21. Okiro EA, Ngama M, Bett A, Nokes DJ. The Incidence and Clinical Burden of Respiratory Syncytial Virus Disease Identified through Hospital Outpatient Presentations in Kenyan Children. PLoS ONE [Internet]. 2012 Dec [cited 2020 May 10];7(12). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530465/

22. Emukule GO, Khagayi S, McMorrow ML, Ochola R, Otieno N, Widdowson M-A, et al. The Burden of Influenza and RSV among Inpatients and Outpatients in Rural Western Kenya, 2009–2012. PLOS ONE [Internet]. 2014 Aug [cited 2020 May 10];9(8):e105543. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0105543

23. Rowlinson E, Dueger E, Taylor T, Mansour A, Van Beneden C, Abukela M, et al. Incidence and Clinical Features of Respiratory Syncytial Virus Infections in a Population-Based Surveillance Site in the Nile Delta Region. The Journal of Infectious Diseases. 2013 Dec;208(suppl_3):S189–96.

24. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, Singleton RJ, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: A systematic review and meta-analysis. Lancet (London, England). 2010 May;375(9725):1545–55.

25. Babakazo P, Lubula L, Disasuani W, Manya LK, Nkwembe E, Mitongo N, et al. The national and provincial burden of medically attended influenza-associated influenza-like illness and severe acute respiratory illness in the Democratic Republic of Congo, 2013-2015. Influenza and Other Respiratory Viruses [Internet]. 2018 [cited 2020 May 10];12(6):695–705. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/irv.12601

26. Muthoni KA, Ngesa OO. Determinants and Spatial Modeling of Acute Respiratory Infections (ARI) Among Children Less Than Five Years in Kenya. American Journal of Theoretical and Applied Statistics. 2017 Mar;6(2):123.

27. Amsalu ET, Akalu TY, Gelaye KA. Spatial distribution and determinants of acute respiratory infection among under-five children in Ethiopia: Ethiopian Demographic Health Survey 2016. PLoS ONE [Internet]. 2019 Apr [cited 2020 Jun 9];14(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6476529/

28. Moïsi JC, Nokes DJ, Gatakaa H, Williams TN, Bauni E, Levine OS, et al. Sensitivity of hospital-based surveillance for severe disease: A geographic information system analysis of access to care in Kilifi district, Kenya. Bulletin of the World Health Organization. 2011 Feb;89(2):102–11.

29. Pfeffermann D. The Role of Sampling Weights When Modeling Survey Data. International Statistical Review / Revue Internationale de Statistique. 1993;61(2):317–37.

Table 2.1: Crude prevalence and incidence estimates for each virus pathogen stratified by age category and overall.

	Age	Denominator	RSV	HRV	HCoV	FLU	PIV	ADV	hMPV
Prevalence	0-59 months	4295	9.31	16.81	5.61	4.84	6.08	3.84	1.70
%	≥5 yrs	3690	2.47	11.11	4.12	6.75	2.93	0.84	0.46
	Overall	7985	6.15	14.18	4.92	5.72	4.62	2.45	1.13
	χ^2		161.2	53.0	9.4	13.4	44.7	74.7	27.3
	p-value		<.001	<.001	0.009	0.001	<.001	<.001	<.001
Incidence	0-59 months	69410	412.04	746.29	252.13	218.99	288.14	159.92	74.92
per 100,000	≥ 5 yrs	364276	14.82	65.06	24.71	40.90	19.49	6.04	1.92
	Overall	433683	78.40	174.09	61.10	69.41	62.49	30.67	13.60
	χ^2		1566.9	1927.1	600.9	296.3	882.8	678.0	283.8
	p-value		<.001	<.001	<.001	<.001	<.001	<.001	<.001

[1]The column named "Denominator" indicates the number of samples tested for prevalence estimates and population estimates for incidence rates estimates.

Table 2.2: Estimates of ARI prevalence and 95% credible intervals (CIs) for each virus pathogen tested by age category

		All ages	Under 5s	Over 5s
RSV	Prevalence (95% CI)	4.32 (1.92-9.46)	3.54 (1.25-9.78)	2.74 (0.93-7.83)
	Model	Poi	Bin	Bin
	DIC (pD)	465.06 (44.48)	458.45 (54.40)	459.44 (54.08)
HRV	Prevalence (95% CI)	13.53 (8.46-20.77)	12.14 (7.14-19.67)	14.13 (8.40-22.69)
	Model	Logit n	Logit n	Logit n
	DIC (pD)	344.91 (19.73)	346.70 (19.56)	334.73 (19.75)
HCoV	Prevalence (95% CI)	8.91 (4.61-16.48)	2.62 (1.18-5.60)	7.19 (3.32-14.82)
	Model	Logit n	Logit n	Logit n
	DIC (pD)	482.94 (16.10)	484.78 (16.11)	477.43 (16.06)
FLU	Prevalence (95% CI)	3.94 (1.78-8.48)	5.63 (2.22-13.25)	5.42 (2.12-13.20) [2]
	Model	Logit n	Logit n	Logit n
	DIC (pD)	488.51 (44.13)	488.38 (47.93)	479.47 (47.50)
PIV	Prevalence (95% CI)	3.94 (1.78-8.48)	2.51 (1.10-5.70)	6.74 (2.99-14.42)
	Model	Poi	Poi	Logit n
	DIC (pD)	482.72 (18.42)	482.08 (18.56)	476.78 (17.34)
ADV	Prevalence (95% CI)	7.36 (3.30-15.72)	2.33 1.99 (0.83-4.66)	3.74 (1.47-9.42)
	Model	Bin	Poi	Bin
	DIC (pD)	445.11 (14.56)	444.97 (14.54)	445.04 (14.67)
hMPV	Prevalence (95% CI)	0.75 (0.14-3.90)	0.94 (0.18-4.46)	0.18 (0.03-1.00)
	Model	Poi	Poi	Poi
	DIC (pD)	269.06 (14.84)	276.73 (18.14)	278.65 (18.20)

[2]The rows named "Model" indicate the model which offered the best fit on the data by DIC values; Logit n= Logit normal, Bin=Binomial, Poi=Poisson, See statistical methods for details.

Table 2.3: Incidence rate estimates per 100000 mid-year population and 95% credible intervals (CIs) for each virus pathogen tested by age category

		All ages	Under 5s	Over 5s
RSV	Incidence (95% CI)	85.09 (28.44-237.78)	501.67 (134.88-1706.98)	14.88 (2.07-96.67)
	Model	Bin	Poi	Bin
	DIC (pD)	412.69 (46.89)	373.72 (42.65)	213.02 (14.83)
HRV	Incidence (95% CI)	650.91 (258.73-1614.67)	3248.78 (1273.14-7879.38)	155.42 (36.48-609.99)
	Model	Logit n	Logit n	Poi
	DIC (pD)	381.26 (22.26)	442.85 (20.08)	412.47 (21.35)
HCoV	Incidence (95% CI)	160.04 (53.53-462.96)	329.61 (94.55-1041.28)	159.49 (34.06-695.75)
	Model	Poi	Poi	Poi
	DIC (pD)	434.26 (48.65)	386.61 (18.89)	278.87 (15.34)
FLU	Incidence (95% CI)	89.43 (27.34-293.23)	1005.32 (278.27-3334.51)	189.78 (39.38-910.21) [3]
	Model	Bin	Poi	Bin
	DIC (pD)	485.21 (66.98)	397.67 (19.46)	370.26 (54.02)
PIV	Incidence (95% CI)	101.34 (39.17-261.59)	578.83 (216.44-1535.82)	33.62 (5.01-186.54)
	Model	Poi	Poi	Poi
	DIC (pD)	440.76 (20.05)	401.21 (18.23)	237.07 (14.18)
ADV	Incidence (95% CI)	267.10 (81.22-862.69)	909.31 (304.55-2758.60)	164.20 (18.35-1312.96)
	Model	Bin	Zib	Poi
	DIC (pD)	365.60 (19.73)	353.090 (10.89)	140.41 (4.85)
hMPV	Incidence (95% CI)	35.75 (4.29-282.46)	476.25 (66.45-2924.85)	13.28 (0.24-474.53)
	Model	Poi	Poi	Poi
	DIC (pD)	225.55 (19.78)	235.36 (8.88)	71.01 (7.04)

[3]The rows named "Model" indicate the model which offered the best fit on the data by DIC values; Logit n= Logit normal, Bin=Binomial, Zib=Zero-inflated binomial, Poi=Poisson. See statistical methods for details. The models in bold offered the best fit

Table S.1: Deviance information criteria (DIC) and effective number of parameters (pD) used for comparisson of models used estimate the prevalence of viral ARI in Kilifi.

Model		RSV	HRV	HCoV	FLU	PIV	ADV	hMPV
All ages								
Poi	DIC	**465.06**	642.94	494.87	548.10	**482.72**	453.33	**279.41**
	pD	**44.48**	21.12	35.96	67.03	**18.42**	7.13	**18.52**
Zip	DIC	469.06	737.85	541.25	632.71	523.76	454.11	269.06
	pD	44.33	32.72	20.07	21.84	19.22	14.66	14.84
Bin	DIC	456.51	640.49	489.98	540.48	483.24	**445.11**	280.25
	pD	51.96	29.71	44.71	71.33	18.550	**14.56**	18.47
Zib	DIC	574.93	736.09	541.84	547.90	634.12	453.49	291.23
	pD	19.07	29.99	19.48	60.73	17.42	14.41	9.49
Logit n	DIC	532.94	**344.91**	**482.94**	**488.51**	488.00	516.14	696.65
	pD	42.38	**19.73**	**16.10**	**44.13**	17.37	6.52	7.84
Under 5s								
Poi	DIC	467.30	643.63	495.15	547.63	**482.08**	**444.97**	**276.73**
	pD	47.42	21.01	42.13	67.77	**18.56**	**14.54**	**18.14**
Zip	DIC	576.46	804.99	667.76	638.16	483.89	447.85	286.90
	pD	19.81	31.95	19.55	21.83	18.64	13.44	9.69
Bin	DIC	**458.45**	639.53	489.04	540.06	482.57	445.03	277.63
	pD	**54.40**	32.48	48.20	71.91	18.69	14.83	18.10
Zib	DIC	461.90	736.47	542.40	645.57	524.06	455.43	288.06
	pD	54.42	29.94	19.45	21.05	18.62	14.36	9.59
Logit n	DIC	535.57	**346.70**	**484.78**	**488.38**	488.43	522.17	679.71
	pD	44.60	**19.56**	**16.11**	**47.93**	17.04	7.27	14.05
Over 5s								
Poi	DIC	468.22	642.26	494.94	547.91	482.29	445.26	**278.65**
	pD	46.92	21.16	43.10	67.81	18.39	14.74	**18.20**
Zip	DIC	572.15	737.160	544.35	638.20	523.67	456.29	286.77
	pD	19.87	32.71	20.07	21.80	19.16	14.56	9.56
Bin	DIC	**459.44**	638.64	488.91	540.17	482.78	**445.04**	279.36
	pD	**54.08**	29.57	48.71	71.98	18.52	**14.67**	18.03
Zib	DIC	585.24	798.73	545.17	645.66	523.63	447.85	287.84
	pD	18.62	28.54	19.47	21.02	18.58	13.55	9.46
Logit n	DIC	525.13	**334.73**	**477.43**	**479.47**	**476.78**	514.77	684.90
	pD	44.07	**19.75**	**16.06**	**47.50**	**17.34**	7.07	6.88

[4]

[4]The column named "Model" indicates the model names; Poi=Poisson, Zip=Zip=Zero-inflated Poisson, Bin=Binomial, Zib=Zero-inflated binomial, Logit n= Logit normal. See statistical methods for details. The logit normal model gave the best fit for most viral pathogens across age categories. The models in bold offered the best fit

Table S.2: Deviance information criteria (DIC) and effective number of parameters (pD) used for comparisson of models used estimate incidence rates of viral ARI per 100000 population in Kilifi.

Model		RSV	HRV	HCoV	FLU	PIV	ADV	hMPV
All ages								
Poi	DIC	413.17	599.96	**434.26**	485.25	**440.76**	365.70	**225.55**
	pD	46.77	24.85	**48.65**	66.56	**20.05**	19.78	**19.78**
Zip	DIC	523.43	698.68	480.14	493.83	480.68	388.54	248.52
	pD	16.50	27.50	16.35	48.37	16.80	11.90	9.92
Bin	DIC	**412.69**	601.48	435.73	**485.21**	440.97	**365.60**	225.56
	pD	**46.89**	31.87	49.00	**66.98**	20.20	**19.73**	19.80
Zib	DIC	523.95	698.88	480.24	558.67	480.73	388.55	253.24
	pD	16.47	27.41	16.34	17.73	16.78	11.89	8.10
Logit n	DIC	568.14	**381.26**	540.57	537.04	513.00	578.45	722.20
	pD	36.34	22.26	**18.96**	39.96	17.61	8.86	9.36
Under 5s								
Poi	DIC	**373.72**	545.18	**386.61**	**397.67**	**401.21**	355.68	**235.36**
	pD	**42.65**	22.52	**18.89**	**19.46**	**18.23**	8.16	**8.88**
Zip	DIC	487.39	631.13	407.50	616.34	425.89	353.13	566.67
	pD	15.31	23.45	13.18	13.38	14.33	10.93	7.17
Bin	DIC	371.98	545.40	386.79	397.89	401.30	355.81	235.56
	pD	43.86	22.55	18.97	19.49	18.12	8.14	8.89
Zib	DIC	489.47	631.91	407.76	617.07	426.10	**353.09**	568.50
	pD	15.13	23.28	13.12	13.20	14.27	**10.89**	7.08
Logit n	DIC	589.17	**422.85**	593.20	612.48	557.37	598.23	715.20
	pD	34.46	**20.08**	15.84	9.02	16.65	8.16	8.90
Over 5s								
Poi	DIC	212.95	**412.47**	**278.87**	371.50	**237.07**	**140.41**	**71.01**
	pD	14.88	**21.35**	**15.34**	54.66	**14.18**	**4.85**	**7.04**
Zip	DIC	664.99	466.59	293.66	411.01	257.90	677.23	550.49
	pD	10.40	15.57	9.76	12.18	8.80	7.72	3.23
Bin	DIC	**213.02**	412.52	278.95	**370.26**	237.19	140.42	77.98
	pD	**14.83**	21.39	15.41	**54.02**	14.33	4.88	10.40
Zib	DIC	242.46	466.64	293.68	411.14	257.89	136.36	551.45
	pD	8.02	15.54	9.76	12.17	8.79	5.74	3.20
Logit n	DIC	706.66	518.11	645.22	605.01	661.40	746.38	768.73
	pD	13.23	18.22	9.14	17.67	6.24	5.44	5.53

[5]

[5]The column named "Model" indicates the model names; Poi=Poisson, Zip=Zip=Zero-inflated Poisson, Bin=Binomial, Zib=Zero-inflated binomial, Logit n= Logit normal. See statistical methods for details. The Poisson and binomial models gave the best fit for most viral pathogens across age categories. The models in bold offered the best fit.

2.1.2 Manuscript Figures

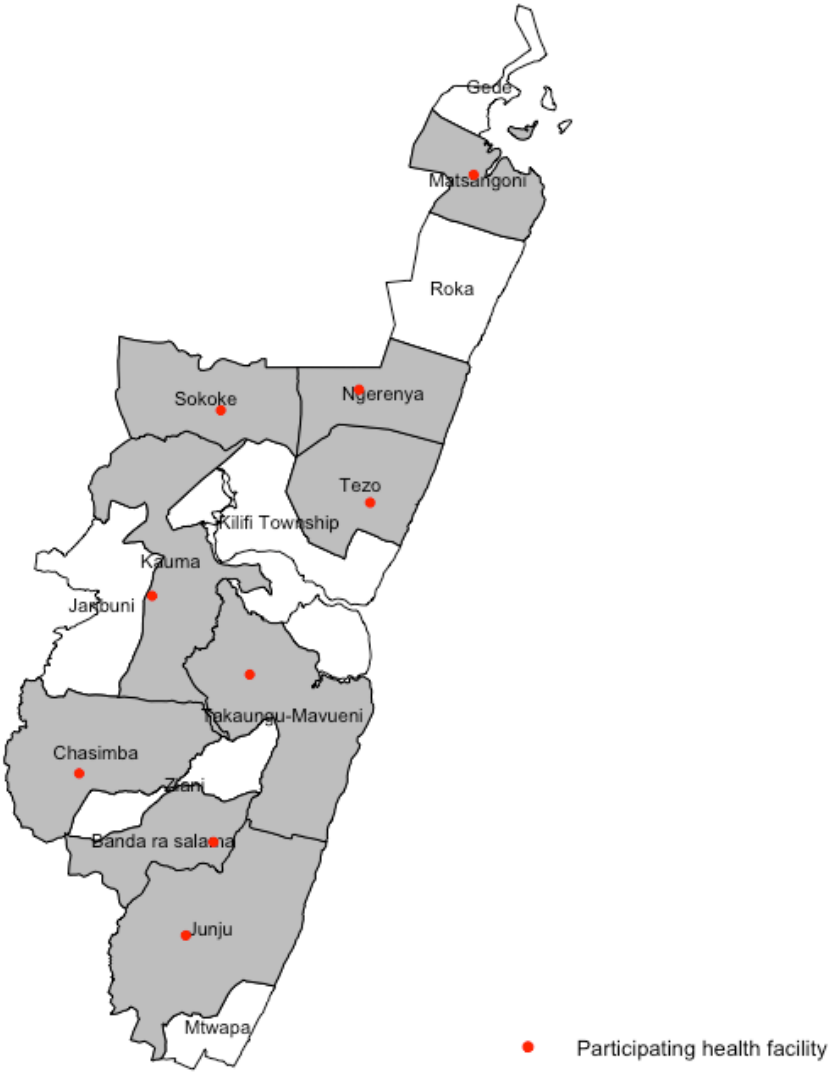


Figure 2.1: Map showing administrative locations within the KHDSS. The grey fill highlights locations where the study was conducted and the red dots show the participating health facilities within the locations.

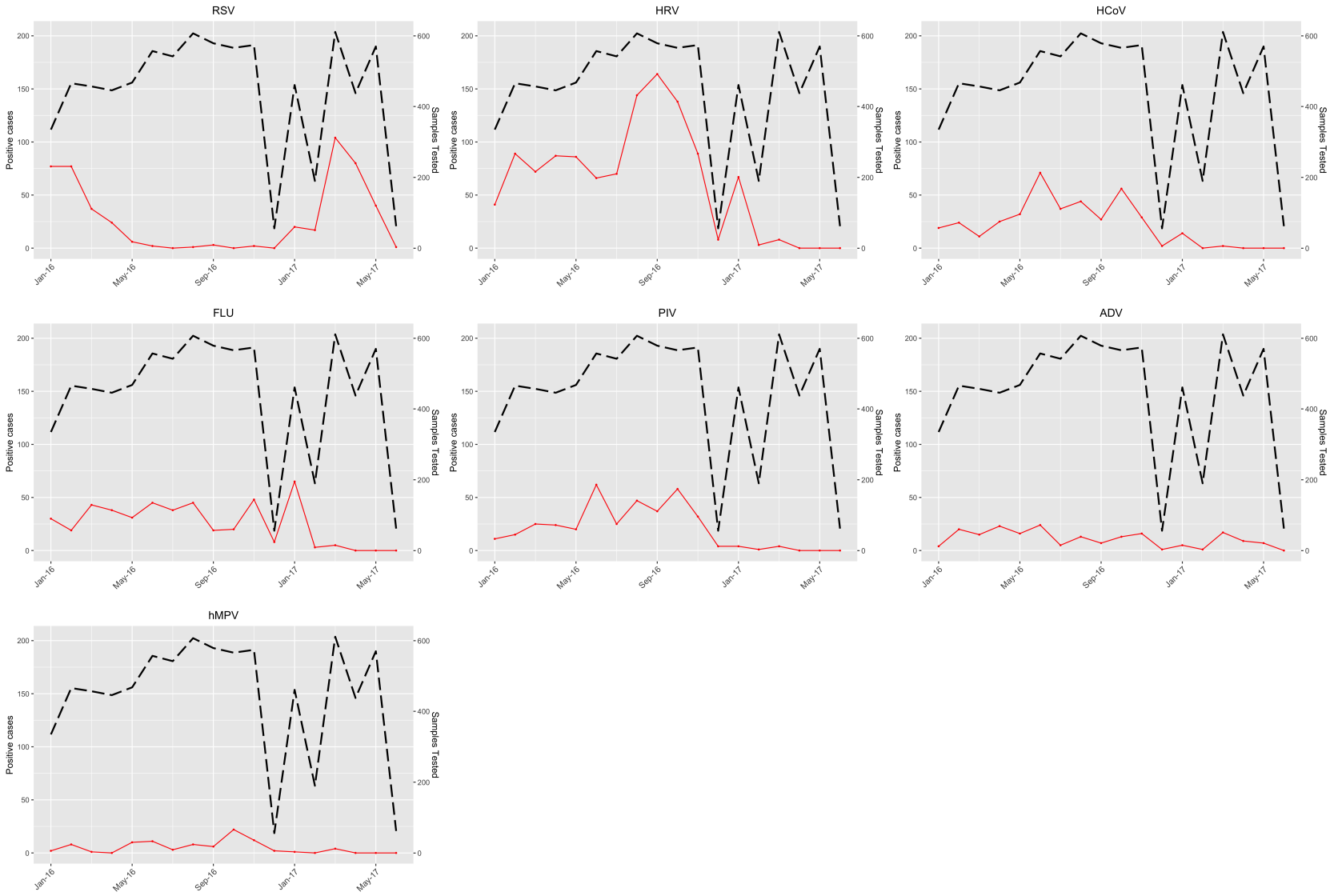


Figure 2.2: Panel of graphs showing the trends of each virus pathogen. The primary y-axis denotes the monthly numbers positive cases for each virus pathogen shown by the red solid line. The secondary y-axis denotes the number of samples tested shown by the black dashed line.

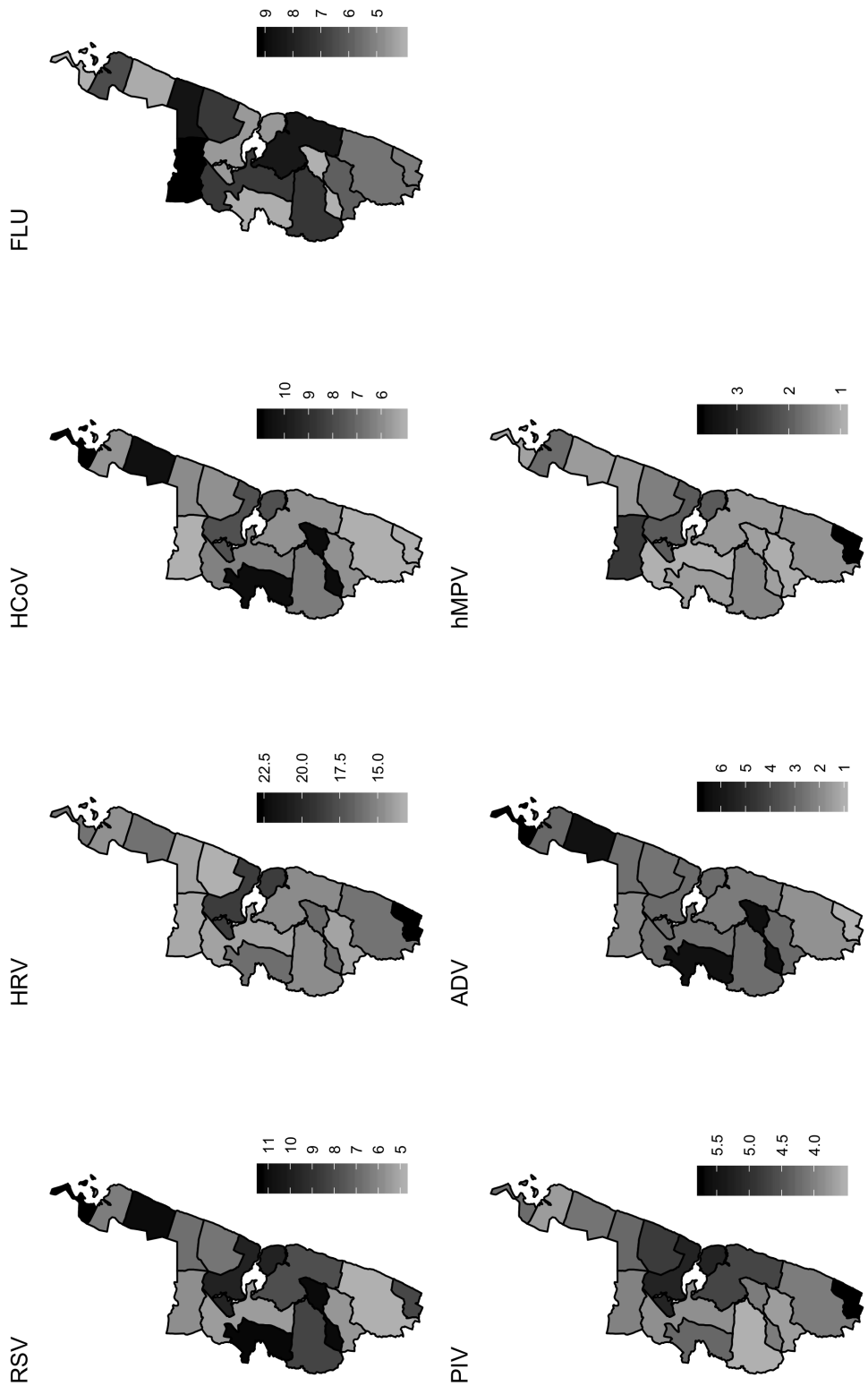


Figure 2.3: Panel of maps showing the estimated prevalence by virus pathogen for all ages within the KHDSS.

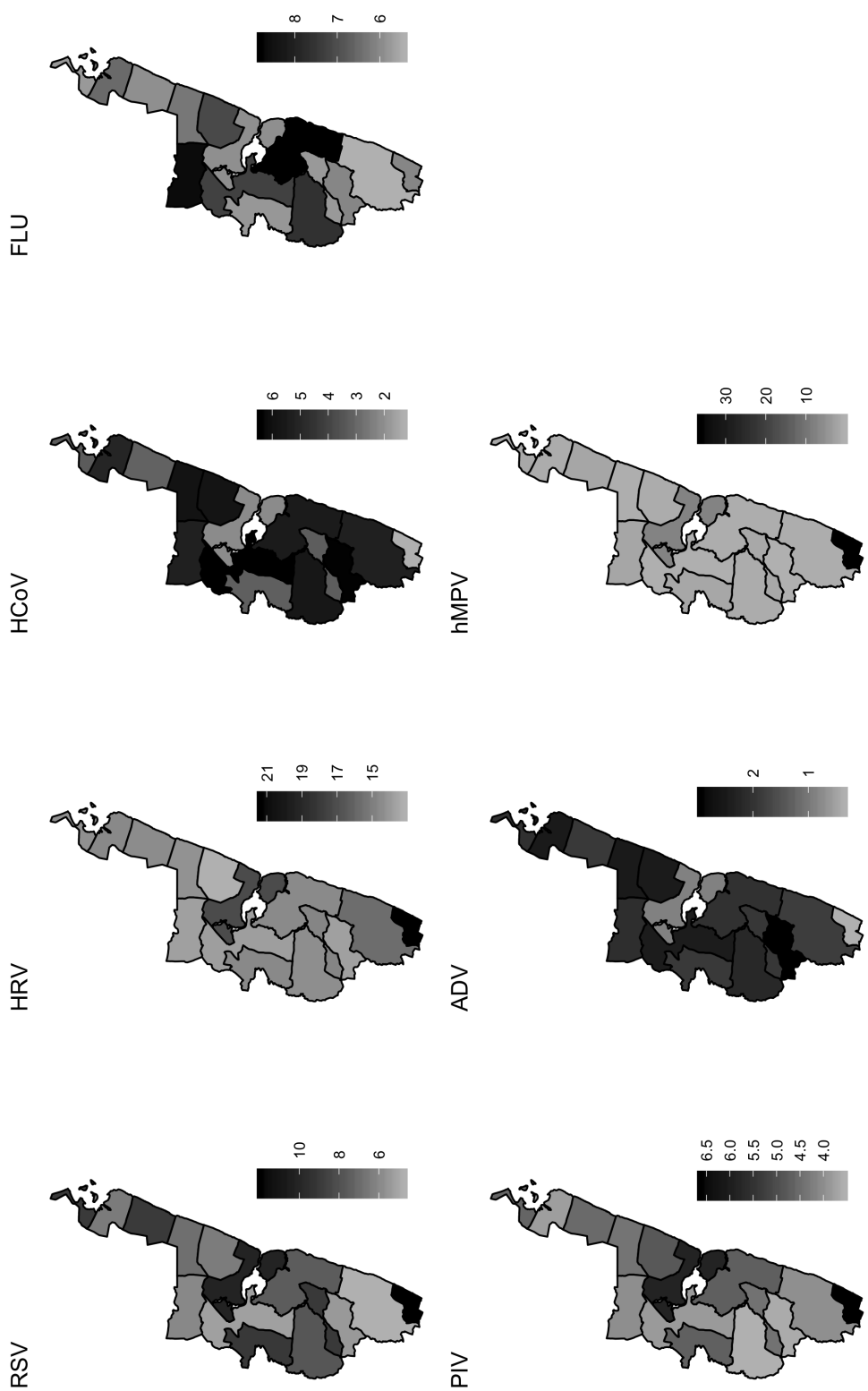


Figure 2.4: Panel of maps showing the estimated prevalence by virus pathogen children aged 0-59 months within the KHDSS.

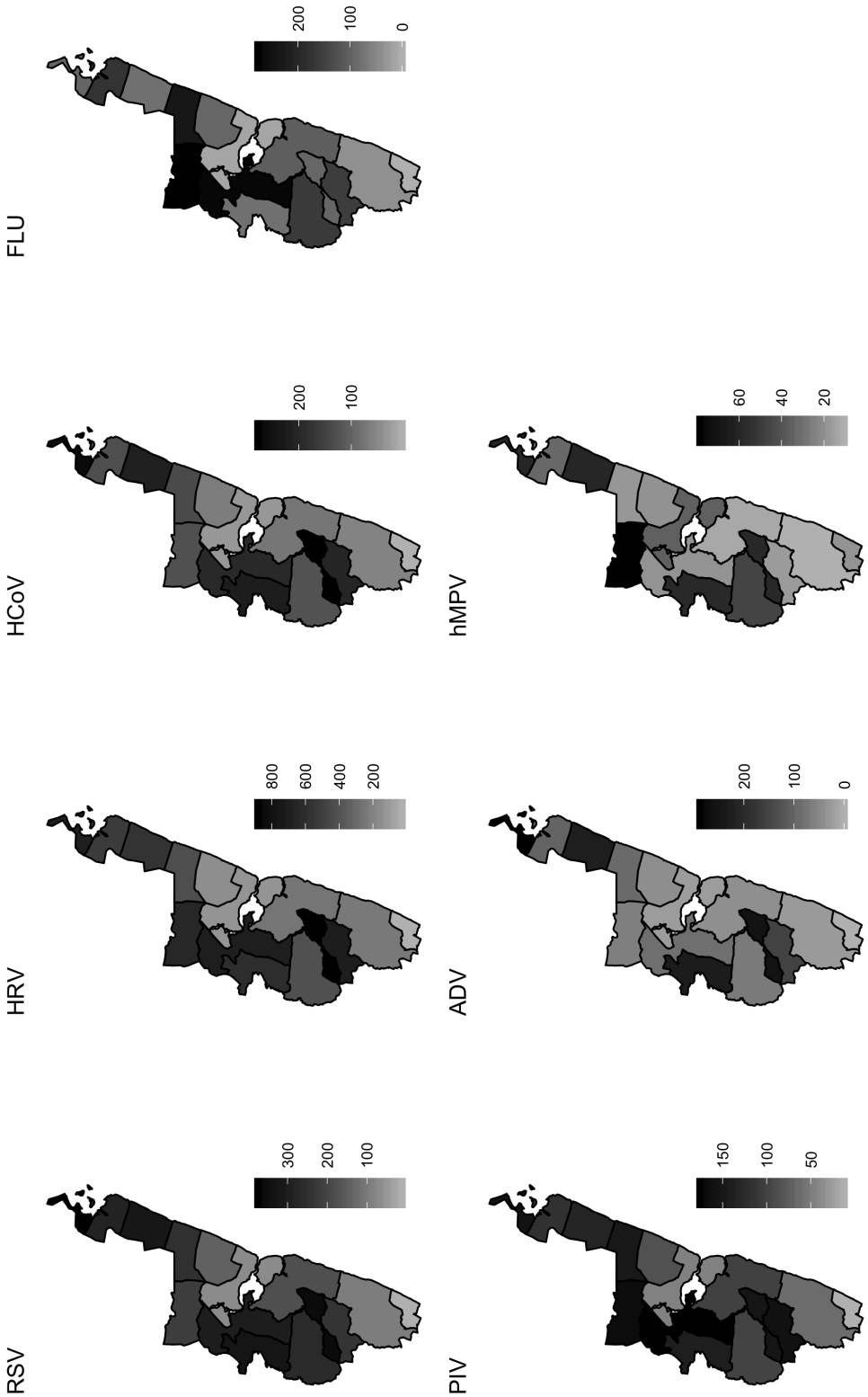


Figure 2.5: Panel of maps showing the estimated incidence rates by virus pathogen for all ages within the KHDSS.

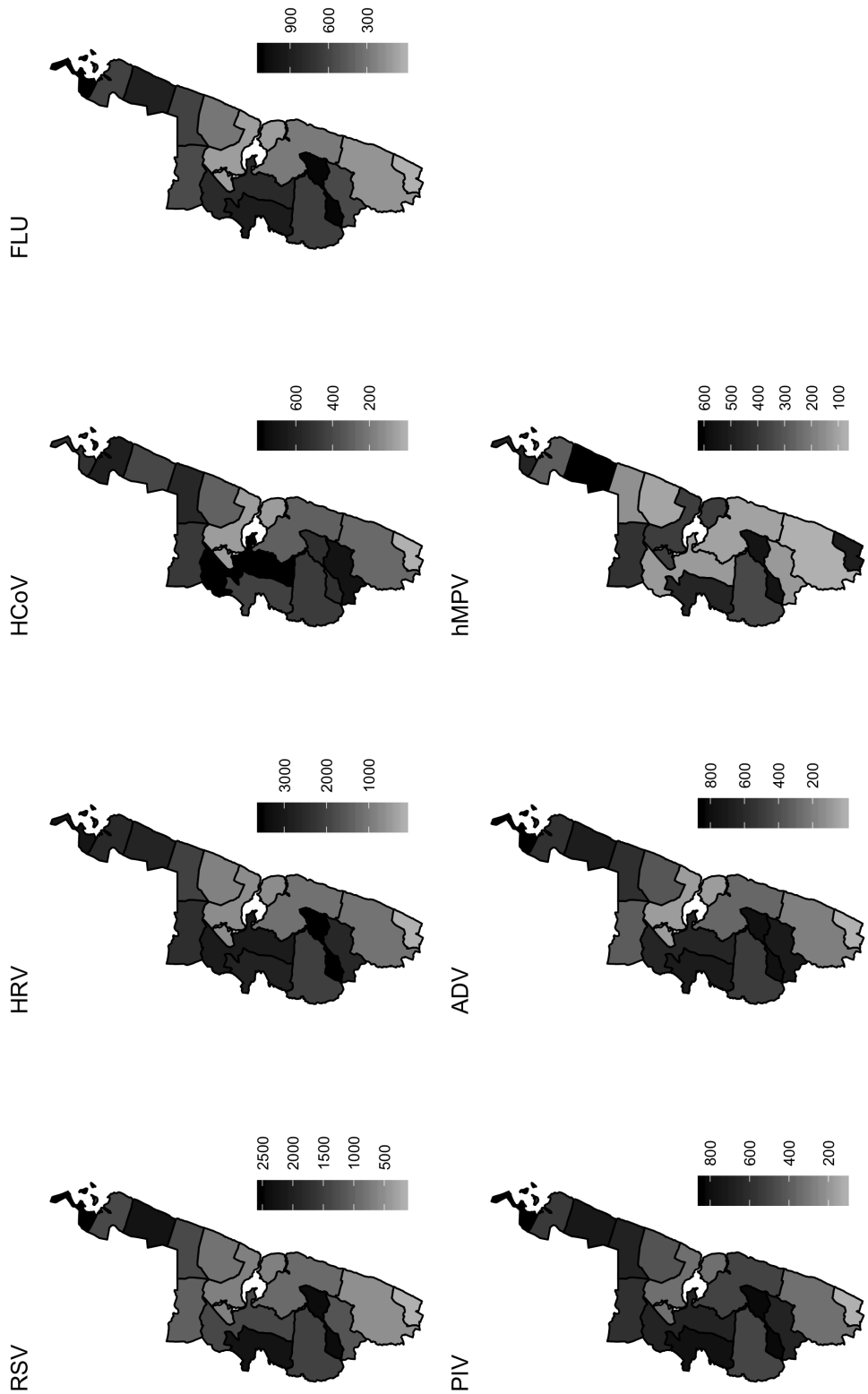


Figure 2.6: Panel of maps showing the estimated incidence rates per 100000 population by virus pathogen for children aged 0-59 months within the KHDSS.

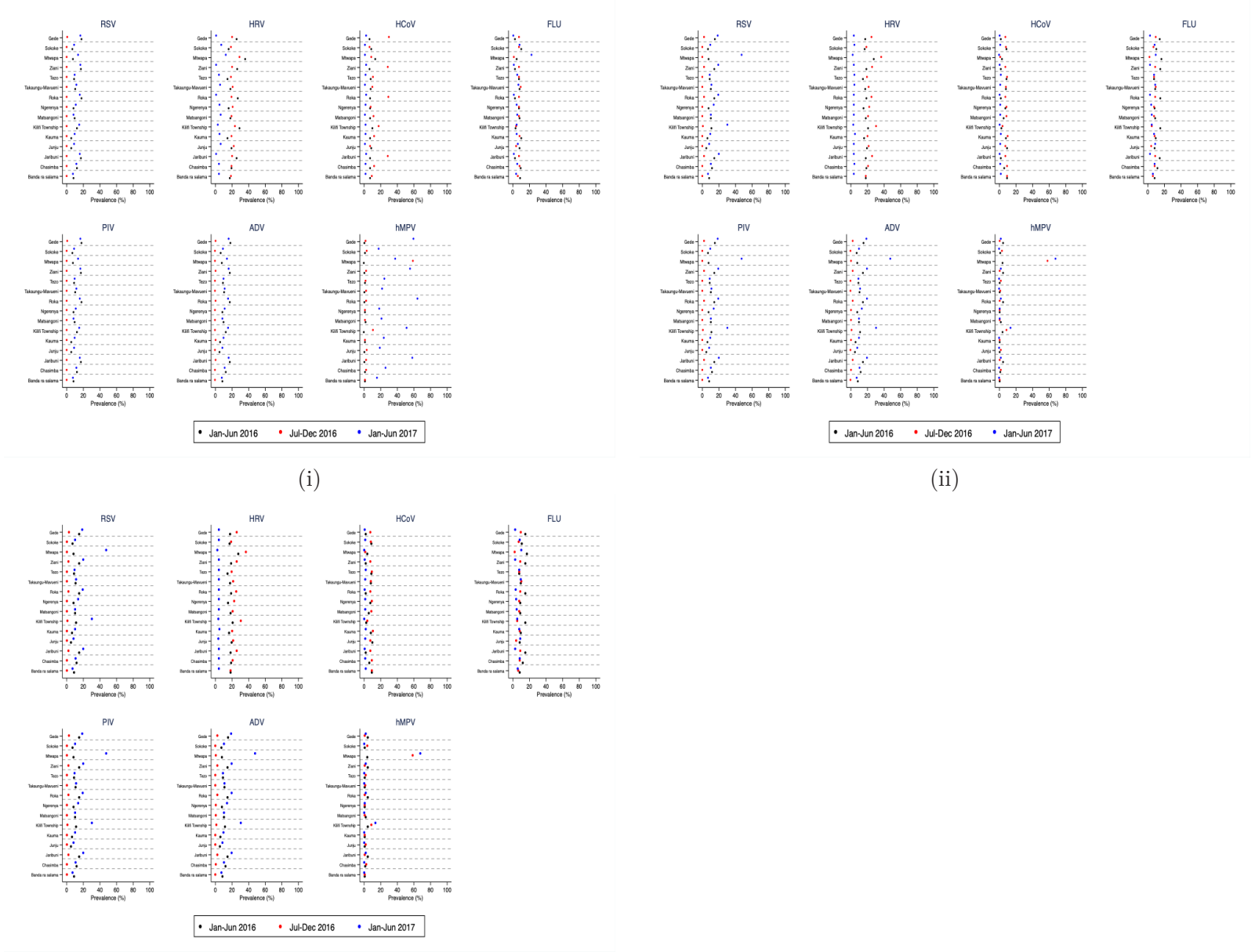
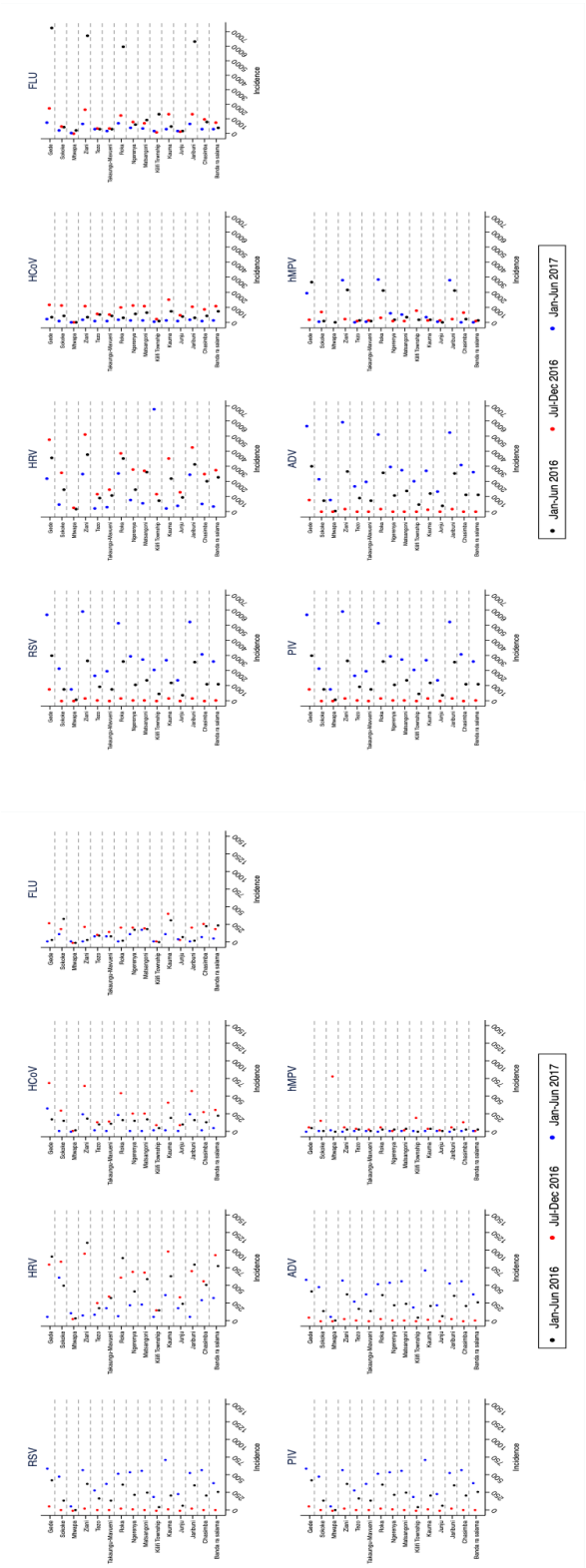


Figure 2.7a: Panel of estimated prevalence for all ages (i), under 5s (ii) and over 5s (iii).



(i) (ii) (iii)

Figure 2.7b: Panel of estimated incidence rates for all ages (i), under 5s (ii) and over 5s (iii).

2.1.3 Supplementary File 1

This file gives details of the statistical methods used for models fitted and selected to estimate the prevalence and incidence rates of viral ARI in Kilifi, Kenya.

Notations

For each of the 7 viral pathogens, let y_{it} denote the counts of positive cases in location i and month t assumed to follow a binomial specification; $y_{it} \sim Binomial(n_{it}, p_{it})$; $i = 1, ..., m$ locations and $t = 1, ..., T$ months, where p_{it} is the probability of a positive viral outcome in location i at time t.

The hierarchical space-time model structure for modeling p_{it} contained the following terms (1);

- Grand intercept α_0
- Spatial component accounting for existence of spatial autocorrelation across locations u_i
- Time component to account for fixed and random time effects by month B_t
- Space-time interactions term to account for residual spatial variation not accounted for by the main time and space effects v_{it}
- A set of auxilliary variables

An exchangeable structure on the intercepts of location, month and space-time interaction effects was assumed (2), hence defining the distribution on the hyperparameters e.g. for the space effect;

$$id \sim N(0, \frac{1}{\sigma^2_{u_i}}) \quad \text{where} \quad log\left(\frac{1}{\sigma^2_{u_i}}\right) \sim logGamma(1, 10^{-5}) \tag{1}$$

where id is an abitrary identifier for each location within the KHDSS

A linearity constraint was imposed on the temporally structured (γ_t) and unstructured (ζ_t) effects constituting the time component (B_t). The temporally structured effect was modeled dynamically using a random walk process order 1 (RW1);

$$\gamma_t|\gamma_{t-1} \sim Normal\left(\gamma_{t-1}, \sigma^2\right) \tag{2}$$

and ζ_t was specified using a Gaussian exchangeable prior; $\zeta_t \sim N\left(0, \frac{1}{\tau_\zeta}\right)$

To account for any spatial correlation, the Besag-York-Mollie (BYM) conditional autoregressive (CAR) process for the location random effects $(u_1, ..., u_m)^{'}$ (1,3) $u_i = \phi_i + \xi_i$, where

$$\phi_i|\phi_j; i \neq j, W, \tau^2 \sim N\left(\frac{\sum_j w_{ij}\phi_j}{\sum_j w_{ij}}, \frac{\tau^2}{\sum_j w_{ij}}\right) \tag{3}$$

with the hyperpriors specified as $\xi_i \sim N(0, \sigma^2)$ and $\tau^2 \sim logGamma(a, b)$. In addition, the proximity matrix W is specified with elements w_{ij} such that;

$$\mathbf{w}_{ij} = \begin{cases} 0 \text{ if } i = j \\ 1 \text{ if location } i \text{ neighbours } j \end{cases}$$

Models specification

Model 1: Poisson distribution

$$y_{it} \sim Poisson(\lambda_{it}) \;\; ; \;\; i = 1, ..., 9 \text{ and } t = 1, ..., T \tag{4}$$

λ_{it} is expressed as $\lambda_{it} = E_{it}\gamma_{it}$ where E_{it} and γ_{it} are the expected counts and probability parameters respectively. The model is specified as;

$$log\ (\lambda_{it}) \;\; = \;\; \alpha_0 + log\ (E_{it}) + x_{it}\beta + B_t + u_i + v_{it} \tag{5}$$

where β is a vector of coefficients, α_0, B_t, u_i and v_{it} are defined as specified above.

There were multiple instances of zero positive cases observed for a variable in a given location. This could be considered as presence of excess zeros hence the use of Poisson model would not suitable. Under such circumstances,the Zero-inflated Poisson (ZIP) model would preferred ; a mixture model of a proportion $1 - p$ of excess zeros, and a proportion p of nonexcess zero and nonzero counts is assumed.

Model 2: Zero-inflated Poisson distribution

Based on specifications of equation (4), $\lambda_{it} = I_{it} \times \mu_{it}$ hence the model is specified as;

$$log\ (\mu_{it}) \;\; = \;\; \alpha_0 + log\ (E_{it}) + x_{it}\beta + B_t + u_i + v_{it} \tag{6}$$

and $I_{it} \sim Bernoulli(p), \quad p \sim beta(1,1)$.

I is an indicator ($I = 0$ for excess zero counts and $I = 1$ for nonexcess zero or nonzero counts) and p is the probability of nonexcess zero counts (4).

Model 3: Binomial distribution

$$y_{it}|\tilde{p_{it}} \sim Binomial(n_{it}, p_{it}) \tag{7}$$

and the linking model given by;

$$logit(\tilde{p_{it}}) = \alpha_0 + x_{it}\beta + B_t + u_i + v_{it}$$

Model 4: Zero-inflated binomial (ZIB) distribution

In this distribution, it is assumed that

$$\mathbf{y}_{it} = \begin{cases} 0, & \text{with probability } \ p_{it} \\ \text{binomial } \ (n_{it}, \theta_{it}), & \text{with probaility} 1 - p_{it} \end{cases} \tag{8}$$

hence $y_{it} \sim ZIB(n_{it}, p_{it}, \theta_{it})$ and the parameters are modeled via a logit link and given by;

$$logit\ (\tilde{p_{it}}) = \alpha_0 + x_{it}\beta + B_t + u_i + v_{it}$$

Model 5: Logistic-Normal

This model was specified as;

$$y_{it} = log\left(\frac{\hat{p_i t}}{1 - \hat{p_i t}}\right) \tag{9}$$

Then by the delta method, the asymptotic distribution of y_{it} is:

$$y_{it}|p_{it} \sim N\left(log\left(\frac{\hat{p_i t}}{1 - \hat{p_i t}}\right), \frac{\hat{Var}(\hat{p_{it}})}{\hat{p_{it}}^2 (1 - p_{it})^2}\right)$$

References

1. Lawson AB. Bayesian Disease Mapping: Hierarchical modeling in Spatial Epidemiology. Chapman & Hall / CRC; 2009.

2. Marta Blangiardo, Michela Cameletti. Spatial and Spatio-temporal Bayesian Models with R-INLA. 2015.

3. Sudipto Banerjee, Bradley P. Carlin, Alan E. Gelfand. Hierarchical Modeling and Analysis for Spatial Data. 2nd ed. Chapman; Hall/CRC; 2014. (Chapman & Hall/CRC Monographs on Statistics and Applied Probability).

4. Ghosh SK, Mukhopadhyay P, Lu J-C. Bayesian analysis of zero-inflated regression models. Journal of Statistical Planning and Inference [Internet]. 2006 Apr [cited 2020 Apr 23];136(4):1360–75. Available from: http://www.sciencedirect.com/science/article/pii/S0378375804004008

CHAPTER 3: EXTENDED METHODS

This chapter provides additional information on the methods used in this work that was not reported in the draft manuscript presented in the previous chapter. Specifically, the aspects of study design, sampling weights and model selection are highlighted and discussed below.

2.1 Study Design

In this study, the same sentinel surveillance study design (1) for entire duration covered by the original study (SPReD KHDSS) was adopted.

2.2 Study site

This study was conducted within the KHDSS, along the coastal region of Kenya (2,3). The KHDSS, nested within KWTRP, was established in 2000 and maintains records of pregnancies, births, child vaccination, deaths and migration (2). The study utilised data from 9 of the 21 government owned outpatient facilities namely; Matsangoni, Ngerenya, Mtondia, Sokoke, Mavueni, Jaribuni, Chasimba, Pingilikani and Junju which were selected to represent the geography of KHDSS across 15 admistrative locations (3).

2.3 Study Population

The study population consisted of all people living within the KHDSS and accessed any of the 9 health facilities for treatment of ARI.

2.4 Sampling

No sampling was done in this study.

2.5 Data collection

Aggregate data from the investigator(s) of the original study at KEMRI-Wellcome Trust Research Programme (KWTRP) in Kilifi, Kenya was used. The obtained data contained weekly counts of ARI episodes, stratified by location, age categories and virus pathogen. In addition, auxilliary data on the geographical locations within KHDSS such as mid-year population estimates per location and weather patterns for the period covered in the study were obtained.

2.6 Data Management and Analysis

In this study, aggregate data was used therefore minimal data cleaning processes were required. Data analysis was conducted in R (version 3.6.1) with the support of various packages listed in appendix A9. R codes of models fitted in this analysis are available in appendix A10.

2.7 Sampling weights

Estimates of viral ARI prevalence and incidence rates presented in the draft manuscript above were obtained from SAE models without considering sampling weights. These weights could be estimated by:

$$wt_{ikjt} = \frac{1}{\text{Probability of selecting an individual i at facility j within location i at time t}} \tag{1}$$

where the probability in the denominator, while assuming independence, can be expressed as the product of;

1. Probability of selecting a location k (9 out of 15 locations participated)

$$= \frac{9}{15}$$

2. Probability of selecting a health facility j in location k (1 facility per location)

$$= \frac{1}{\text{Number of health facilities in location } \ k}$$

3. Probability of selecting an individual i at facility j in location k at time t

$$= \frac{\text{Number of study recruited ARI cases at facility j in location k at time t}}{\text{Number of ARI cases that were served at facility j in location k at time t}}$$

However, the number of ARI visits that were served in a health facility (denominator of part 3) was not readily available through out the study period hence sampling weights were not calculated.

2.8 Model Fitting

Additional information providing a general outline of the hierarchical Bayesian specification of the SAE models used in this study is presented in the appendix A5. In the methods section of the manuscript, it is indicated model selection was based on low values of DIC and effective number of parameters. The DIC, defined as $DIC = \bar{D(\theta)} + 2pD$, is the sum of the posterior mean of deviance $D(\theta)$ given by $\bar{D(\theta)} = E[D(\theta)|y]$ and twice the difference between posterior mean deviance and the deviance evaluated at the posterior mean of the parameters which is given by $pD = \bar{D(\theta)} + D(E[\theta|y])$. A DIC difference of 5-10 units between competing models was considered good (4,5).

2.9 Ethical Approval

Only count and weather data was obtained from KWTRP hence no personally identifiable information availed; there was no potential risk of loss of confidentiality among study participants. Data from the primary study is hosted in an online data repository (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZX7NS4) and was made available upon request from KWTRP. The primary study was approved by the KEMRI-Scientific and Ethical Review Unit (SERU# 3103) and the University of Warwick Biomedical and Scientific Research Ethics Committee (BSREC# REGO-2015-6102) (3) while this study (protocol number W-CBP-200317) was granted ethical waiver by the Human Research Ethics Committee of the University of Witwatersrand (Appendix A3)

References

1. Miquel Porta. A Dictionary of Epidemiology. 5th ed. Oxford University Press,Inc; 2008.

2. Scott JAG, Bauni E, Moisi JC, Ojal J, Gatakaa H, Nyundo C, et al. Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). International Journal of Epidemiology. 2012 Jun;41(3):650–7.

3. Nyiro JU, Munywoki P, Kamau E, Agoti C, Gichuki A, Etyang T, et al. Surveillance of respiratory viruses in the outpatient setting in rural coastal Kenya: Baseline epidemiological observations. Wellcome Open Research. 2018 Jul;3:89.

4. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002 Oct;64(4):583–639.

5. Spiegelhalter DJ, Best NG, Carlin BP, Linde A van der. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2014;76(3):485–93.

CHAPTER 4: EXTENDED RESULTS

In this chapter, additional results not reported in the manuscript above are presented. These results were obtained from the selected models by DIC values and discussed in the submissible manuscript (chapter 2) for each virus pathogen. In addition, some model comparisson outputs which helped inform model selections are also presented in this chapter. Spatio-temporal models were fitted, however in this section the two components of space and time are discussed separately

4.1 Temporal Component

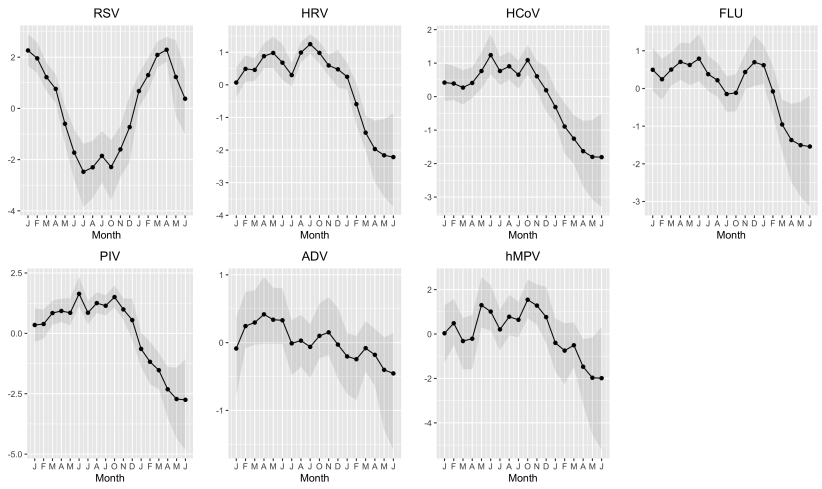
4.1.1 Exploratory Data Analysis

Temporal analysis was conducted using monthly counts data of each viral respiratory pathogen. The auto correlation plots, for the period January 2016 to June 2017, showed that each virus had peaks at different periods (Appendix A6). Using the portmanteau test for white noise, it was established that temporal pattern was unique for RSV (p-value=0.0886), HCoVs (p-value=0.0043), PIV (p-value=0.0028) and HRV (p-value=0.058). However, the same evidence was not observed for ADV (p-value=0.6465), FLU (p-value=0.7139) and hMPV (p-value=0.1133) which had low positive numbers.

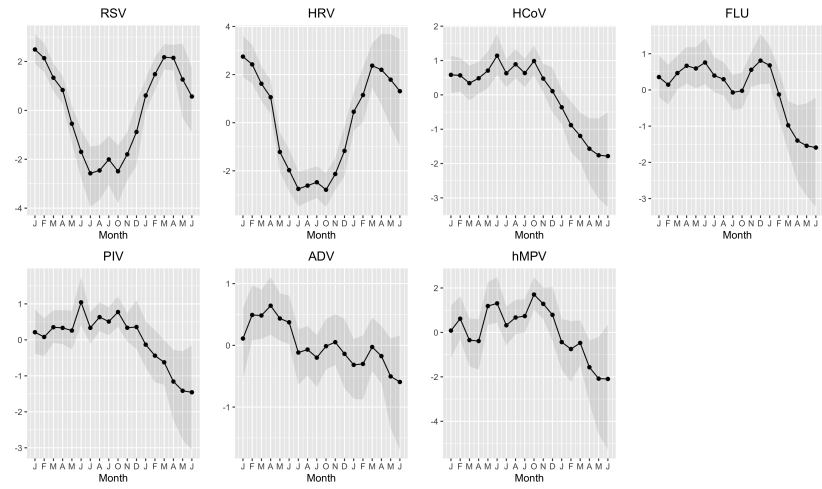
4.1.2 Non-linear effects of time

This section gives results on the non-linear effects of time in months on the prevalence and incidence rates of viral ARI in kilifi based on the model of best fit selected for each virus pathogen.

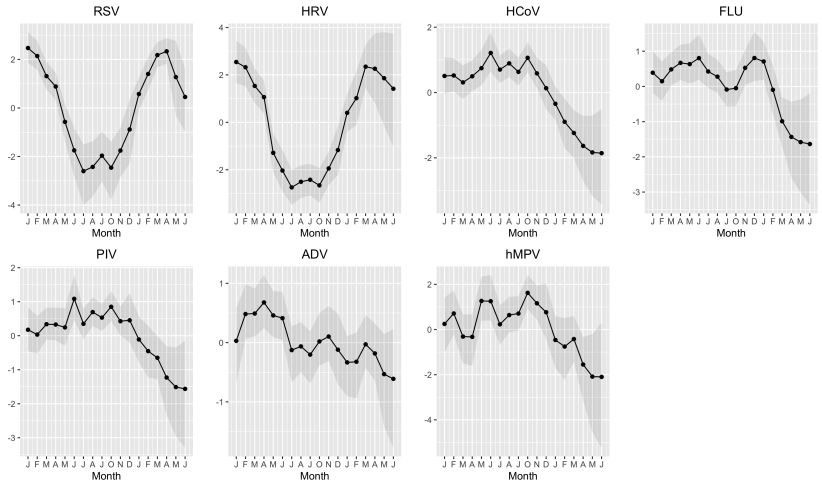
The overall non linear effects of time in months on ARI prevalence are presented in figure 4.1. A non linear influence of time in months on the prevalence of viral ARI was observed across all ages collectively and when stratified by either under or over 5 year olds. When the data was stratified by age, wide confidence intervals were observed for ADV. Similarly, a non -linear effect of time in months was observed when estimating the incidence rates per 100000 population (Figure 4.2). In adition, wide confidence intervals were observed across all ages (FLU, ADV and hMPV), under 5s (FLU,PIV and ADV) and over 5s (FLU,ADV and hMPV). The wide CIs are attributed to low numbers of positive cases for the respective viral pathogens.



(a)



(b)



(c)

Figure 4.1: Panel of graphs showing non-linear effects of time in months on the prevalence of viral ARI for all ages (a), under 5 year olds (b) and over 5 year olds (c). Shown are the posterior means together with 95intervals.

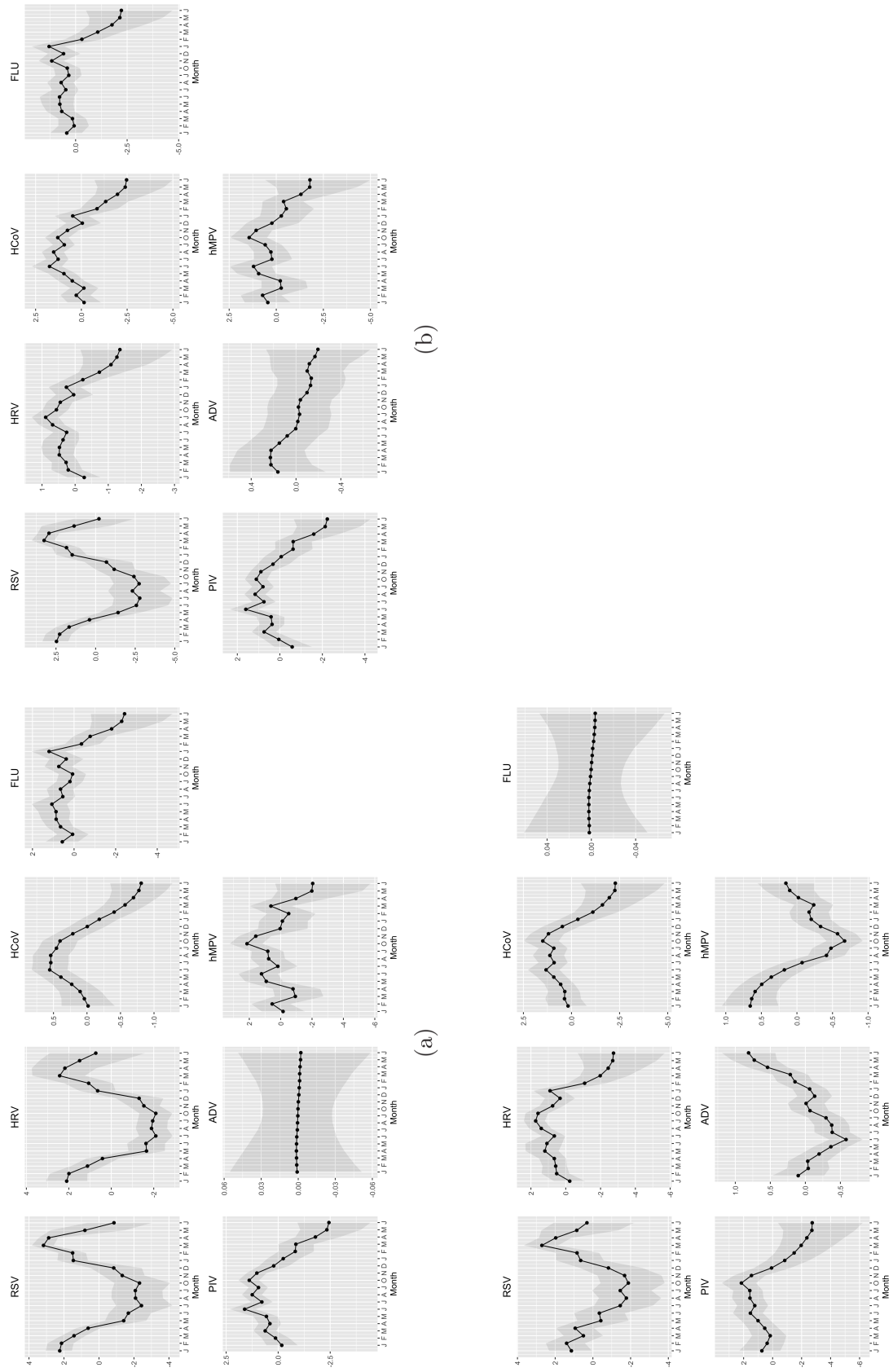


Figure 4.2: Panel of graphs showing non-linear effects of time in months on the incidence of viral ARI for all ages (a), under 5 year olds (b) and over 5 year olds (c). Shown are the posterior means together with 95intervals.

4.2 Spatial Component

4.2.1 Spatial Autocorrelation

The results indicated that there was no statisticial evidence of a global spatial autocorrelation on viral ARI cases in Kilifi (Global Moran's I index=0.06, p-value=0.2752). This could be attributed to the possible exisitence of significant spatial autocorrelation in a small section of the KHDSS locations masked in the context of the entire region and the fact that the study was conducted in selected locations. The local Moran's I index showed marginal significance of spatial autocorrelation ($I_i = 1.22$, p-value=0.082) indicating potential presence of spatial correlations in viral ARI at local levels.

4.2.2 Spatial random effects

A map of spatial random effects was generated for each virus pathogen using the best fit model selected in order to identify any spatial trends. A visual inspection of the maps did not show any strong spatial signals for both prevalence (Figures 4.3 a and b) and incidence rates (Figures 4.4 a and b) hence no trend was observed.

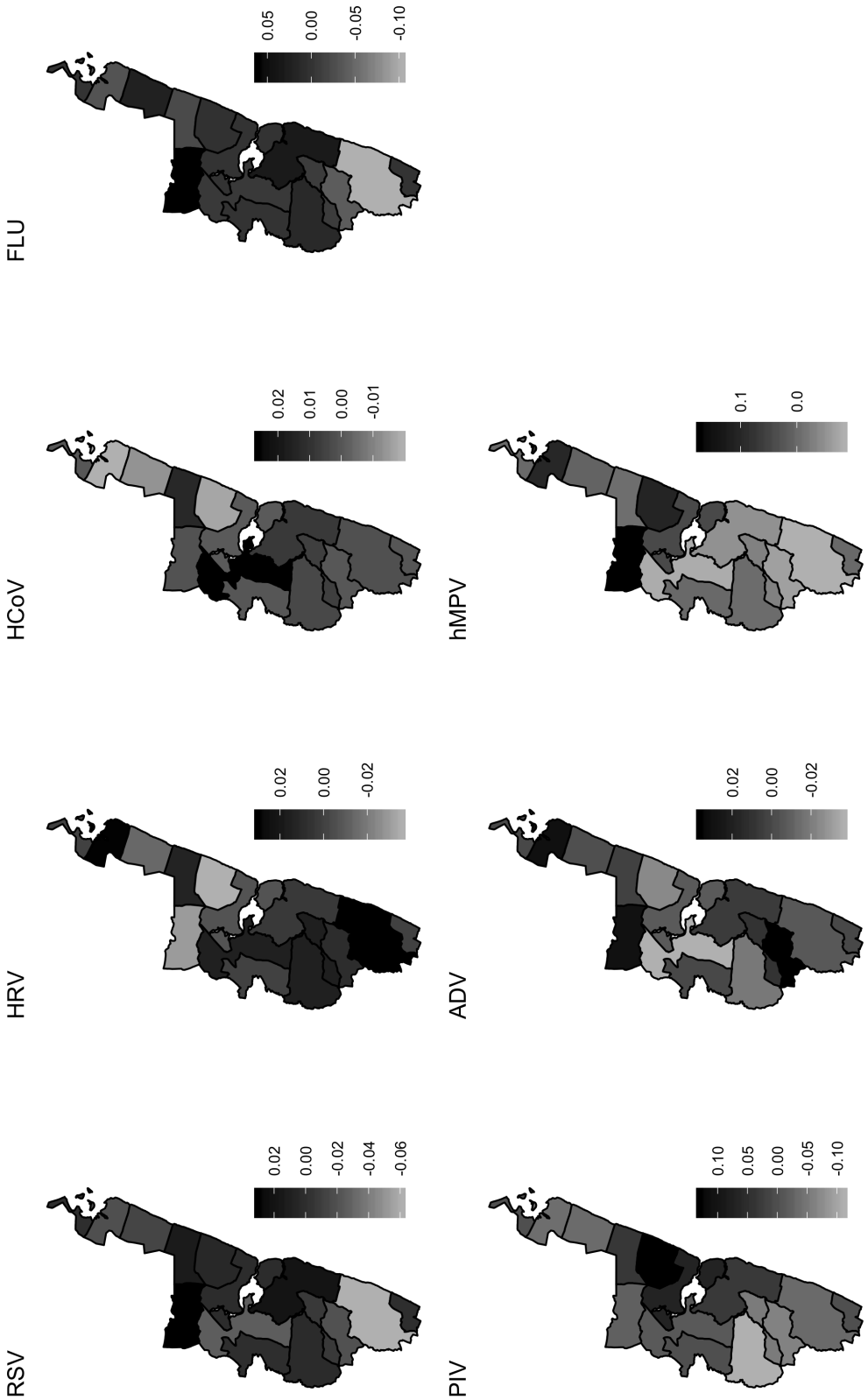
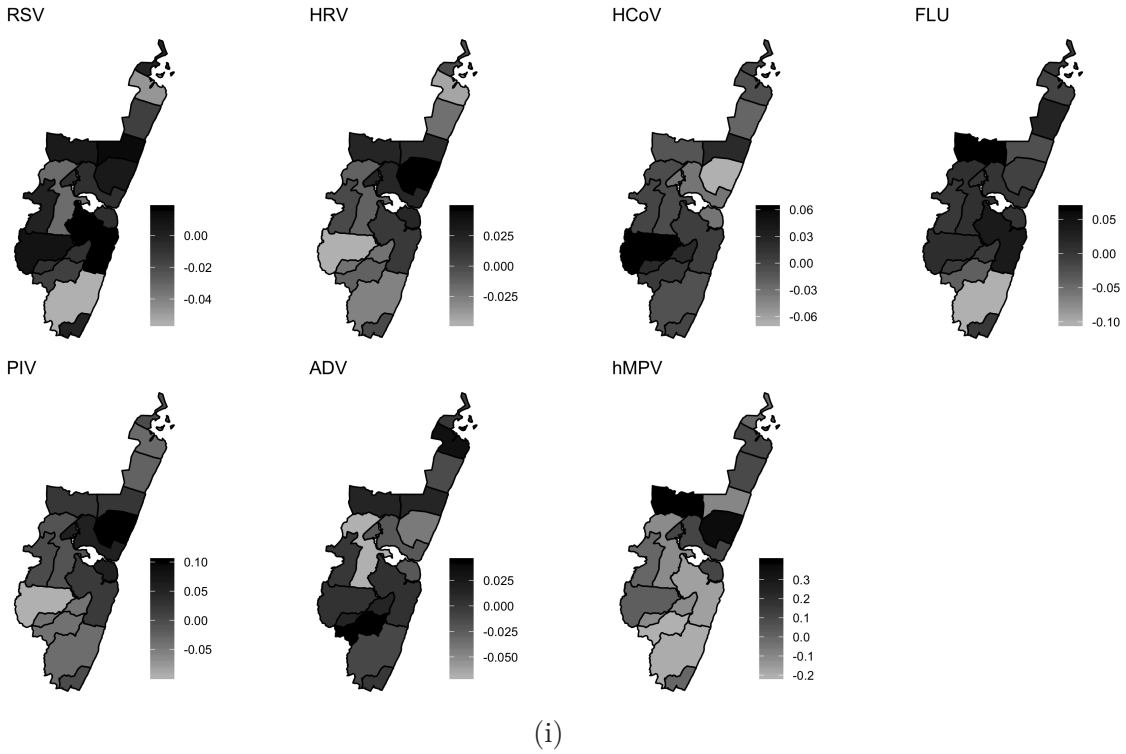
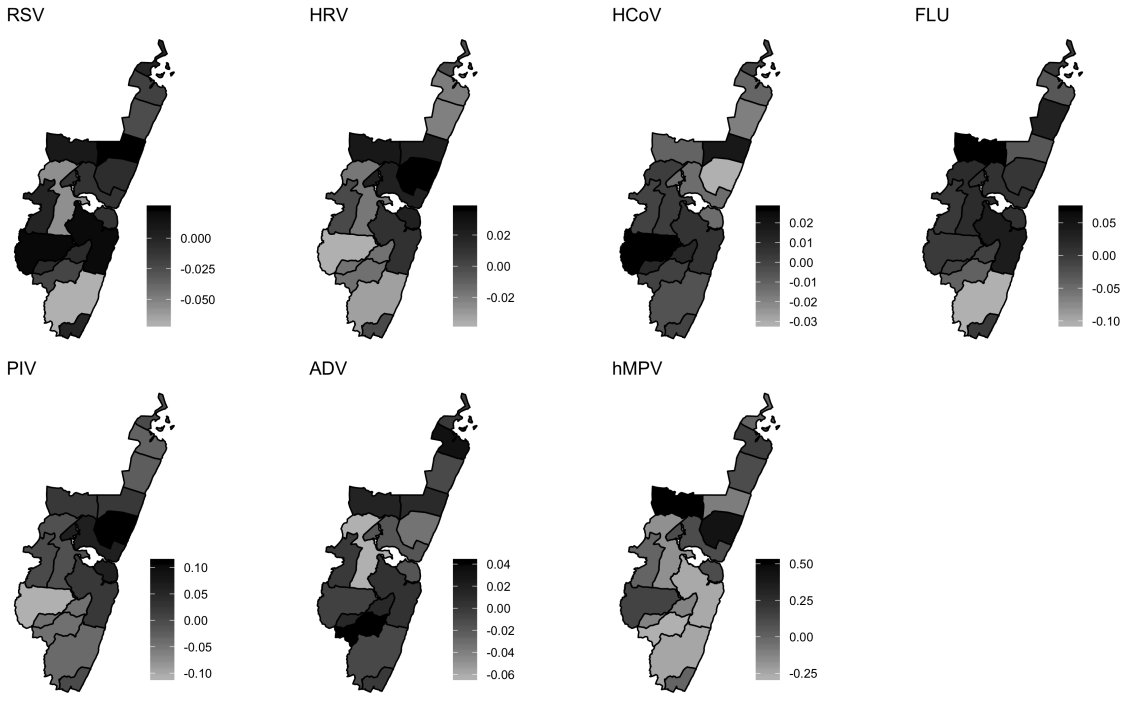


Figure 4.3a: Maps showing the residual spatial locations effects on the prevalence of viral ARI in Kilifi for all ages



(i)



(ii)

Figure 4.3b: Panel of maps showing the residual spatial locations effects on the prevalence of viral ARI in Kilifi for under 5 year olds (i) and over 5 year olds (ii).

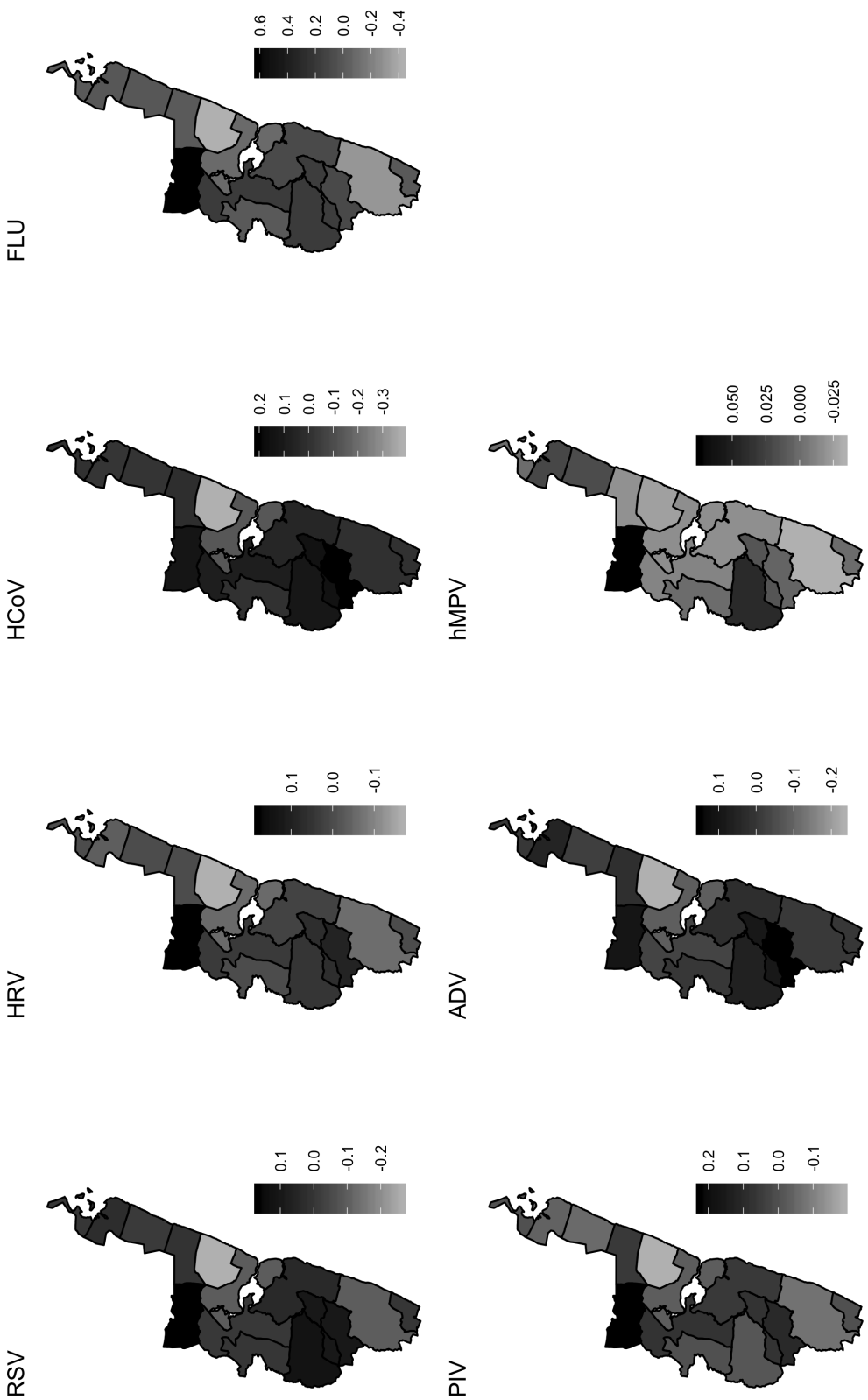
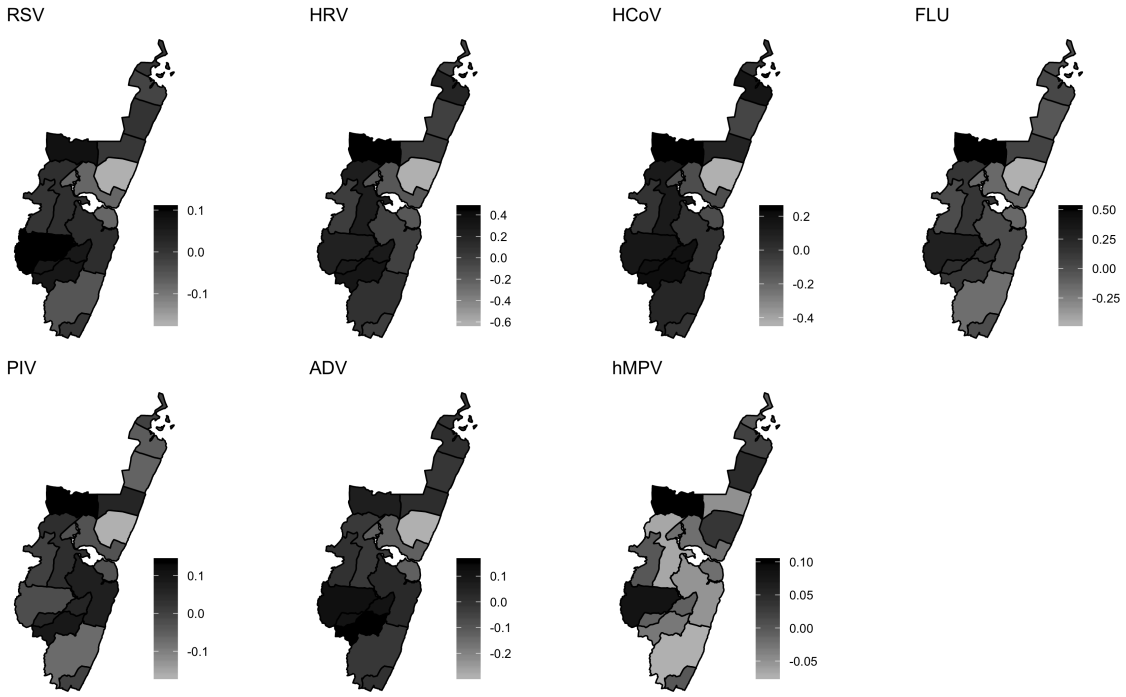
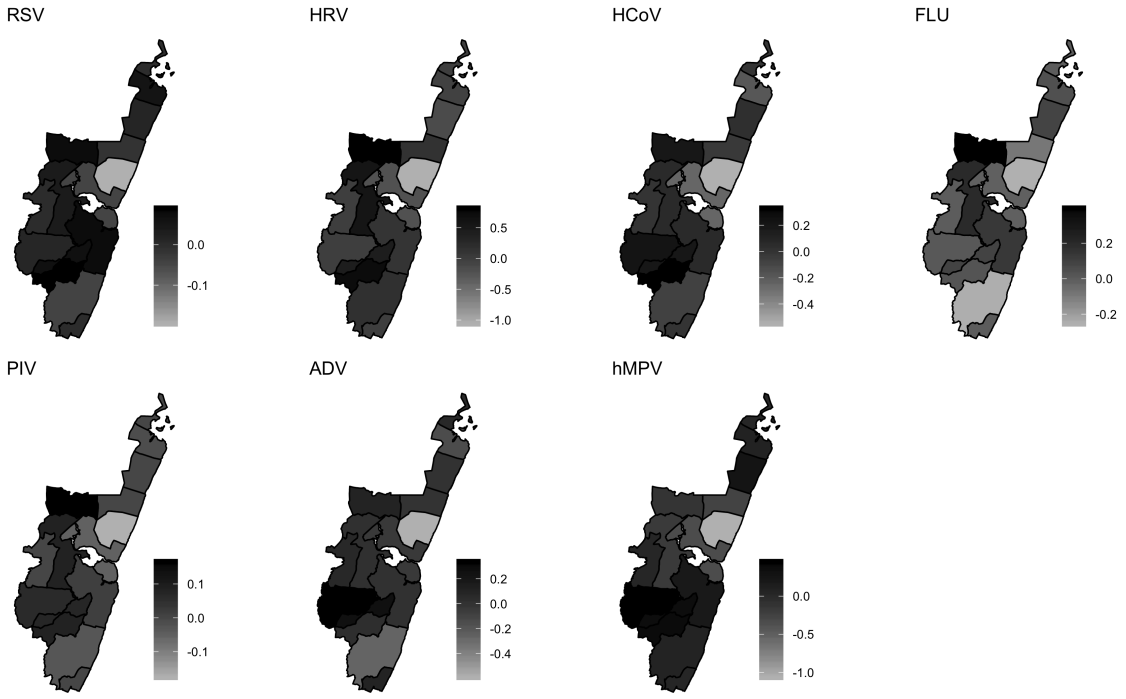


Figure 4.4a: Maps showing the residual spatial locations effects on the incidence rates of viral ARI in Kilifi for all ages



(i)



(ii)

Figure 4.4b: Panel of maps showing the residual spatial locations effects on the incidence rates of viral ARI in Kilifi for under 5 year olds (i) and over 5 year olds (ii).

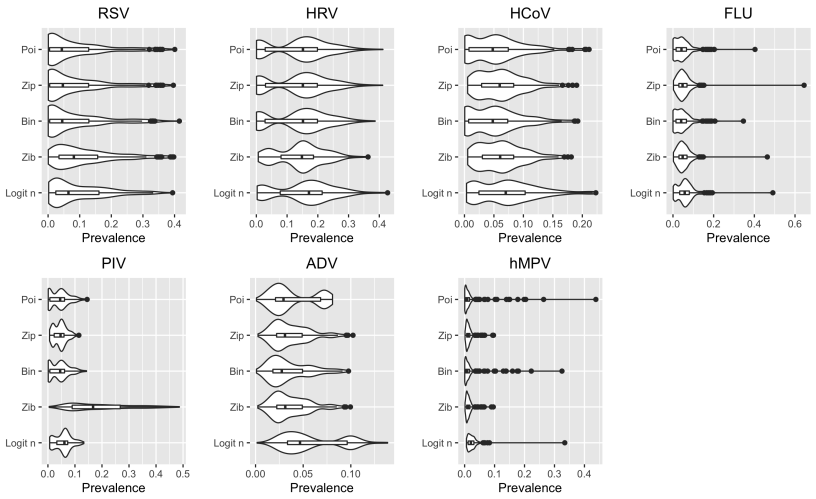
4.3 Models comparisson

4.3.1 Model fit

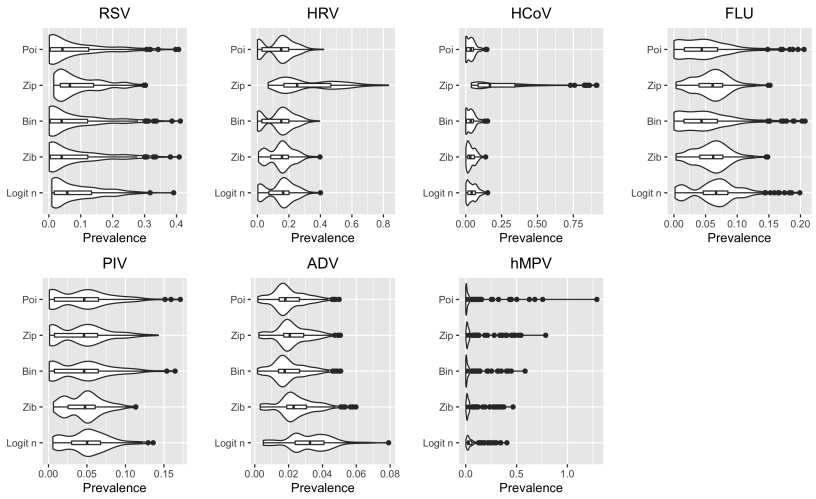
In the draft manuscript in chapter 2, it is indicated that model selection was based on low DIC. Due to the heavy criticism of the DIC as a measure of model fit (see chapter 5) especially under complex models, Watanabe–Akaike information criterion were also checked. Use of WAIC for model comparisson led to selection of the same models as selected using the DIC (Appendices A6 and A7).

4.3.2 Model estimates

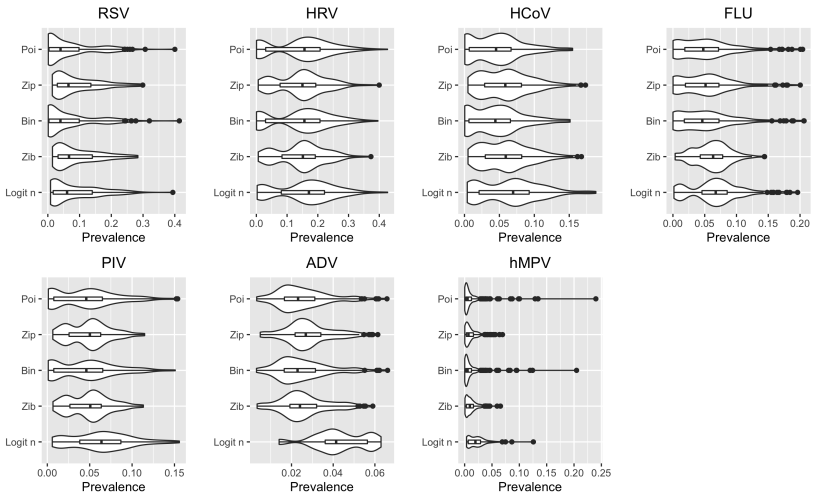
From each model fitted, small area predicted of prevalence and incidence rates were obtained. Figures 4.5 and 4.6 present violin plots of predicted prevalence and incidence rates respectively. Generally, it was observed that predicted estimates from the best perfoming model, by DIC values, for each virus pathogen were less heterogenious compared to models with higher DIC values. The shapes of the violin plots across different models for each pathogen were generally similar. By comparison, competing models with small DIC differences (mostly between 5-10) had very close resemblence of violin plots indicating similarity of predicted values.



(a)

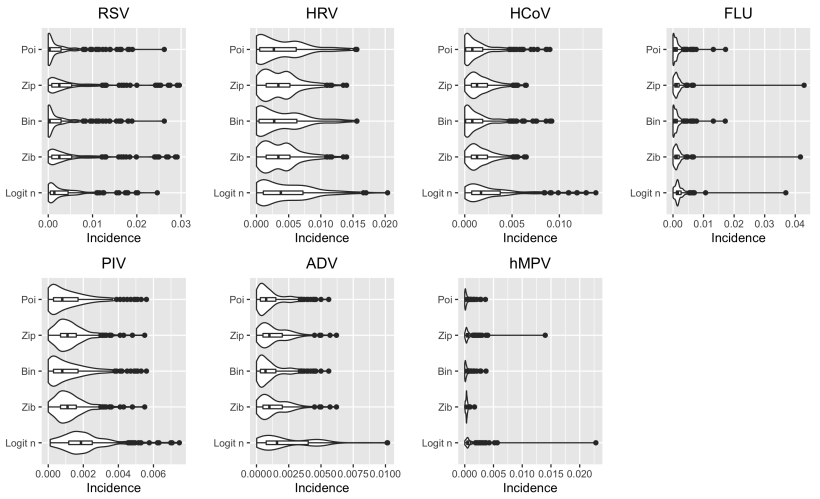


(b)

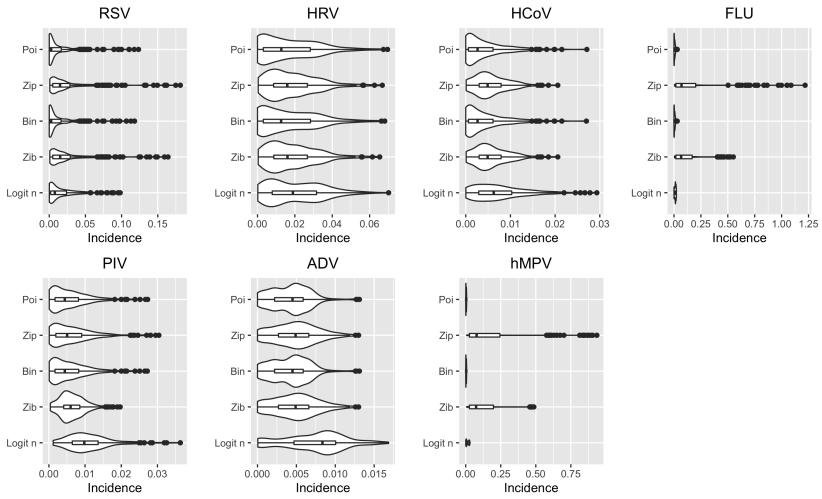


(c)

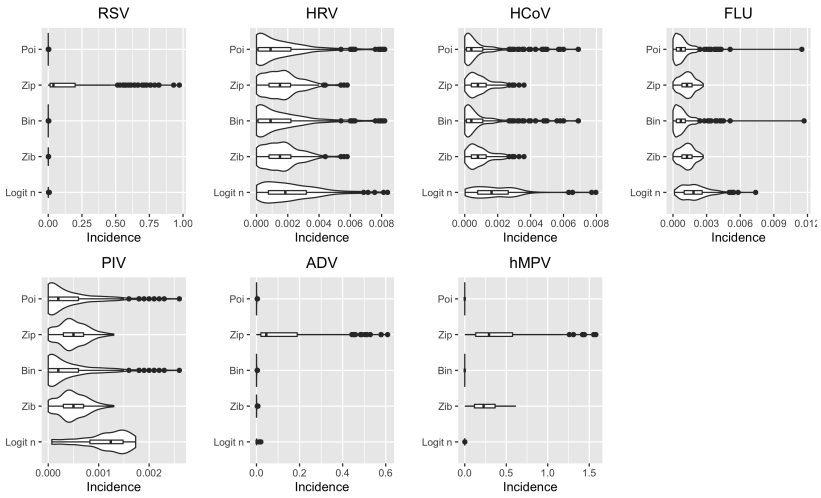
Figure 4.5: Panel of violin plots showing the estimated prevalence by virus pathogen tested from each model fitted on the data. The panel plots are presented for all ages (a), under 5 year olds (b) and over 5 year olds (c).



(a)



(b)

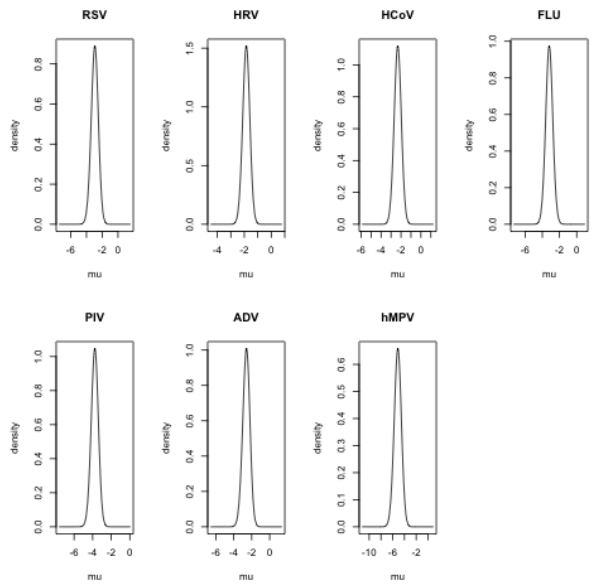


(c)

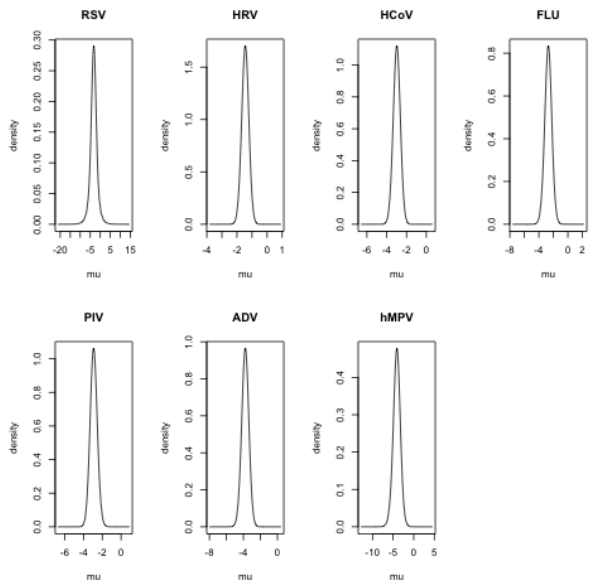
Figure 4.6: Panel of violin plots showing the estimated incidence rates by virus pathogen tested from each model fitted on the data. The panel plots are presented for all ages (a), under 5 year olds (b) and over 5 year olds (c).

4.3.2 Model fit diagnostics

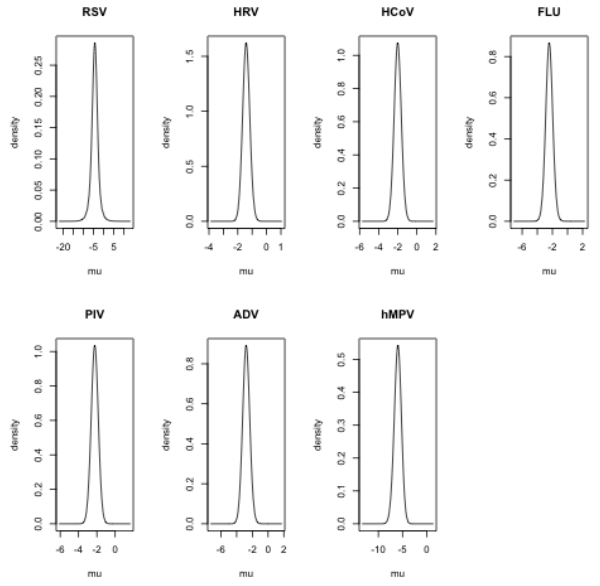
The final models selected were investigated for convergence using density plot for the posterior means. Symmetry was observed on the convergence plots for the prevalence and incidence rates estimates indicating that the models captured the data well. This was evident in models used to estimate the prevalence (Figure 4.7) and incidence rates (Figure 4.8).



(a)

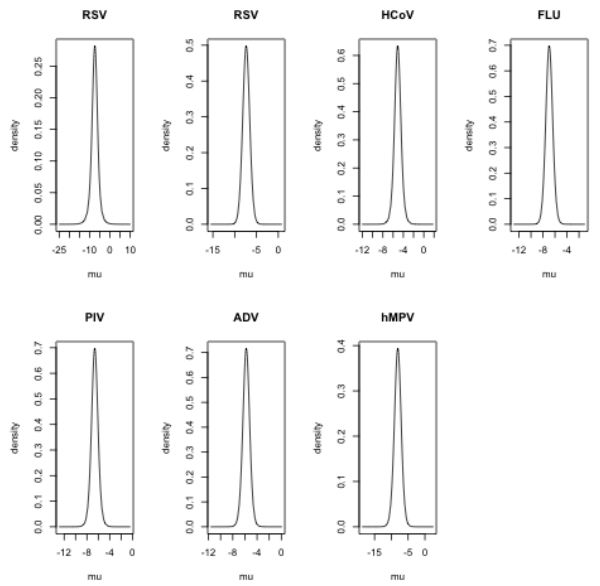


(b)

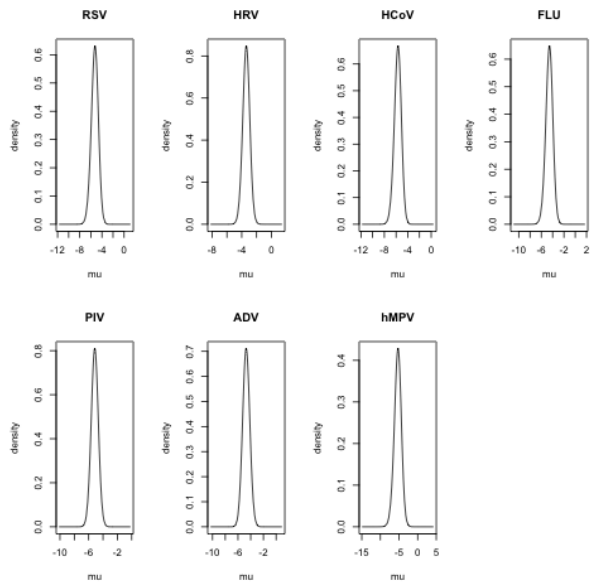


(c)

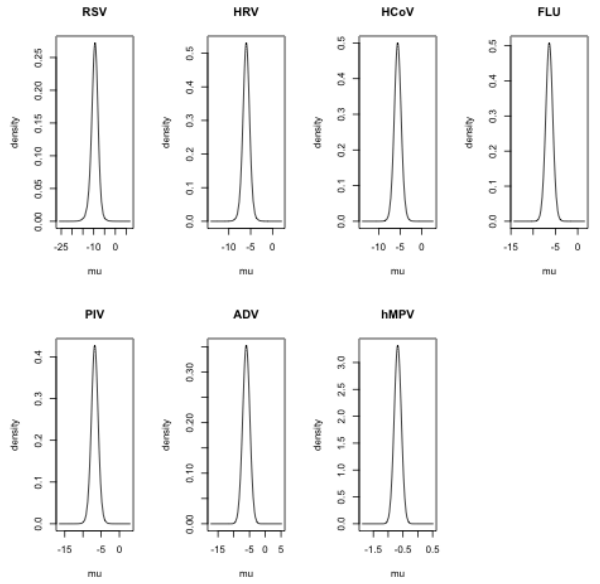
Figure 4.7: Panel of posterior density plots from models used to estimate ARI prevalence by virus pathogen tested. The panel plots are presented for all ages (a), under 5 year olds (b) and over 5 year olds (c).



(a)



(b)



(c)

Figure 4.8: Panel of posterior density plots from models used to estimate incidence rates by virus pathogen tested. The panel plots are presented for all ages (a), under 5 year olds (b) and over 5 year olds (c).

CHAPTER 5: OVERALL DICUSSIONS

This chapter gives a brief discussion of the study findings, the overall conclusions and recommendations based on the key findings of the study. In addition, the strengths and limitations of this study are presented.

5.1 Discussions

This study, conducted in a rural area along the northern coastal region of Kenya, estimated the prevalence and incidence of viral ARI using data from an outpatient surveillance at nine government owned health facilities. This was achieved by fitting hierarchical Bayesian spatio-temporal models, which assumed different probability distributions, on each of the 7 viral respiratory pathogens tested in Kilifi. Comparissons were made on two age bands (children aged 0-59 months (under 5s) and those aged above 5 years) and overall (all ages).

There was a fluctuating pattern of the number of cases that tested positive for each virus pathogen. Among the 7 virus pathogens tested, RSV had peaks in the first quarter of the year coinciding with RSV epidemics which have been reported to begin in last quater of a year through to the first quarter of the subsequent year (1). Model estimates showed that the burden of ARI was quite substantial among children aged under 5 years. Estimates of disease prevalence were shown to vary across age and viral pathogens.

In addition, there were spatio-temporal variations of both prevalence and incidence rates estimates for all age bands. Spatial heterogeneity of ARI reported in this study, is consistent with results from larger studies conducted in Ethiopia and Nigeria (2,3). Compared to other outpatient studies conducted at KCH on under 5s, we observe lower prevalence of RSV (4% versus 7%) and HRV (12% versus 24%) (4,5). This difference could be as a result of these studies happening at a time prior to recent developments of health care systems in Kenya. These development initiatives, especially in rural areas, have resulted in improved services at local public health facilities (6) hence possibly reduce the need to visit KCH for care.

There is limited data on the epidemiology of ARI in Africa (7). Over a five year period (2012-2016), the reported prevalence of hMPV (3.7%) in a smaller age band (1 month-46 years) in Senegal (8) was considerably larger than estimates (less than 1%) for all ages reported in this study. A surveillance study at 5 health facilities in Uganda, using influenza like illness case definition, estimated a higher prevalence of influenza A or B (10.4%) for all ages compared to estimates from this study which were nearly half their estimate (9). In Nigeria, higher prevalence for FLU (A and B only) were reported though with a smaller sample size a different case definition compared this study (10).

Hierarchical bayesian SAE models were fitted on each virus pathogen assuming different distributions and the model with the lowest DIC considered of best fit. Although the DIC is widely used to assess model fit and complexity, it has been criticised to have several limitations such as invariance of pD to parametisaton and inconsistency of the DIC (11). In light of these limitations, it was investigated whether using a different criteria would lead to different decision(s) on model selection. In this study, model comparison using WAIC led to selection of the same models as using the DIC (Appendices A7 and A8).

5.2 Strengths and Limitations

The available data on measures of ARI's disease burden within developing economies, such as Kenya, is limited (7) and have mostly been reported at national or higher sub-national levels (12,13). This study's main strength is the ability to provide estimates of ARIs disease burden (prevalence and incidence) at low sub-national small areas (locations) within a rural setting along northern coastal region of Kenya. Spatio-temporal hierarchical Bayesian models were used to estimate the prevalence and incidence of medically attended ARI, from outpatient health facilities within the KHDSS, attibuted to seven respiratory viral pathogens. This estimation process was anchored on the ability of SAE models to borrow strength through space and time (14) hence predictions were made for both sampled and unsampled locations. In addition, application of SAE methods in the epidemiology of ARI has not been widely

explored in SSA and this study provides additional evidence to the applicability of SAE in epidemiology. The hierarchical bayesian approach adopted provides improvements to model estimates by specifying prior distributions in the models while accounting for potential variations in space and time. However, there exists a possibility that model estimates reported in this study could change if sampling weights were included in the models (15). It was not possible to include sampling weights due to limited data on the study design. As a result of the complex structure of our data, the models were fitted in INLA alone hence MCMC was not utilised due to it's high computational requirements. This study was ecological in nature hence our ARI disease burden estimates bear the potential limitations of ecological fallacy attributable to such studies (16). This limitation can be overcomed by conducting individual level analysis which was however not possible in this study due to some data limitations. Therefore, it is recommended that future SAE modeling studies be conducted at individual level to further strengthen the evidence presented in this study.

5.3 Conclusion

To understand the epidemiology of ARI, especially in Africa, and implement effective control measures, it is essential to capture the burden of ARI which is currently sparsely reported. Using outpatient surveillance data, evidence on the considerable burden of ARI in a tropical setting within SSA has been provided. A substantial part of this burden is borne by children aged less than 5 years. This has implications on future interventons such as vaccination, localised therapies, surveillance and planning. Additional data and further research can be conducted to estimate the burden of ARI in Kenya and other SSA countries, preferrably using individual level SAE models, thereby help support localised healthcare policy. SAE models have been used to estimate disease burden in both sampled and unsampled regions hence giving a clearer understanding of the geographical spread of ARI within the KHDSS.

References

1. Nokes DJ, Ngama M, Bett A, Abwao J, Munywoki P, English M, et al. Incidence and Severity of Respiratory Syncytial Virus Pneumonia in Rural Kenyan Children Identified through Hospital Surveillance. Clinical Infectious Diseases. 2009 Nov;49(9):1341–9.

2. Amsalu ET, Akalu TY, Gelaye KA. Spatial distribution and determinants of acute respiratory infection among under-five children in Ethiopia: Ethiopian Demographic Health Survey 2016. PLOS ONE. 2019 Apr;14(4):e0215572.

3. Adesanya OA, Darboe A, Mendez Rojas B, Abiodun DE, Beogo I. Factors contributing to regional inequalities in acute respiratory infections symptoms among under-five children in Nigeria: A decomposition analysis. International Journal for Equity in Health. 2017 Aug;16(1):140.

4. Okiro EA, Ngama M, Bett A, Nokes DJ. The Incidence and Clinical Burden of Respiratory Syncytial Virus Disease Identified through Hospital Outpatient Presentations in Kenyan Children. PLOS ONE. 2012 Dec;7(12):e52520.

5. Onyango CO, Welch SR, Munywoki PK, Agoti CN, Bett A, Ngama M, et al. Molecular Epidemiology of Human Rhinovirus Infections in Kilifi, Coastal Kenya. Journal of Medical Virology. 2012 May;84:823–31.

6. Marx M, Nitschke C, Nafula M, Nangami M, Brodowski M, Marx I, et al. If you can't measure it- you can't change it – a longitudinal study on improving quality of care in hospitals and health centers in rural Kenya. BMC Health Services Research. 2018 Apr;18(1):246.

7. Madhi SA, Klugman KP. Acute Respiratory Infections. In: Jamison DT, Feachem RG, Makgoba MW, Bos ER, Baingana FK, Hofman KJ, et al., editors. Disease and Mortality in Sub-Saharan Africa [Internet]. 2nd ed. Washington (DC): World Bank; 2006 [cited 2019 Oct 21]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK2283/

8. Jallow MM, Fall A, Kiori D, Sy S, Goudiaby D, Barry MA, et al. Epidemiological, clinical and genotypic features of human Metapneumovirus in patients with influenza-like illness in Senegal, 2012 to 2016. BMC Infectious Diseases. 2019 May;19(1):457.

9. Wabwire-Mangen F, Mimbe DE, Erima B, Mworozi EA, Millard M, Kibuuka H, et al. Epidemiology and Surveillance of Influenza Viruses in Uganda between 2008 and 2014. PLOS ONE. 2016 Oct;11(10):e0164861.

10. Odun-Ayo F, Odaibo G, Olaleye D. Influenza virus A (H1 and H3) and B co-circulation among patient presenting with acute respiratory tract infection in Ibadan, Nigeria. African Health Sciences. 2018 Dec;18(4):1134–43.

11. Spiegelhalter DJ, Best NG, Carlin BP, Linde A van der. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2014;76(3):485–93.

12. Troeger C, Blacker B, Khalil IA, Rao PC, Cao J, Zimsen SRM, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory

infections in 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. The Lancet Infectious Diseases. 2018 Nov;18(11):1191–210.

13. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Swartz S, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: A systematic analysis for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases. 2017 Nov;17(11):1133–61.

14. J. N. K. Rao, Isabel Molina. Small Area Estimation. Second. John Wiley & Sons, Inc; 2015.

15. Pfeffermann D. The Role of Sampling Weights When Modeling Survey Data. International Statistical Review / Revue Internationale de Statistique. 1993;61(2):317–37.

16. Cohen HW. Limitations of an Ecological Study: A Review. American Journal of Hypertension. 2005 Jun;18(6):750–0.

APPENDIX

A1. Plagiarism Statement

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG | FACULTY OF HEALTH SCIENCES

Plagiarism declaration for written work

I, **Grieven Paul Otieno**, Student number:**1846979** is a postgraduate student registered for the degree/programme of Master of Epidemiology (Biostatistics) at the Wits School of public Health. I am submitting my written work for the assessment for the module COMH7175A: Research Report. I hereby declare the following:

- I am aware that plagiarism is the use if someone else's work without their permission and/or without acknowledging the original source.
- I am aware that plagiarism is wrong
- I confirm that the work submitted for the stated course is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required convention in referencing the thoughts and ideas of others.
- I understand that the university of the Witwatersrand may take discplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words of others in my writing.

Signature:  Date: April 30, 2021

A2. Turnitin Report

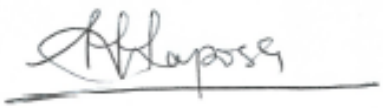
1846979:Msc_Biostats_report_GrievenOtieno.pdf

ORIGINALITY REPORT

13%	10%	10%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	wellcomeopenresearch.org Internet Source	2%
2	hdl.handle.net Internet Source	1%
3	Hae-Ryoung Song, Andrew Lawson, Ralph B. D'Agostino, Angela D. Liese. "Modeling type 1 and type 2 diabetes mellitus incidence in youth: An application of Bayesian hierarchical regression for sparse small area data", Spatial and Spatio-temporal Epidemiology, 2011 Publication	<1%
4	stacks.cdc.gov Internet Source	<1%
5	Tivey, Maurice A., H. Paul Johnson, Marie S. Salmi, and Michael Hutnak. "High-resolution near-bottom vector magnetic anomalies over Raven Hydrothermal Field, Endeavour Segment, Juan de Fuca Ridge. : Raven field vector magnetic anomalies", Journal of Geophysical Research Solid Earth, 2014.	<1%



A3. Ethics clearence

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

17/03/2020

Ref: W-CBP-200317-01

TO WHOM IT MAY CONCERN:

Waiver: This certifies that the following research does not require clearance from the Human Research Ethics Committee (Medical).

Investigator: Mr G Otieno
Student No. (if appropriate): 1846979
Staff No. (if appropriate):

Supervisor: Dr I Maposa

School: Public Health
Epidemiology and Biostatistics
Medical School
University

Project title: *Hierarchical Bayesian estimation of the prevalence of viral acute respiratory tract infections in Kilifi, Kenya*

Reason: Review of information in the public domain.
No human participants will be involved in the study.

Dr CB Penny
Chairperson: Human Research Ethics Committee (Medical)

Research Office Secretariat:
Physical address: Phillip Tobias Building, 3rd Floor, Office 302, Corner York Road and Princess of Wales Terrace, Parktown, Johannesburg 2193.
Postal address: Private Bag 3, Wits 2050
Tel Nos. +27 (0)11-717-1234/2656/2700/1252
Office E-mail: HREC-Medical.ResearchOffice@wits.ac.za
Website:http://www.wits.ac.za/research/about-our-research/ethics-and-research-integrity/

A4. Declaration: Student's contribution to article and agreement of co-author

I, Grieven Paul Otieno, student number 1846979, declare that this Research Report is my own work and that I contributed adequately towards research findings published in the article stated below which are included in my Thesis/Dissertation/Research Report.

Signature of Student:  **Date:** 02/08/2020

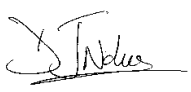
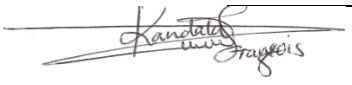
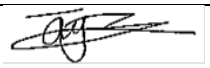
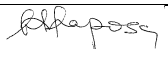
Name of primary supervisor: Innocent Maposa, PhD

Signature of primary supervisor: 

Agreement by co-authors: By signing this declaration, the co-authors listed below agree to the use of the article by the student as part of his Research Report. In cases where the student is not the 1st author of a published article, the primary supervisor must explain (under comments) why the student is entitled to use the paper for his/her degree purposes.

Article Title: Estimating the Burden of Viral Acute Respiratory Infections (ARI) in Kilifi, Kenya.

Journal name, year, volume and page numbers: The target journal is the Journal of Global Health Reports

	Name	**Signature**	**Date**
1st author	Grieven P. Otieno		02/08/2020
2nd author	D. James Nokes		02/08/2020
3rd author	Ngianga-Bakwin Kandala		02/08/2020
4th author	Joyce Nyiro		03/08/2020
4th author	Innocent Maposa		02/08/2020

Comments by primary supervisor: N/A

A5. Hierarchical Bayes SAE specification

In this study, count data were used to estimate the prevalence and incidence rates of viral respiratory infections in Kilifi, Kenya. This section provides a general perspective of the fitted models discussed in section 2.1.3.

The number of cases (y_{it}) that tested positive for any of the 7 respiratory pathogens studied where $i = 1...m$ locations and $t = 1...T$ months. Assuming that y_{it}'s are conditionally independent and follow an expondential family distribution with pdf;

$$f(y_{it}|\theta_{it}, \zeta_{it}) = exp\left[\phi_{it}^{-1}(y_{it}\theta_{it} - \psi(\theta_{it})) + \rho(y_{it}; \zeta_{it})\right] \tag{A1}$$

The density (A1) is parameterized with respect to the scale parameters ζ_{it} and the canonical / natural parameters θ_{it}. The scale parameters are assumed to be known while the canonical parameters were first modeled as;

$$h(\theta_{it}) = X_{it}^T B + B_t + u_i + v_{it} \quad (i = 1, ..., m \text{ locations and } t = 1, ..., T \text{ months}) \tag{A2}$$

where $X_{it}(p\times 1)$ are known design vectors of auxilliary variables, $\beta(p\times 1)$ is the unkown regression coefficient,B_t is the time component, u_i are the spatial random effects and v_{it} are the space-time interaction terms. It is assumed that u_i and v_{it} are mutually independent with $u_i \overset{iid}{\sim} N(0, \sigma_u^2)$ and $v_{it} \overset{iid}{\sim} N(0, \sigma^2)$. Further, let $R_u = \sigma_u^{-2}$, $R = \sigma^{-2}$, $\boldsymbol{\theta} = (\theta_{11}, ...\theta_{mt})^T$ and $\boldsymbol{U} = (u_1, ...u_m)^T$. The prior distributions, conditional on $\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\beta}, R_u = r_u$ and $R = r$ of the hierarchical model are given by the following:

1. $Y_{it}|\ \boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\beta}, r_u, r$ are independent with $f(y_{it}|\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\beta}, r_u, r) = exp\left[\phi_{it}^{-1}\left(y_{it}\theta_{it} - \psi(\theta_{it})\right) + \rho\left(y_{it}; \phi_{it}\right)\right]$
2. $h(\theta_{it})|\ \boldsymbol{u}, \boldsymbol{\beta}, r_u, r \overset{ind}{\sim} N(X_{it}^T B + u_i r_u^{-1} + v_{it}, r^{-1})$
3. $u_i|\ r_u \propto r_u^{\frac{m}{2}} exp\left[-\frac{r_u}{2}\sum_{i<j} w_{ij}(u_i - u_j)^2\right]$ where the w_{ij} are stritly positive if locations i and j are contiguous and $w_{ij} = 0$ otherwise.
4. B_t comprised of a temporaty structured (γ_t) and unstructured random (ζ_t) effects with a linearity constraint imposed on them. The temporally structured effect was modeled dynamically using a random walk process order 1 (RW1) $\gamma_t|\gamma_{t-1} \sim Normal\left(\gamma_{t-1}, \sigma^2\right)$ and a Gaussian exchangeable prior specified for the temporally structured effects; $\zeta_t \sim N\left(0, \frac{1}{\tau_\zeta}\right)$

5. $\boldsymbol{\beta}, R_u,$ and R are mutually independent with their respective prior distributions specified as:

   (a) $\boldsymbol{\beta} \sim uniform(\boldsymbol{R}^p)$

   (b) $R_u \sim gamma(\frac{1}{2}a, \frac{1}{2}b)$

   (c) $R \sim gamma(\frac{1}{2}c, \frac{1}{2}d)$

Of interest is the estimation of the posterior means, variances and covariances. These are given by:

1. $\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{u}, r_u, r, \boldsymbol{y} \sim N((\Sigma_{i,t}\boldsymbol{X}_{it}\boldsymbol{X}_{it}^T)^{-1}(\Sigma_{it}\boldsymbol{X}_{it}(h(\theta_{it}) - u_i)), r^{-1}(\Sigma_{it}\boldsymbol{X}_{it}\boldsymbol{X}_{it}^T)^{-1})$
2. $u_i|\boldsymbol{\theta}, \boldsymbol{u}, r_u, r, \boldsymbol{y} \overset{ind}{\sim} N((rn_i + r_u)^{-1}r\Sigma_k(h(\theta_{it}) - \boldsymbol{X}_{it}\boldsymbol{\beta}), (rn_i + r_u)^{-1})$
3. $R|\boldsymbol{\theta}, \boldsymbol{u}, r_u, r, \boldsymbol{y} \sim gamma\left(\frac{1}{2}(c + \Sigma_i\Sigma_t(h(\theta_{ik}) - X_{it}^T\beta - u_i)^2), \frac{1}{2}(d + \Sigma_1^m n_i)\right)$
4. $R_u|\boldsymbol{\theta}, \boldsymbol{u}, r_u, r, \boldsymbol{y} \sim gamma\left(\frac{1}{2}(a + \Sigma_i u_i^2), \frac{1}{2}(b + m)\right)$
5. $\theta_{it}|\boldsymbol{\theta}, \boldsymbol{u}, r_u, r, \boldsymbol{y} \overset{ind}{\sim} \pi\left(\theta_{it}|\boldsymbol{\theta}, \boldsymbol{u}, r_u, r, \boldsymbol{y}\right) \propto exp\left[(y_{it} - \theta_{it} - \psi(\theta_{it}))\phi_{it}^{-1} - (\frac{r}{2})(h(\theta_{it}) - X_{it}^T\beta) - u_{it})^2\right] h'(\theta_{it})$

A6: Autocorrelation Plots

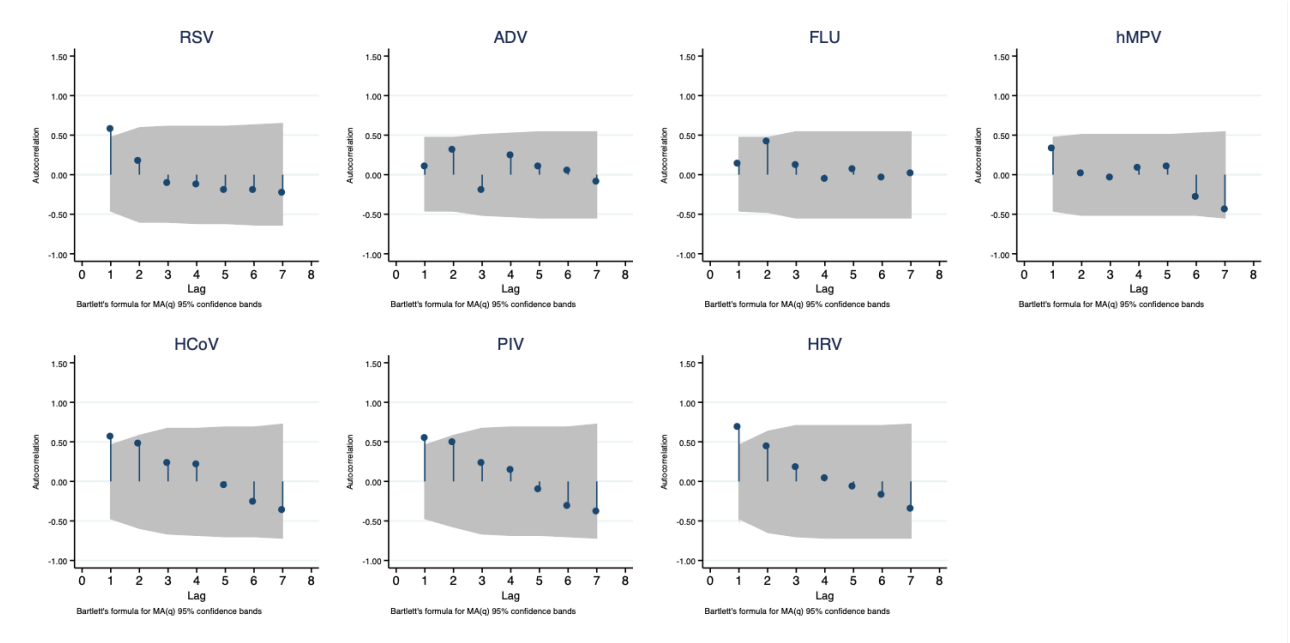


Figure 4.9: Autocorrelation plots for each virus pathogen tested in Kilifi.

A7. Prevalence WAIC values

Table A.1: Watanabe–Akaike information criterion (WAIC) used for comparisson of models used estimate the prevalence of viral ARI in Kilifi.

Model	RSV	HRV	HCoV	FLU	PIV	ADV	hMPV
All ages							
Poi	467.50	645.63	500.61	543.91	489.13	456.28	289.54
Zip	471.43	739.45	544.30	649.69	525.13	453.67	275.39
Bin	456.23	646.44	493.15	534.84	491.10	449.60	291.04
Zib	595.13	744.65	546.47	544.94	643.15	453.57	298.14
Logit n	530.48	347.40	483.09	486.87	487.24	516.54	700.84
Under 5s							
Poi	469.17	646.18	499.10	542.91	488.00	448.58	287.67
Zip	594.62	812.05	681.84	658.00	488.61	450.19	294.29
Bin	457.68	645.10	490.89	533.96	489.96	448.85	288.98
Zib	460.86	744.96	547.59	668.86	526.21	454.93	295.81
Logit n	532.08	349.19	485.48	484.86	487.12	522.10	677.49
Over 5s							
Poi	470.22	644.62	498.03	543.30	488.40	449.60	289.50
Zip	588.25	738.50	548.42	657.03	524.79	456.05	293.14
Bin	458.82	644.09	490.40	534.17	490.34	449.68	290.92
Zib	606.75	810.93	550.84	667.84	526.18	451.33	294.57
Logit n	521.91	337.05	478.15	476.08	475.53	515.91	688.43

[6]

[6]The row named "Model" indicate the model which offered the best fit on the data by DIC values; Logit n= Logit normal, Bin=Binomial, Zib=Zero-inflated binomial, Poi=Poisson, Zip=Zero-inflated Poisson.

A8. Incidence WAIC values

Table A.2: Watanabe–Akaike information criterion (WAIC) used for comparisson of models used estimate the incidence rates of viral ARI in Kilifi.

Model	RSV	HRV	HCoV	FLU	PIV	ADV	hMPV
All ages							
Poi	413.72	607.71	435.37	480.56	447.19	368.01	232.200
Zip	540.76	708.83	486.54	494.58	482.53	389.21	254.240
Bin	413.29	607.61	437.11	480.23	447.44	367.96	232.190
Zib	541.45	709.25	486.69	574.62	482.62	389.24	259.030
Logit n	566.41	383.83	541.90	533.45	508.96	576.86	724.260
Under 5s							
Poi	376.71	550.05	395.13	411.35	407.16	357.21	241.72 [7]
Zip	503.97	637.38	410.92	639.66	426.51	351.88	604.61
Bin	374.39	550.80	395.49	411.81	407.43	357.39	241.95
Zib	506.79	638.74	411.32	640.36	426.86	351.90	606.52
Logit n	589.08	425.21	593.64	619.95	552.41	595.98	714.96
Over 5s							
Poi	217.38	415.04	281.07	369.00	237.51	141.39	1125.71
Zip	731.86	468.90	293.51	418.28	256.74	783.42	622.22
Bin	217.49	415.12	281.10	368.12	237.50	141.41	6600.09
Zib	247.61	469.00	293.54	418.44	256.74	137.49	623.83
Logit n	705.98	514.72	647.44	605.17	659.77	747.28	773.56

[7]The row named "Model" indicate the model which offered the best fit on the data by DIC values; Logit n= Logit normal, Bin=Binomial, Zib=Zero-inflated binomial, Poi=Poisson, Zip=Zero-inflated Poisson.

A9. R packages

Below is a list of R packages that were used in this analysis:

tidyr

Hadley Wickham and Lionel Henry (2019). tidyr: Tidy Messy Data. R package version 1.0.0. https://CRAN.R-project.org/package=tidyr

reshape2

Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.

dplyr

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.3. https://CRAN.R-project.org/package=dplyr

purrr

Lionel Henry and Hadley Wickham (2019). purrr: Functional Programming Tools. R package version 0.3.3. https://CRAN.R-project.org/package=purrr

rgeos

Roger Bivand and Colin Rundel (2019). rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-2. https://CRAN.R-project.org/package=rgeos

maptools

Roger Bivand and Nicholas Lewin-Koh (2019). maptools: Tools for Handling Spatial Objects. R package version 0.9-8. https://CRAN.R-project.org/package=maptools

broom

David Robinson and Alex Hayes (2019). broom: Convert Statistical Analysis Objects into Tidy Tibbles. R package version 0.5.2. https://CRAN.R-project.org/package=broom

rgdal

Roger Bivand, Tim Keitt and Barry Rowlingson (2019). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-4. https://CRAN.R-project.org/package=rgdal

rgeos

Roger Bivand and Colin Rundel (2019). rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-2. https://CRAN.R-project.org/package=rgeos

haven

Hadley Wickham and Evan Miller (2019). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.1.1. https://CRAN.R-project.org/package=haven

RColorBrewer

Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer

classInt

Roger Bivand (2019). classInt: Choose Univariate Class Intervals. R package version 0.4-2. https://CRAN.R-project.org/package=classInt

spdep

Bivand, Roger S. and Wong, David W. S. (2018) Comparing implementations of global and local indicators of spatial association TEST, 27(3), 716-748. URL https://doi.org/10.1007/s11749-018-0599-x

Roger S. Bivand, Edzer Pebesma, Virgilio Gomez-Rubio, 2013. Applied spatial data analysis with R, Second edition. Springer, NY. http://www.asdar-book.org/

ggplot2

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

cowplot

Claus O. Wilke (2019). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.0.0. https://CRAN.R-project.org/package=cowplot

gridExtra

Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

INLA

Havard Rue, Sara Martino, and Nicholas Chopin (2009), Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion), Journal of the Royal Statistical Society B, 71, 319-392.

Thiago G. Martins, Daniel Simpson, Finn Lindgren and Havard Rue (2013), Bayesian computing with INLA: New features, Computational Statistics and Data Analysis, 67(2013) 68-83

Finn Lindgren, Havard Rue, and Johan Lindstrom (2011). An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach (with discussion), Journal of the Royal Statistical Society B, 73(4), 423-498.

Finn Lindgren, Havard Rue (2015). Bayesian Spatial Modelling with R-INLA. Journal of Statistical Software, 63(19), 1-25. URL http://www.jstatsoft.org/v63/i19/.

A10. INLA models R Codes

```
# Poisson models
# model formula
formula<-y~1 + age + avg_rain + facilities +
        f(locid,model="bym",graph=H, scale.model=T,adjust.for.con.comp=TRUE,
          hyper=list(prec.unstruct=list(prior="loggamma",param=c(1,0.01)),
                     prec.spatial=list(prior="loggamma",param=c(1,0.01)))) +
        f(month,model="iid") +
        f(locmonth,model="rw1") +
        f(locmonth1,model="iid") +
        f(loctime,model="iid")


model<-inla(formula,family="poisson",data=sub.dat,E = E,
              control.results=list(return.marginals.random=TRUE,
                                   return.marginals.predictor=TRUE),
              control.predictor=list(compute=TRUE,link=1),
              control.compute=list(dic=TRUE,cpo=TRUE,waic=TRUE))

# Zero inflated Poisson model
model<-inla(formula,family="zeroinflatedpoisson1",data=sub.dat,E = E,
              control.results=list(return.marginals.random=TRUE,
                                   return.marginals.predictor=TRUE),
              control.predictor=list(compute=TRUE,link=1),
              control.compute=list(dic=TRUE,cpo=TRUE,waic=TRUE),
              control.inla = list(tolerance = 1e-30, h = 1e-25))

# Binomial models

model<- inla(formula, family = "binomial", data = sub.dat, Ntrials=N,
              control.results=list(return.marginals.random=TRUE,
                                           return.marginals.predictor=TRUE),
              control.predictor=list(compute=TRUE,link=1),
              control.compute=list(dic=TRUE,cpo=TRUE,waic=TRUE))

# Zero-inflated binomial models

model<- inla(formula, family = "zeroinflatedbinomial1", data = sub.dat, Ntrials=N,
                         control.results=list(return.marginals.random=TRUE,
                                              return.marginals.predictor=TRUE),
                         control.predictor=list(compute=TRUE,link=1),
                         control.compute=list(dic=TRUE,cpo=TRUE,waic=TRUE),
                         control.inla = list(tolerance = 1e-30, h = 1e-25))


# Logistic Normal model
```

```
formula = p.logit ~ 1 + age + avg_rain + facilities +
                    f(locid,model="bym",graph=H, scale.model=T,
                      hyper=list(prec.unstruct=list(prior="loggamma",
                                                    param=c(0.1,0.001)),
                      prec.spatial=list(prior="loggamma",param=c(0.1,0.001)))) +
                      f(month,model="ar1") +
                      f(month1,model="iid") +
                      f(locmonth,model="iid")


model <- inla(formula,family = "gaussian", data =sub.dat,
          control.predictor=list(compute=TRUE),
          control.family=list(hyper=list(prec=list(initial=log(1),fixed=TRUE))),
          control.compute=list(dic=TRUE, cpo=TRUE, waic=TRUE),
          scale=logit.prec)
```

A11. Instruction for Authors

Authorship criteria

JoGHR subscribes to the authorship criteria developed by the International Committee of Medical Journal Editors (ICMJE) : "The ICMJE recommends that authorship be based on the following 4 criteria: 1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; and 2) Drafting the work or revising it critically for important intellectual content; and 3) Final approval of the version to be published; and 4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved." However, we are aware of the limitations of the ICMJE definition and its practical applications, and ask our authors to write in their own words why they think they deserve authorship of the submitted manuscript. Authors' declared contributions to the research are published at the end of the article.

Changes in authorship

Requests for changes in authorship will only be permitted where valid reasons are provided and all authors are in agreement with the change. Requests for changes to authorship must be directed to the Editors-in-Chief. We follow COPE guidelines for authorship changes in manuscripts before and after publication.

Conflicts of interest

JoGHR uses the ICMJE uniform disclosure form for reporting all financial and personal relationships that might bias the authors' work. We ask authors to fill out the form available at the ICMJE site. The conflict of interest declarations are held by the corresponding author, who must make them available on request. All conflict of interests are declared and published in each manuscript, and where no conflict of interest is declared, this is included too. JoGHR also asks its reviewers to declare possible conflicts of interest (including financial and personal relationships) related to the manuscripts under review. See Editorial Research for policy on editorial research and conflicts of interest.

Research integrity

The Editorial Board of JoGHR is devoted to the promotion of scientific integrity as a vital component of the research process. JoGHR follows the ethics flowcharts developed by the Committee on Publication Ethics (COPE) for dealing with cases of possible misconduct. The COPE flowcharts are available here.

Results and data integrity

Authors are required to be honest in presenting the results and conclusions of their research. Data falsification (including fabrication and selective reporting of data) is unethical. Cases of suspected research misconduct are thoroughly investigated by JoGHR editorial board, including contacting the institution where the research was performed. Our journal staff and editors are members of relevant editorial associations with necessary experience on how to detect and manage possible misconduct.

Plagiarism

Reproducing text from other papers without properly crediting the source (plagiarism) or producing many papers with almost the same content by the same authors (self-plagiarism) is not acceptable. Submitting the same results to more than one journal concurrently is also unacceptable. Please note that submitted manuscripts may be subject to checks in order to detect instances of overlapping and similar text. For more information about plagiarism checks, please visit Crossref site.

Handling allegations of misconduct

JoGHR is aware of the potential impact an allegation of ethical misconduct can have upon a researcher's career, as well that there are differing definitions of scientific misconduct. All allegations of ethical misconduct are taken seriously, and a full investigation will take place. The Editors-in-Chief should always be the first point of contact, and will seek clarification from all affected parties, in accordance with the guidance from COPE flowcharts. (provide link here). The Editor-in-Chief will not be influenced by other parties and will make decisions in an unbiased and objective manner. At all times, the Editor-in-Chief will remain neutral in tone, acting with integrity, and educating where possible. Where the allegation is made against the Editor-in-Chief, this should be sent to the Publisher for further investigation. If JoGHR is approached by a third party with an allegation of plagiarism, the Editor-in-Chief will always seek a response from the original author(s) or copyright holder(s) before a recommendation is made. JOGHR reserves the right not to proceed with a case if the complainant presents a false name or affiliation, or acts in an inappropriate or threatening manner towards the journals editors and staff.

Corrections and retractions

All authors have an obligation to inform and collaborate with journal editors to provide prompt retractions or correction of errors in published works. JOGHR editors are committed to preserve the integrity of the published record. The journal will issue retractions if:

- The findings have previously been published elsewhere without proper crossreferencing, permission or justification;
- There is clear evidence that the findings are unreliable, either as a result of misconduct or honest error;
- The published article has high text similarity to previously published material;
- The published articles report unethical research.

The journal will issue correction if:

- A small portion of an otherwise reliable publication proves to be misleading (especially because of honest error);
- The author list is incorrect (i.e. a deserving author has been omitted or somebody who does not meet authorship criteria has been included).

Data availability

JoGHR follows COPE guidelines on data availability. The results of research should be recorded and maintained, and data retained for a reasonable period. JoGHR encourages authors to make their data available upon reasonable request following publication and include a statement on this

in their submission. Exceptions may be appropriate in certain circumstances to preserve privacy, or assure patent protection.

Research involving human subjects

If the work involves the use of human subjects, the author should ensure that the work described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans; Uniform Requirements for manuscripts submitted to Biomedical journals. Authors should include a statement in the manuscript that Informed Consent was obtained for research, studies or experimentation with human subjects. The privacy rights of human subjects must always be observed in the conduct of research.

LICENCE

JoGHR publishes under CC BY 4.0 licence, which means that all published articles can be:

- Shared - copy and redistribute the material in any medium or format; and/or
- Adapted - remix, transform and build upon the material for any purpose, even commercially.

JoGHR's publisher cannot revoke this, provided that these terms are followed:

- Attribution. You must give appropriate credit, provide a link to the licence, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the publisher endorses you or your use.
- No additional restrictions. You may not apply legal terms or technological measures that legally restrict others from doing anything the CC BY 4.0 licence permits.

COPYRIGHT

The copyright for all published articles remains with the author(s), and journal copyright remains with the journal.