Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Sarah Hamid	8 April 2022
	Date

Etiology and patterns of acute respiratory infections in low- and middle-income countries: implications for vaccination strategies and environmental interventions

By

Sarah Hamid
Doctor of Philosophy
Epidemiology

Thomas F. Clasen
Advisor

Benjamin Lopman
Committee Member

Robert Lyles
Committee Member

Eric McCollum
Committee Member

John P. McCracken
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D, MPH
Dean of the James T. Laney School of Graduate Studies

Date

Etiology and patterns of acute respiratory infections in low- and middle-income countries: implications for vaccination strategies and environmental interventions

By

Sarah Hamid
A.B., Brown University, 2007
M.P.H., Yale University, 2012

Advisor: Thomas F. Clasen, J.D., PhD

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Epidemiology
2022

Abstract

Acute respiratory infections (ARI) are a leading cause of mortality, especially in low- and middle-income countries (LMIC). Respiratory syncytial virus (RSV) and *Streptococcus pneumoniae* are recognized as leading causes of lower respiratory infection morbidity and mortality. Household air pollution (HAP) from cooking with solid fuels is the leading environmental risk factor. Quantifying the etiologic role of respiratory pathogens, describing their patterns, and identifying effective interventions targeting the environmental drivers of disease are essential to reduce respiratory disease burden. The overarching goal of this dissertation is to shed light on the etiology and patterns of ARI in LMIC in order to better inform vaccine strategies and environmental interventions.

In **Aim 1**, we characterized RSV seasonality in Guatemala. We found substantial variability in the timing of seasonal epidemics such that two differential patterns of RSV seasonality were identified: an early season starting in June-July and a late season starting in October-November. This variability suggests that age-based vaccination would be more effective than seasonal vaccination.

In **Aim 2**, we assessed whether prenatal HAP exposure is associated with respiratory illness in two-year-old children using data from a liquified petroleum gas stove intervention during gestation and the first year of life in Guatemala, India, and Rwanda. In an intent-to-treat analysis, we did not find an effect of the intervention on illness with a cough in two-year-old children. Similarly, we did not find evidence of an association between HAP and illness with cough in an exposure-response analysis.

In **Aim 3**, we estimated the fraction of hospitalized ARI attributable to *S. pneumoniae,* and assessed whether a semi-quantitative measure of bacterial load (PCR quantification cycle [Cq] values) could improve understanding of the etiologic role of *S. pneumoniae* in hospitalized ARI in adults in six LMIC. Population attributable fraction estimates that incorporated Cq values were higher than those that relied on qualitative PCR. The proportion of hospitalized ARI attributed to *S. pneumoniae* varied across countries, ranging from 0.1% to 18.5%.

These findings further our understanding of the etiology, patterns, and environmental risk factors of ARI in low-resource settings, and can inform vaccine strategies, environmental interventions, and healthcare management practices.

Etiology and patterns of acute respiratory infections in low- and middle-income countries: implications for vaccination strategies and environmental interventions

By

Sarah Hamid
M.P.H., Yale University, 2012
A.B., Brown University, 2007

Advisor: Thomas Clasen, JD, PhD

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Epidemiology
2022

Acknowledgements

I received a great deal of support and assistance during my doctoral studies. I would first like to thank my dissertation committee, who provided both methodological and moral support throughout this process: Tom Clasen, my committee chair, for generously welcoming me into his research group, giving me the freedom to pursue my interests, helping me overcome hurdles in the dissertation process, and providing unwavering encouragement; John McCracken, for his close mentorship and support with the conceptual development of this dissertation; Bob Lyles and Ben Lopman, for generously providing their time and methodological feedback; and Eric McCollum, for sharing his clinical expertise. Additionally, I would like to thank the Department of Epidemiology at Emory University, especially Penny Howards, who supported me during a difficult time in my training.

I would like to extend my sincere thanks to the VICo team in Guatemala, TAC study investigators at CDC, and the HAPIN investigators, without whom this dissertation would not have been possible. Special thanks to Anaite Diaz, Oscar de Leon, and others at Universidad del Valle de Guatemala, who hosted me while I explored dissertation topics and provided access to data. Thanks to Jenny Milucky, Chris Van Beneden, and Jonas Winchell for sharing data from the TAC study and for their helpful suggestions and insight into laboratory methods. Thanks also to Jennifer Peel, who provided advice and feedback on my HAPIN analysis plan.

Finally, I would like to thank my family and friends: my parents, who supported my early academic achievements and always had my back; my twin sister for being my pandemic companion; my fellow Epidemiology PhD students--especially Katie Labgold, Zerleen Quader, and Katie Ross, who gave their friendship, emotional support, and a peer accountability structure during my dissertation years. Finally, I would like to thank Yuzo Arima, who supported my decision to pursue doctoral studies and patiently encouraged me every step of the way. He has been my partner, friend, coach, and sounding board throughout my doctoral training.

Table of Contents

Tables and Figures

Chapter 1. Introduction

Acute respiratory infections (ARI), comprising both upper and lower respiratory tract infections, exert a considerable burden and cost on health care systems.[1,2] In 2019, lower respiratory tract infections (LRI), defined as pneumonia or bronchiolitis, resulted in 489 million incident cases and an estimated 2.5 million deaths and among all ages, making them the fourth leading cause of mortality for all ages and the second leading cause of death among children younger than 5 years worldwide.[2] While upper respiratory infections (including cough, acute nasopharyngitis, sinusitis, pharyngitis, tonsillitis, laryngitis, tracheitis, epiglottitis, rhinitis, rhinosinusitis, rhinopharyngitis, and supraglottitis) typically do not result in severe disease, they can significantly impair quality of life and productivity. In 2019, there were an estimated 17.2 billion incident cases of upper respiratory tract infections worldwide, contributing to 9,460 deaths and 6.39 million DALYS.[2]

Low- and middle-income countries bear a disproportionate burden of LRI, which strongly correlates with poverty. Historically, poverty was blamed for LRI, and it remains an important underlying cause.[3] While microorganisms are a necessary (but not sufficient) cause of LRI, people at highest risk for contracting or dying from LRI often lack access to adequate nutrition, clean cooking fuel, vaccines, and WASH (water, sanitation and hygiene).[4] Thus, the causes and risk factors for LRIs can be thought of in terms of the traditional epidemiologic triad: the host, the microorganism, and the environment. Disease results from the interaction between the microorganism and the susceptible host in an environment that supports transmission of the microorganism and contributes to the vulnerability of the host.

In terms of microorganisms, modeling studies have attributed the majority of LRI deaths to four etiologies, namely *Haemophilus influenzae* type B (HiB), *Streptococcus pneumoniae*, influenza virus, and respiratory syncytial virus (RSV).[5] However, the etiological agents of LRI include a wide range of

bacterial and viral pathogens, and in a given case of LRI, it is difficult to determine the causal microbial agent.[6] This difficulty arises because many pathogens, particularly bacteria, colonize the upper respiratory tract, and it is difficult to distinguish pathogen carriage from infection. For example, *S. pneumoniae* is known to be an important cause of LRI, but a case-control study of severe acute respiratory infections (SARI) in adults found that it was detected with similar frequency in SARI cases and asymptomatic adults, suggesting it has a minor etiologic role in SARI.[7]

An approach proposed for differentiating pathogen carriage from clinically significant infection is the quantification of pathogen load.[8,9] Higher pathogen load in the upper respiratory tract has been associated with pneumonia, and for some respiratory pathogens, it has been associated with more severe outcomes.[10-16] For diarrheal illness, the quantity of nucleic acid in a specimen is also thought to distinguish clinical infection from asymptomatic shedding in children.[17-22] Since 2017, re-analyses of two major studies of the etiology and burden of pediatric diarrheal infections have used quantitative PCR to adjust population attributable fractions for the high prevalence of asymptomatic pathogen carriage.[17,23] A similar approach could be used to improve understanding of respiratory disease etiology.

In terms of environmental causes, exposure to household air pollution (HAP), largely from the use of solid fuels for cooking, is a leading risk factor for childhood pneumonia.[24,25] HAP exposure likely begins to impact lung development during gestation, increasing the risk of future respiratory disease.[26] To date, however, interventions to reduce HAP have struggled to show reductions in pneumonia incidence.[27,28] A possible reason for the inability of past interventions to show an effect on respiratory outcomes is that they did not achieve sufficient exposure contrast between intervention and control arms.[28] The recently completed Household Air Pollution Intervention Network (HAPIN) trial has overcome this limitation, achieving a substantial reduction in personal

exposure to fine particulate matter ($PM_{2.5}$).[29] HAPIN provided liquefied petroleum gas (LPG) stoves and fuel to intervention households with pregnant women in four countries. HAPIN children from three of the original four study sites are participating in a cohort study to assess the longer-term health impacts of the intervention. Analysis of these data will improve our understanding of the impact of gestational and infant HAP exposure on respiratory health in early childhood.

Other important environmental drivers of respiratory infection are climate and weather patterns. While we cannot intervene to change these factors, understanding the seasonality of respiratory pathogens can guide decisions about when to deploy vaccines, non-pharmaceutical interventions, and therapeutics. RSV is the most common pathogen identified in children with pneumonia and is an important future vaccine target, with many candidate vaccines in clinical development.[30-32] To identify optimal vaccination strategies and provide a baseline to assess possible future vaccine effects, it is important to characterize RSV seasonality. An eventual vaccine strategy will likely include a combination of maternal and infant immunization. A maternal vaccine would provide passive immunity to infants but the protection conferred would be of limited duration. Understanding RSV seasonality is important to ensure that maternal vaccines are administered during periods that will provide protection to infants during the RSV season.[33] Global reviews have served as a guide to RSV seasonality[34] but characterizing local seasonality patterns is needed to inform effective national vaccine strategies. Thus, WHO has identified the description of local seasonality patterns in RSV incidence as a priority research activity.[35]

Reducing the global burden of ARI requires effective interventions targeting the different causes of disease. The most effective interventions for reducing ARI are vaccinations against respiratory pathogens. In order to prioritize pathogens for vaccine development and to guide national vaccine policy, it is important to estimate the proportion of disease attributable to specific etiologic agents.

Once vaccines are developed, effectively deploying them and assessing their impact requires knowledge of pathogen seasonality. Reducing the burden of ARI also requires identifying effective interventions to reduce environmental factors, such as HAP, that affect the infectious agent and host defenses. The **overarching research goal** of this dissertation is to shed light on the etiology and patterns of acute respiratory infections in low- and middle-income settings in order to better inform vaccine strategies, environmental interventions, and healthcare management practices.

Specific aims

AIM 1. To characterize RSV seasonality in Guatemala.

Aim 1 uses data from *Vigilancia Integrada Comunitaria* (VICo), an integrated infectious disease surveillance system in Guatemala. Among other syndromes, VICo includes surveillance of hospitalized ARI. ARI cases were tested for a range of pathogens including RSV.

AIM 2. To assess whether prenatal HAP is associated with the prevalence of illness with a cough in the second year of life.

Aim 2a. To estimate the longer-term effect of an LPG stove intervention during gestation and the first year of life on the prevalence of illness with a cough in children at 24 months of age.

Hypothesis: children born in intervention households have lower prevalence of illness with a cough at age 2 years relative to those born in control households.

Aim 2b. To estimate the effect of prenatal HAP exposure on the prevalence of illness with a cough in children at 24 months of age.

Hypothesis: exposure to HAP pollutants during gestation (a critical developmental period) is positively associated with the prevalence of illness with a cough at age 2 years.

Aim 2 uses data from the HAPIN trial, which provided LPG stoves and fuel to intervention households with pregnant women in four countries, and followed children until 1 year of age to assess a range of health outcomes, including pneumonia. Control households cooked primarily with solid biomass fuels. To assess longer-term health impacts, including the prevalence of respiratory infections, three of the original four study sites (Guatemala, India, and Rwanda) are participating in a cohort study to follow HAPIN children through age 2 years.

AIM 3. To determine whether bacterial load (as measured by real-time PCR Cycle threshold (Ct) values) can improve understanding of SARI etiology.

AIM 3a. To assess the association between *S. pneumoniae* load in the upper respiratory tract and SARI among adults in six low- and middle-income countries.

Hypothesis: On average, among adults who test positive *for S. pneumoniae* by PCR, adults with SARI have lower Ct values (i.e. higher bacterial loads) than asymptomatic adults.

AIM 3b. To estimate the proportion of SARI attributable to *S. pneumoniae* among adults in six low- and middle-income countries, incorporating the association between PCR Ct values and SARI.

Hypothesis: Compared to a conventional analysis using qualitative PCR results, using PCR Ct values from quantitative PCR will yield a higher estimated population attributable fraction for *S. pneumoniae*.

Aim 3 uses data from the TAC study, a case-control study of community acquired pneumonia among adults in Bangladesh, China, Egypt, Guatemala, Kenya, and Thailand. Upper respiratory tract samples were collected from SARI cases and asymptomatic adults in each country and tested for a panel of respiratory pathogens using the Taqman Array Card.

Chapter 2. Background and Literature review

Case definitions

Pneumonia is of keen interest to public health researchers because of its high disease burden and severity. However, it is not straightforward to define so studies use different, usually non-specific case definitions depending on their purpose (e.g. clinical management, research, surveillance).[3] The various commonly-used case definitions cover a broad spectrum of disease severity. The definitions below provide a general description of commonly used case definitions in research and surveillance but these case definitions vary over time and across studies. This dissertation uses data from three different studies, each with its own case definition of acute respiratory infections.

Lower respiratory tract infection (LRI)

LRI refer to all infections that extend into the chest. The Global Burden of Disease Study defines lower respiratory tract infections as pneumonia or bronchiolitis.[36]

Pneumonia

Pneumonia is the most severe manifestation of lower respiratory tract infections. It is an infection of the lung caused predominantly by bacteria, viruses, or fungi. Pathological definitions of pneumonia are generally considered the gold standard and usually define pneumonia as alveolar inflammation in which the alveoli fill with pus and other liquid. However, pathological examinations of the lung are often not feasible.

Acute respiratory infection (ARI)

ARI covers all infections of the respiratory system with a duration less than 2 weeks. Most of these infections are limited to the upper respiratory tract (i.e. the nose and throat) such as pharyngitis, tonsillitis, sinusitis, and laryngitis. ARI is commonly used for surveillance purposes but the specific case definition has varied over time and across surveillance systems. For surveillance purposes, it is

typically defined as illness with a sudden onset of symptoms and at least one respiratory system (shortness of breath, cough, sore throat, or coryza).[37,38]

Severe acute respiratory infection (SARI)

SARI case definitions have changed over time but refer to a subset of ARI requiring hospitalization.[39]

Disease burden

Worldwide, LRIs contributed to 2.49 million (2.27–2.74) deaths, representing over 4% of global deaths among all ages in 2019.[40] Among children younger than 5 years of age, they were the second leading cause of death after neonatal disorders, accounting for over 13% of deaths in this age group in 2019.[41] Proportionate mortality is u-shaped, declining with increasing age after early childhood and increasing among older adults. In people aged ≥70 years, LRI comprise over 4% of deaths (1.23 million [1.06-1.32]). Upper respiratory infections comprise a much smaller proportion of global mortality, contributing to 9,460 deaths (5,540–14,900) among all ages in 2019.[40] However, they result in substantial morbidity and burden on healthcare systems. In 2019, there were an estimated 17·2 billion (95% UI 15·4–19·3) incident cases and 237 million (212–265) prevalent cases of upper respiratory infections.[40]

Between 2000 and 2019, under-5 LRI mortality decreased markedly following reductions in risk factors for LRI.[2,42] Where they have been introduced, *Haemophilus influenzae* type b vaccine and pneumococcal conjugate vaccine have shifted the epidemiology and etiology of respiratory infections. However, parallel improvements have not been seen in adults for whom the etiology of and risk factors for pneumonia differ. The causes and risk factors for LRIs can be thought of in terms of the traditional epidemiologic triad: the host, the microorganism, and the environment.

Disease results from the interaction between the microorganism and the susceptible host in an environment that supports transmission of the microorganism from a source to the host.

Host risk factors among all age groups include sex, malnutrition, comorbidities (e.g. HIV), and smoking/secondhand smoke.[43] In infants, inadequate nutrition and feeding practices (e.g. non-exclusive breastfeeding), short gestation, and low birth weight are also recognized as risk factors. Environmental risk factors include exposure to household air pollution from solid fuels, ambient particulate matter, crowding, poor hygiene, and the climate or seasonality of ARI. In terms of the microorganism or etiologic agent of disease, key risk factors for pediatric pneumonia include incomplete vaccination against respiratory pathogens including *Haemophilus influenzae* type b, pneumococcal infections, measles, and pertussis. In adults, pneumococcal and influenza vaccination prevent disease.[43]

Etiologic agents

Identifying the etiologic agents of ARI is important in order to select vaccine targets and vaccination strategies. Although modeling studies have attributed the majority of LRI deaths to four etiologies (HiB, *Streptococcus pneumoniae*, influenza virus, and RSV),[5] respiratory infections can be caused by many bacterial and viral pathogens. For this reason, a typical respiratory pathogen panel tests for many viral and bacterial targets, including adenovirus, coronavirus (229E, HKU1, NL63, OC43), human metapneumovirus, human rhinovirus/enterovirus, influenza viruses, parainfluenza viruses, respiratory syncytial virus, *Bordetella pertussis*, *Haemophilus influenzae* type B (HiB), *Streptococcus pneumoniae*, and *Mycoplasma pneumoniae*. Before the advent of molecular diagnostics, understanding the etiology of lower respiratory tract infections was limited because testing for multiple viruses and bacteria required an assortment of assay modalities, which were cumbersome and resource-intensive.[44] Improved diagnostics, including multiplex real-time PCR assays, have allowed for more

comprehensive investigations of the etiology of pneumonia as demonstrated by the Pneumonia Etiology Research for Child Health (PERCH) study and the Etiology of Pneumonia in the Community (EPIC) study.

The PERCH study was designed to shed light on the etiology of childhood pneumonia in developing countries using different types of specimens and laboratory tests.[44] From 2011 to 2014, the study enrolled children under 5 years of age with pneumonia and asymptomatic controls in seven countries and tested them for over 30 potential pathogens.[30] A small set of pathogens (respiratory syncytial virus, parainfluenza virus, human metapneumovirus, influenza virus, *S. pneumoniae, Haemophilus influenzae type b (Hib), H influenzae non-type b*, and *Pneumocystis jirovecii*) accounted for most hospitalized pneumonia cases but the etiological fraction of pneumonia differed by age. Viruses accounted for 61.4% of pneumonia cases, bacteria for 27.3%, and *M. tuberculosis* for 5.9%. The study greatly improved our understanding of pneumonia etiology in children, providing an evidence base for strategies to accelerate reductions in pediatric pneumonia morbidity and mortality.[45]

The EPIC study investigated the etiology of pneumonia among both children and adults in the U.S. between 2010 and 2012.[46] The study prospectively enrolled 2,024 patients with community acquired pneumonia and 759 asymptomatic controls to compare the prevalence of 13 viruses in the upper respiratory tract of cases and asymptomatic controls. The study found that detections of influenza, RSV, and human metapneumovirus indicate an etiologic role in community acquired pneumonia but no evidence for a causal role of parainfluenza, coronaviruses, rhinovirus, and adenovirus. These findings suggested that the viral etiology of pneumonia may have been previously underestimated because of a limited range of diagnostic methods.[47] Although EPIC and several other studies that used molecular diagnostics provided important information on the etiology of community acquired

pneumonia in adults, most were conducted in the developed world and their findings may not be generalizable to developing countries.[48-50]

The "Study of the etiology of community-acquired pneumonia in adults: Use of TAC multiple pathogen detection platforms in the International Emerging Infections Program (IEIP) sites" (TAC study) was a prospective case-control study that used a real-time PCR-based multiple pathogen detection platform to understand the etiology of severe respiratory disease among hospitalized adults across six low- and middle-income countries in Africa, Asia, and the Americas. The main study analysis provided estimates of the proportion of SARI in adults caused by specific pathogens using a Bayesian analytic method developed for the PERCH study.[51] Nevertheless, the accurate determination of the specific causes of SARI was limited because of the high prevalence of asymptomatic carriage of some pathogens—particularly bacteria--and ability of PCR to detect small amounts of nucleic acid in upper respiratory tract samples of both SARI cases and adults without respiratory symptoms. While the most commonly detected pathogens among TAC study participants were *Streptococcus pneumoniae* and *Haemophilus influenzae,* these bacterial pathogens were not found to be the most common causes of SARI because there were also high numbers of detections in asymptomatic adults, indicating colonization.

Streptococcus pneumoniae

The aforementioned etiology studies demonstrated that most cases of pneumonia in high-income and low- and middle-income countries are caused by viruses. However, the proportion of pneumonia cases caused by bacteria increases with the severity of disease.[3] In children in low- and middle-income countries, and in high-income countries in the much of the 20th century, the most common causes of bacterial pneumonia are/were *Haemophilus influenzae* and *Streptococcus pneumoniae.*[3] However, the epidemiology of pneumonia is evolving with the introduction of vaccines against these

pathogens. It is estimated that pneumococcal deaths declined by 51% (7–74) and Hib deaths by 90% (78–96) from 2000 to 2015.[52] Globally, an estimated 294 000 pneumococcal deaths (uncertainty range [UR] 192 000–366 000) and 29 500 Hib deaths (18 400–40 700) occurred in HIV-uninfected children aged 1–59 month in 2015.[52] Approximately 50% of all pneumococcal deaths in 2015 occurred in four countries in Africa and Asia: India (68 700 deaths, UR 44 600–86 100), Nigeria (49 000 deaths, 32 400–59 000), the Democratic Republic of the Congo (14 500 deaths, 9300–18 700), and Pakistan (14 400 deaths, 9700–17 000]).[52]

Vaccine probe studies suggest that pneumococcus causes a substantial proportion of radiologically confirmed pneumonia, a smaller proportion of severe, non-radiologically confirmed pneumonia, and a smaller proportion still of non-severe pneumonia.[53] Results from a pediatric PCV trial in the Gambia estimated the vaccine effectiveness for radiologically confirmed pneumonia was 37%, suggesting that at least a third of all pneumonia episodes are caused by the nine serotypes of pneumococci included in the vaccine[54]

In adults, *Streptococcus pneumoniae* is the most common bacterial cause of pneumonia. Although pneumococci were found in just 5% of adults in the EPIC study,[55] CDC estimates that pneumococci account for 10% to 30% of adult community-acquired pneumonia.[56] PCV13 trials suggest that pneumococci are responsible for 6% to 11% of hospitalized community acquired pneumonia, and between 4% and 12% of primary and secondary care pneumonia outcomes.[57] Possible reasons for the variability in estimates could relate to differences in the populations studied including rates of asymptomatic carriage, outcome definitions, and diagnostics used.

Accurately estimating the etiologic role *of S. pneumoniae* in pneumonia is complicated by the high rate of carriage of pneumococci in the respiratory tract. While the absence of pneumococci excludes it as

an etiological agent, its detection in non-sterile sites (e.g. the upper respiratory tract) could be attributed to either infection or asymptomatic carriage. Pneumococci can be isolated from the nasopharynx of 5% to 90% of healthy persons. Rates of asymptomatic carriage vary with age, environment, and the presence of upper respiratory infections. Among school-age children, colonization ranges from 20% to 60%. In contrast, only 5% to 10% of adults without children are colonized, although in some settings and populations it has been found to be much higher. For example, colonization in military service personnel has been estimated to range between 50% and 60%.[56]

While many surveillance studies rely on PCR using upper respiratory tract samples for the detection of *S. pneumoniae*, a definitive diagnosis of infection with *S. pneumoniae* generally relies on isolation of the organism from blood or other normally sterile body sites (e.g., CSF, middle ear fluid, joint fluid, and peritoneal fluid).[56] A urinary antigen test is available to detect the C-polysaccharide antigen of *S. pneumoniae* as a cause of community-acquired pneumonia among adults (but not children).[56] The test is rapid and has the ability to detect pneumococcal pneumonia after antibiotic therapy has been started.[56]

Associations of Pathogen Load with Disease

The challenge of differentiating between pathogen carriage in the nasopharynx and infection of the lower respiratory tract speaks to a fundamental problem in pneumonia etiology research—the difficulty of sampling the lung and reliance instead on upper respiratory tract samples as a proxy for the site of infection.[45] For diagnostic purposes, bronchoalveolar lavage has high sensitivity for detection of bacteria and viruses in the lower respiratory tract but it requires bronchoscopy, which is an invasive procedure.[58] For surveillance and research purposes, nasopharyngeal (NP) and oropharyngeal (OP) swabs are the most logistically feasible specimen to collect, but they have low

specificity due to the high prevalence of bacterial colonization of the nasopharynx. While case-control studies can demonstrate which pathogens are more strongly associated with disease, it is difficult to ascribe etiology to pathogens that are known to colonize the upper respiratory tract because they are frequently detected by PCR in upper respiratory tract specimens of both cases and controls.[9]

An approach proposed for differentiating pathogen carriage from clinically significant infection is the quantification of pathogen load.[8,9] Real-time PCR (also called quantitative PCR—qPCR) can be used to quantify pathogen load.[59] In real-time PCR, a fluorescent marker is used to monitor the rate of generation of amplification target. After each amplification cycle, the intensity of the fluorescent signal reflects the quantity of DNA amplicons in the sample at that specific time. The point at which the fluorescence intensity increases above the threshold or background level corresponds proportionally to the initial number of template DNA molecules in the sample. This point is called the quantification cycle (Cq). The Cq value can be used to determine the absolute quantity of target DNA in the sample by constructing a calibration curve from serially diluted standard samples with known concentrations or copy numbers. In the absence of serially diluted samples, Cq values can be considered semi-quantitative measures of the amount of pathogen in a clinical specimen. Cq levels are inversely proportional to the amount of nucleic acid in a sample, with etiological relevance assumed when Cq values are low, suggesting high pathogen loads.[60]

Higher pathogen load in the upper respiratory tract has been associated with pneumonia, and for some pathogens, it has been associated with more severe outcomes.[10-16] However, there is still limited information on the association between colonization density of the upper respiratory tract and pneumonia, and no definitive thresholds in pathogen density identify a pathogen as causing a given case of pneumonia.[9] Findings from previous studies on the association between colonization

density and respiratory disease differ across populations and pathogens.[10-12,61-69] In pediatric populations, the PERCH study evaluated differences in bacterial colonization densities between pneumonia cases and controls, and found evidence that colonization densities of *Streptococcus pneumoniae* and *Haemophilus influenzae* are associated with microbiologically-confirmed pathogen-specific pneumonia.[10,70] The researchers did not find meaningful differences between pneumonia cases and community controls in nasopharyngeal colonization densities of *M. catarrhalis*, *S. aureus*, or *P. jirovecii*.[70] In adults, a study in Kenya found an association between Ct values for influenza A virus and respiratory illness, but did not find a difference in Ct values of other respiratory viruses among respiratory patients and asymptomatic controls.[12] Other pathogens for which colonization density has been linked to disease status or severity include RSV, parainfluenza virus 2, and human rhinovirus.[13-16] For *Streptococcus pneumoniae,* the association between bacterial density in the nasopharynx and pneumonia is fairly consistent across studies.[10,11,61-63,66] The focus of most studies has been the identification of bacterial load cutoffs that could be useful for diagnostic purposes, and the consensus is that Cq values are not useful for diagnostic purposes in individual patients. To our knowledge, no studies have used the association between bacterial load and disease to understand the etiologic role of S. *pneumoniae* in respiratory infections at the population level.

In the absence of assays to determine pathogen density in the upper respiratory tract directly, examining the association between Cq values (a proxy for pathogen density) and case-control status could help to distinguish clinically relevant infection from colonization, thereby improving our understanding of the etiology of respiratory infections. This approach has been applied in studies of diarrheal disease etiology, where the high prevalence of pathogen carriage also poses a methodological challenge.[17-22] Since 2017, re-analyses of two major studies of the etiology and burden of diarrheal infections in children have used PCR Ct values to estimate population

attributable fractions adjusted for the high prevalence of asymptomatic pathogen carriage.[17,23] A similar approach could contribute to our understanding of the etiology of respiratory infections.

Respiratory Syncytial Virus

Disease burden

RSV is the major viral respiratory tract infection of early infancy and the most common cause of hospitalization in infants globally.[6] Infection occurs with the greatest frequency during the first 2 years of life but infection during the first month of life is rare, presumably due to maternal immunity.[71] Exposure to RSV does not lead to long-lasting protection so people can have many infections over their lifetimes, although subsequent infections tend to be less severe. Infection mainly leads to mild disease, and mortality in children admitted to hospitals in low- and middle-income countries with RSV-associated acute lower respiratory infections (ALRI) is low (1-3%).[3] However, in very young children, older adults, and immunocompromised patients it can result in serious disease or death.[3] In infants, it is the major causative agent of bronchiolitis but it also causes pneumonia, croup, bronchitis, otitis media, and febrile upper respiratory tract illness.[71] Moreover, it can lead to persisting abnormalities in gas exchange and to wheezing; about half of children admitted to hospital with RSV disease experience later episodes of wheezing.[3]

Globally, RSV has been estimated to cause about 34 million episodes of acute lower respiratory infections in children under 5 years of age each year, with over 3 million severe enough to cause hospitalization.[72] In terms of mortality, RSV is associated with an estimated 66,000–199,000 deaths in children under 5 years old, the vast majority in low- and middle-income countries.[72] Data on the burden of RSV-associated ARI among older adults in low- and middle-income countries is limited. Based on available data, RSV-associated ARI causes between 4,800 and 50,500 in-hospital deaths and between 186,000 and 614,000 hospitalizations in older adults aged ≥65 years.[73]

In 2015, the WHO Product Development for Vaccines Advisory Committee highlighted RSV as a pathogen for which there is major vaccine pipeline activity and likelihood of a candidate emerging for licensure in the near term.[74] However, a number of obstacles have impeded vaccine development.[75] In the 1960s, a formalin-inactivated RSV vaccine induced an exaggerated clinical response to wild-type RSV infections in infants who were RSV naïve before vaccination. In one study, nearly 80% of vaccine recipients were hospitalized with lower respiratory illness compared to 5% of controls.[75] Since then, considerable progress has been made in the development of RSV subunit and nanoparticle vaccines.[75] Although the burden of RSV is greatest in infants, the primary targets of these vaccines are older adults and pregnant women. The latter are targeted for vaccination because very young infants may not respond adequately to vaccination due to immunologic immaturity and because maternal antibodies may interfere with the immune response to vaccination.[75]

Although the exact characteristics of a future maternal or infant vaccine are unknown, seasonality might influence the deployment of these vaccines, as with seasonal influenza vaccine.[76] Evidence from previous studies in temperate regions suggests that birth month is associated with the risk of RSV-associated hospitalizations, with infants born just before the start of the RSV season having the greatest risk of RSV-associated hospitalization.[77] Thus, infants born just before or during the RSV season could potentially benefit from maternal RSV immunization, while infants born outside this period are unlikely to do so.[33] In temperate regions, RSV infection tends to occur in predictable annual epidemics, while in tropical regions RSV seasonality tends to be more variable and more prolonged. Thus, a year-round maternal immunization strategy might be appropriate in tropical regions. However, in temperate regions, the optimal maternal immunization strategy might target infants due to be born just before the RSV season, depending on the duration of immune protection.[33] This highlights the importance of understanding and characterizing the temporal and

geographic patterns of RSV circulation at the local, regional, and national levels to target vaccination strategies once a vaccine is approved.[78] In 2017, WHO's *RSV Vaccine Research and Development Technology Roadmap* identified the description of local seasonality patterns as a priority activity to inform the seasonal versus aged-based vaccination strategies.[35]

Global RSV seasonality

RSV seasonality is largely dependent on geographic location and climate. In temperate regions, RSV activity tends to occur in the coldest months and the season duration is about 5-6 months. In contrast, in subtropical and tropical climates, peak activity typically occurs during the warmest months and RSV seasons last longer, up to 10 months.[79,80] However, seasonality can be variable within countries, particularly those spanning large geographical areas.[81] The start, end, and/or peak of RSV activity usually differs by 1-3 weeks from season to season but there are exceptions to this general pattern in seasonality. For example, in Finland, RSV epidemics follow a 2-year cycle with a small epidemic in the spring in one season followed by a major epidemic that starts in November-December and extends into spring. In Mexico, a 2-season year is followed by a milder year, where the outbreak starts in spring and activity is maintained almost all year round with no clear peaks.[80] In Germany, two differential patterns of RSV seasonality have been detected: an early season starting in October-November and finishing in March–April and a late season starting in December and finishing in May, with both seasons having similar duration.

A recent global overview of RSV seasonality found that most countries with available data on RSV have consistent seasonal patterns.[80] However, seasonal parameters have changed over time in various places and most countries also showed major variations of 1 month at least once during the period studied.[80] For example, an analysis of RSV data in England between 1981 and 2004 showed that the RSV season shortened and ended earlier over time.[82] In Sao Paulo, Brazil, RSV epidemics

have shifted to earlier onsets.[83] Numerous explanations for these seasonal patterns have been proposed, including the possibility that inclement weather modifies human behavior, increasing indoor crowding that enhances exposure to and transmission of RSV, or that low temperatures and absolute humidity increase the risk of disease. Regardless of the cause, identifying seasonal patterns is important when planning prevention strategies such as vaccination.[33] As RSV seasons can vary substantially by year and location, local RSV data are needed to accurately define the onset and offset of RSV seasons to inform the timing of local prevention measures.

LRI burden and RSV seasonality in Guatemala

According to the Global Burden of Disease Study 2017, Guatemala has the highest modeled LRI incidence in children under 5 years of age in the world.[42] However, estimates vary substantially across disease burden modeling studies depending on the methodology and data sources used.[84] A systematic analysis of national, regional, and global pneumonia morbidity and mortality in children under 5 years ranked Guatemala as having one of the lowest pneumonia incidence rates among low- and middle-income countries in 2015.[85] An estimate of childhood pneumonia incidence in 2010 ranked Guatemala's incidence at about the median for low- and middle-income counties.[86] The methods used to produce these estimates incorporate data on the prevalence of known risk factors for childhood pneumonia.

Few Central American countries have published RSV-related research.[87-91] A recent global overview of RSV seasonality included two years (2015-2017) of data from Guatemala. This analysis defined the onset of the RSV season as the first two consecutive weeks when at least 10% of total tested samples for respiratory pathogens were positive for RSV. The season offset was defined similarly as the time when RSV positivity was less than 10% for two consecutive weeks. During this two-year period, the RSV season started between weeks 16-23 (April/May), peaked between weeks 30-34

(July/August), and ended between weeks 43-47 (October/November). The epidemic period was between 25-31 weeks with no evidence of regional variability.[80] As RSV seasons can vary both within countries and over time, a longer period of observation is needed to accurately characterize national RSV seasonality.

Since 2007, Guatemala has conducted surveillance for hospitalized ARI under a collaboration between the CDC's International Emerging Infections Program, the Guatemala Ministry of Public Health and Welfare, and the Universidad del Valle de Guatemala. Epidemiologic analyses of data collected under this surveillance system between 2007 and 2012 described the epidemiology of RSV infection as well as trends in ARI.[92,93] Based on visual inspection of epidemiologic curves during this period, the peak of the RSV season occurred between July and November in most years. While visual inspection of past data to define epidemic thresholds (the level of virus activity that signals the start and end of the annual epidemic season) is easy to implement and understand, it can be overly simplistic and does not capture trend changes over time.[94] Since these analyses were published, RSV vaccine development has rapidly progressed and other methods for characterizing respiratory virus epidemics have been developed.[76,94,95] As shifts in RSV seasonality have occurred in other locations, it is timely to characterize RSV seasonality using more recent surveillance data.

Methods for characterizing seasonality

There is no standard method for characterizing respiratory virus seasonality. Averaging methods have been used to determine epidemic thresholds for influenza that divide the season into pre-epidemic, epidemic, and post-epidemic periods.[94] These methods involve calculating the pre- and post-epidemic rates for historical seasons. Examples include a WHO method proposed in 2012 and the moving epidemic method (MEM), which has been used to define influenza epidemic thresholds in Europe.[94,96] The WHO method determines an average epidemic curve by aligning data from past

seasons around their peaks to identify the average amplitude of a peak. The WHO epidemic threshold is defined as the annual median amplitude of the plotted data. The MEM calculates the epidemic threshold as the one-tailed 95% confidence interval of the arithmetic mean of a subset of the highest 30 pre-epidemic weekly values of past seasons. The MEM method has recently been used to describe the seasonality of RSV in the Netherlands and in Slovenia.[97,98]

In the United States, RSV seasonality used to be defined on the basis of weeks during which antigen-based tests detect RSV in >10% of specimens. However, with the increase in PCR tests for RSV, in 2017, the United States switched to the retrospective slope 10 (RS10) method to define RSV seasonality.[76] The method is based on normalized RSV detections and defines the season onset as the second of two consecutive weeks when the slope of the epidemic curve exceeds 10 normalized detections per week, provided that the slope exceeds 10 from that week forward. This method was chosen because it consistently captures 96-98% of annual detections nationally within a season.

Recent studies of global respiratory virus seasonality (including RSV) used an "average annual percentage" method to determine seasonality.[34,99] This method calculates the average annual percentage of cases for each week, then sorts weeks in descending order, and classifies as epidemic weeks the first weeks to add up to at least 75% of the annual average percentage. The start and end of the epidemic are then identified as the first and last week of the longest consecutive epidemic weeks with a 2-week gap allowed.

Household air pollution (HAP)

According to the most recent estimates from the Global Burden of Disease Study published in 2020, approximately 2.3 million deaths each year can be attributed to household air pollution.[41,100] The burden of disease from HAP exposure is particularly high in young children. HAP nearly doubles

the risk of pneumonia in children and is responsible for close to half of all pneumonia deaths in children under 5 years of age.[100] It also increases the risk of acute lower respiratory tract infections in adults, and contributes to 28% of all adult pneumonia deaths.[100]

In nearly all populations worldwide, people spend the majority of their time indoors.[101] Exposure to high levels of household air pollution is widespread in low- and middle-income countries, where up to 90% of rural households burn coal or biomass fuel (wood, dung, crop waste) on traditional stoves to meet their basic energy needs.[3,100,102] Globally, it is estimated that 2.6 billion people rely on solid fuels for cooking and heating.[103] Burning biomass fuels in open fires or poorly functioning stoves is a major source of exposure to toxic particles such as fine particulate matter, carbon monoxide, nitrogen dioxide, and polycyclic aromatic hydrocarbons.[104]

Air pollution increases the risk of childhood pneumonia through several mechanisms.[3] Smoke paralyzes the cilia that normally clear pathogens, causing the upper airways to become lined with a thick layer of mucus in which bacteria grow readily. Particulate matter in smoke penetrates deep into the lungs and is taken up by alveolar macrophages, which are then less efficient in phagocytosing and killing bacteria.[105] When the number of bacteria in the lungs is large or when the immune system is not functioning well, bacterial multiplication surpasses the body's defenses, and the child develops pneumonia.[3] Often it is the combination of air pollution injury to the lung followed by a viral infection that permits this bacterial invasion. In pregnant women, smoke can also interfere with blood flow to the placenta, leading to smaller infants with smaller airways that put them at increased risk of pneumonia during infancy.[3]

Evidence for the association between air pollution and ARI largely comes from observational studies, which have varied considerably in terms of design and quality.[27] To date, nine review articles on the subject have reported that HAP is an important risk factor for childhood pneumonia.[27,106-113]

A systematic review published 2008 reported a pooled odds ratio estimate of 1.79 (95% CI: 1.46, 2.21) for the association of high versus low indoor air pollution exposure with acute lower respiratory infections. A 2013 review (by some of the same authors as the 2008 review) used survey data on the *self-reported use of solid fuel* for cooking as the main indicator of household air pollution exposure to estimate the intervention effect of reducing household air pollution.[107] Using data from 26 studies, the authors of the 2013 review estimated a pooled odds ratio of 1.73 (1.47, 2.03) for the association between child acute lower respiratory infections (all severities) and HAP exposure (according to varying exposure definitions).[107]

The dependence on self-report (rather than on direct exposure measurements) to understand how HAP exposure relates to health outcomes is an inherent limitation of previous research studies included in early reviews. Confounding by poverty is of particular concern in observational studies as fuel use is highly correlated with poverty, which is a strong determinant of ARI (including pneumonia). More recently, the technology to measure HAP has improved, enabling researchers to physically measure pollutant levels rather than relying on self-reported biomass fuel use as a proxy. A recent 2019 review confirmed the positive association between exposure to solid fuel use and childhood pneumonia, but found that in the few studies where individual pollutants (i.e. carbon monoxide and $PM_{2.5}$) were physically measured, there was no evidence of an association.[106]

In contrast to findings from observational studies and despite biologic plausibility for the effect of household air pollution on respiratory health, RCTs of clean cookstove interventions have not definitively demonstrated an effect on respiratory outcomes. Although one of the first randomized trials of a clean cook stove intervention (RESPIRE) estimated a 22% reduction in physician-diagnosed pneumonia incidence, this effect was not statistically significant.[114] A subsequent intervention in Malawi, the Cooking And Pneumonia Study (CAPS), showed no effect of a cleaner

burning biomass-fueled cookstove intervention on WHO Integrated Management of Childhood Illness–defined pneumonia incidence.[115,116] Similarly, the Ghana Randomized Air Pollution and Health Study (GRAPHS) three-arm household-level randomized controlled trial of LPG versus a cleaner-burning biomass-fueled cookstove compared with control conditions (traditional stove), saw no effect on the incidence of child pneumonia.[117]

These findings from RCTs might indicate that HAP is not as harmful as previously thought or that HAP reduction interventions may not be sufficient to yield beneficial health effects in the context of high ambient air pollution exposure. However, limitations of these trials might also explain their null findings. One such limitation is that previous interventions may not have achieved sufficient exposure reductions to yield health benefits, either because cleaner-burning cookstoves are not clean enough or because intervention households continued to use traditional stoves alongside the interventions, a practice known as stacking. Although the chimney stove intervention used in the RESPIRE trial reduced carbon monoxide exposures by 50% on average, the exposure distributions of the intervention and control arms overlapped substantially.[114] The CAPS intervention had no effect on carbon monoxide exposure, which was low in the study population overall.[116] The improved biomass stove used in GRAPHS did not meaningfully reduce CO or PM2.5 exposure compared to three-stone fires. While the GRAPHS LPG intervention lowered carbon monoxide exposure significantly, post-intervention exposures still exceeded health-relevant targets.[118]

Another limitation of these trials is that they used non-specific case definitions for pneumonia.[28] There is some suggestion that exposure to HAP has an effect on severe outcomes rather than on milder respiratory infections.[28,114] Bruce et al.'s 2013 review reported pooled effect estimates for the association of HAP with all ALRI (21 studies), severe ALRI (4 studies), and fatal ALRI (4 studies); the magnitude of effect estimates increased with the severity of ALRI.[107] RESPIRE did not find a

statistically significant reduction in fieldworker- or physician-assessed WHO-defined pneumonia but did find a statistically significant reduction in WHO-defined severe pneumonia. The primary outcome in CAPS was WHO Integrated Management of Childhood Illness (IMCI)-defined pneumonia in children under 5 years of age. Although secondary outcomes included severe IMCI-defined pneumonia and severe pneumonia with oxygen saturation <90%, only a small number of participants met these case definitions. In general, case definitions with low specificity lead to non-differential outcome misclassification, and thereby an expectation of bias towards the null. Using more specific pneumonia case definitions designed to capture severe disease could improve the ability of studies to detect an effect of HAP on pneumonia.

The recent Household Air Pollution Intervention Network (HAPIN) trial has overcome limitations of previous trials by using a highly specific case definition capturing severe pneumonia and by achieving a strong exposure contrast. HAPIN randomized households with pregnant women in four countries to receive an LPG stove and fuel during pregnancy and until the child reached 1 year of age. The study measured personal exposures to carbon monoxide, black carbon, and $PM_{2.5}$ on a quarterly basis during gestation and the child's first year of life. The main analysis assesses the intervention's effect on severe pneumonia in children through 1 year of life, the period of highest incidence of pneumonia in children. However, evidence suggests that exposure experienced during gestation and early life is linked to a range of longer-term outcomes.[119] Although the mechanisms involved in the long-term effect of gestational HAP exposure on pneumonia risk have not been fully examined, there are critical periods of vulnerability to numerous adverse outcomes in early development. Fetal lung development occurs across gestation and $PM_{2.5}$ can cross the placenta and may disrupt biological mechanisms that regulate fetal growth, maturation and development.[119-122] Nevertheless, relatively few studies have examined the effect of gestational exposure to pollution on infant lung function and early life respiratory infections.[123-125] Findings of GRAPHS showed that

prenatal HAP exposure impairs lung function in infants potentially increasing their risk of pneumonia and severe pneumonia in the first year of life.[125] Other studies have demonstrated that infants born to mothers with increased particulate matter exposure during pregnancy have higher respiratory rates and altered breathing flows.[126-128]

To assess the longer-term health impacts of HAP exposure, three of the original four HAPIN study sites are participating in a cohort study to follow children through age 2 years. The study will not assess severe pneumonia outcomes at 2 years, but data on care-giver reported illness with a cough is being collected. Although illness with a cough is a non-specific outcome, which has been posited as a contributor to null findings in previous trials, we might expect to observe an effect of the intervention on this outcome because of the strong exposure contrast achieved by the HAPIN intervention. Notably, a recent meta-analysis that included 23 articles (including 7 RCTs) reported a relative risk of 0.59 (95% CI: 0.45, 0.77) for the non-specific outcome of ARI in people using improved cookstoves compared with traditional stoves.[129]

ARIs, which include both upper and LRIs, exert a considerable burden and cost on health care systems.[1] Upper respiratory infections alone account for some 17.2 billion illnesses each year.[1] Generating up-to-date estimates of the association between biomass fuel exposure and ARI is important for estimating the global burden of respiratory infections. Global disease burden estimates often use models based on the prevalence and effect sizes of risk factors for childhood pneumonia.[86,130] Exposure to HAP or solid fuel use for cooking are risk factors used in these models. The prevalence and distribution of ARI and HAP/solid fuel use (as determined through DHS) and the estimated association between them are used to parameterize these models. As the global epidemiology of respiratory infections continues to evolve, it is important to generate up-to-date estimates of the association between HAP/solid fuel use and ARI.

Chapter 3. Respiratory syncytial virus seasonality in Guatemala, 2008-2018

Abstract

Background. Local respiratory syncytial virus (RSV) seasonality can inform prevention strategies including vaccination. WHO has identified the description of local seasonality patterns in RSV incidence as a priority research activity. We characterized RSV seasonality in Guatemala using the moving epidemic method (MEM).

Methods. We used absolute counts of RSV-associated acute respiratory infections (ARI) from hospital surveillance in Santa Rosa and Quetzaltenango departments of Guatemala. We identified attributes of RSV seasons including the onset week, offset week, epidemic duration, and epidemic threshold–the level of virus activity that signals the onset of a seasonal epidemic.

Results. From week 17 of 2009 through week 16 of 2018, 8,222 ARI cases tested positive for RSV by rRT-PCR. Season onsets varied up to 5 months such that two differential patterns of RSV seasonality were observed: an early season starting in June-July and finishing in September-November, and a late season starting in October-November and finishing in March-April, with both seasons having similar durations ranging from 4 to 6 months. MEM epidemic thresholds calculated prospectively using previous seasons' data ranged from 2.1 to 4.0 RSV-associated ARI cases/week and captured between 70% and 96% of annual RSV detections. Seasonal patterns diverged between surveillance sites; onset weeks differed by 2 to 10 weeks each season and offset weeks differed by 2 to 16 weeks.

Conclusion. Our results demonstrate subnational differences in seasonality and substantial variability in the timing of seasonal RSV epidemics in Guatemala. This variability speaks to the difficulty in precisely predicting the timing of seasonal RSV epidemics based on onset weeks from

past seasons, and suggests that maximal reduction in RSV disease burden would be achieved through a year-round vaccination program.

Introduction

Respiratory syncytial virus (RSV) is the major viral respiratory tract infection of early infancy and the most common cause of hospitalizations in infants globally.[6] It has been estimated to cause about 34 million episodes of acute lower respiratory tract infections in young children each year, with over 3 million severe enough to cause hospitalization.[72] In terms of mortality, RSV is associated with an estimated 66,000–199,000 deaths in children under 5 years old, the vast majority in developing countries.[72]

As a result, the World Health Organization (WHO) has recognized RSV as the most important future new vaccine target.[31] In 2015, the WHO Product Development for Vaccines Advisory Committee highlighted RSV as a pathogen for which there is major vaccine pipeline activity and likelihood of a candidate emerging for licensure in the near term.[74] Although the exact characteristics of future RSV vaccines are unknown, seasonality—i.e., cyclical patterns in infection incidence--might influence the deployment of these vaccines, as is the case with seasonal influenza vaccine.[76]

To identify optimal vaccination strategies and provide a baseline to assess possible future vaccine effects, it is important to characterize RSV seasonality. For example, potentially effective maternal vaccines against RSV would provide passive immunity of limited duration to infants (maternal antibodies wane by about 6 months) so understanding RSV seasonality is important to ensure that maternal vaccines are administered during periods that will provide protection to infants during the RSV season.[33] Global reviews have served as a guide to RSV seasonality.[34] However, seasonal patterns are often variable within as well as between countries so characterizing local seasonality patterns is needed to inform effective national vaccine strategies and the timing of prevention

measures.[99] Thus, WHO has identified the description of local seasonality patterns in RSV incidence as a priority research activity.[35]

RSV seasonality is correlated with geographic location and climate. In temperate regions, RSV activity tends to occur in colder, drier months and seasonal epidemics last about 5-6 months.[80] Numerous explanations for these patterns have been proposed, including the possibility that inclement weather modifies human behavior, increasing indoor crowding that enhances exposure and transmission of RSV, or that low temperatures and absolute humidity increase the risk of disease.[131-133] In contrast, in subtropical and tropical climates, peak activity typically occurs during the warmest months and RSV seasons last longer, up to 10 months.[79,80] A global overview of RSV seasonality found that most countries with available data on RSV have consistent seasonal patterns, but seasonal parameters can change over time.[80] While the specific drivers of RSV seasonality remain unclear, identifying seasonal patterns and defining epidemic thresholds (the level of RSV activity that signifies the onset of a seasonal epidemic) are important when planning prevention strategies such as vaccination.[33]

Few Central American countries have published research on RSV seasonality, and previous studies relied on only a few years of data, making it difficult to detect long-term trends. Analyses of hospitalized acute respiratory infections in Guatemala between 2007 and 2012 described temporal patterns of RSV infections using visual inspection of epidemiologic curves.[92,93] A recent global overview of RSV seasonality included two years (2015-2017) of data from Guatemala. [80] However, as patterns in RSV seasonality can shift over time, it is important to characterize seasonality over many years to assess potential evolutions in seasonality.[80]

There is no standard approach to characterizing respiratory virus epidemics. However, for influenza, a WHO average curve method proposed in 2012 and the moving epidemic method (MEM) have

been used to determine epidemic and alert thresholds that respectively signal the onset and severity of seasonal influenza epidemics.[22,94,96] The purpose of these methods is to define a baseline based on several years of data and to establish an epidemic threshold above which weekly virus activity is considered to be in the epidemic period. While these methods were designed primarily for influenza epidemic and pandemic severity assessment, they can also be used with any data that have a seasonal accumulation of cases to characterize key aspects of seasonality including season onsets, offsets, and durations. The MEM has recently been used to describe the seasonality of RSV in the Netherlands and in Slovenia.[97,98]

The objective of this study was to characterize RSV seasonality in Guatemala from 2008 to 2018 using the MEM and WHO methods.

Methods

Study setting and population

The main data source for this analysis is *Vigilancia Integrada Comunitaria* (VICo), an integrated infectious disease surveillance system in Guatemala established through a collaboration between the US Centers for Disease Control and Prevention International Emerging Infections Program, the Guatemala Ministry of Public Health and Welfare, and the Universidad del Valle de Guatemala. The surveillance system has been described previously.[92] Briefly, it was established in November 2007 in Santa Rosa Department and subsequently expanded to Quetzaltenango Department in February 2009. Among other syndromes, VICo includes surveillance of hospitalized acute respiratory tract infections (ARI). We analyzed data on ARI from surveillance hospitals in Santa Rosa and Quetzaltenango, which consistently reported cases from February 2009 through June 2018.

Surveillance for hospitalized ARI in Santa Rosa (coastal lowlands, with temperatures typically ranging from 15°C to 30°C[134]) was conducted at the department's only hospital, the Cuilapa

National Hospital, a 176-bed regional referral government hospital. In Quetzaltenango (western highlands, with temperatures ranging from 6°C to 21°C[134]), surveillance was at Hospital Regional de Occidente, a 435-bed hospital. Both facilities include pediatric and adult intensive care units. While the age distributions of the surveillance populations in the two departments are similar, demographic characteristics and health-seeking behaviors differ. The populations of Santa Rosa and Quetzaltenango are 46% and 62% urban respectively, and the populations have different ethnic compositions.[135] Healthcare utilization surveys carried out in Santa Rosa and Quetzaltenango in 2007 and 2009, respectively, found that among those reporting pneumonia in the past year, 33% of those aged <5 years and 75% of those aged ≥5 years in Santa Rosa were admitted to the surveillance hospital whereas in Quetzaltenango 75% of those aged <5 years and 50% of those aged ≥5 years were admitted to the surveillance hospital.[92,136]

Case definitions and study procedures

A case of ARI was defined as a hospitalization with at least one sign or symptom of respiratory disease and evidence of acute infection within the first 24 hours of admission (**Table 3.1**). Study nurses identified eligible patients by reviewing the ward registers and emergency department logs for patients presenting with respiratory diagnoses. Study staff sought consent and enrollment of hospitalized patients meeting the ARI case definition at participating hospitals. All participants were asked to provide nasopharyngeal and oropharyngeal swabs, which were tested by rRT-PCR for a respiratory panel that included RSV. Samples were processed within 72 hours of collection.

Study data and season attributes

To describe RSV seasonality in Guatemala, we used the absolute numbers of RSV detections from ARI surveillance in Santa Rosa and Quetzaltenango from week 17 of 2009 to week 16 of 2018, covering nine consecutive surveillance seasons (defined as the period between troughs in RSV activity) with consistent data collection. Using methods described below, we identified attributes of a

given RSV season, including the onset (start) week, peak week, time from onset to peak week, offset (end) week, epidemic duration, percentage of annual RSV detections that occur within the epidemic period, and the epidemic threshold (i.e. the number of RSV detections that marks the start of the seasonal epidemic) (**Fig 3.1**).

Moving epidemic method

We used the MEM to define attributes of RSV seasonality including an epidemic threshold. The MEM is described elsewhere and was applied using the 'mem' R package and MEM Shiny Web Application.[96,137] Briefly, the MEM divides each season separately into pre-epidemic, epidemic, and post-epidemic weeks. First, the optimum epidemic duration is estimated by drawing a smoothed curve of the weekly maximum accumulated percentage of total season RSV cases, starting with the epidemiologic week where the slope of the epidemic curve is greatest. The MEM includes several settings for determining the optimum season duration from the maximum accumulated percentage curve. We used the slope method, which estimates the optimum epidemic duration of a season as the number of weeks where the slope of the maximum accumulated percentage curve (i.e. the change in the cumulative weekly percentage of total season cases) is above the mean weekly slope. Once the optimum number of epidemic weeks in a given season is identified, epidemic weeks are classified as the consecutive weeks that include the greatest change in slope; pre-epidemic and post-epidemic weeks are those preceding and following the epidemic weeks, respectively.

After identifying the pre-epidemic, epidemic, and post-epidemic weeks in each season, the epidemic threshold is calculated using a set of pre-epidemic values. For each season, the highest n values from the pre-epidemic period are taken, where n=30/number of seasons. The epidemic threshold is the one-tailed 95% mean confidence interval of the arithmetic mean of this subset of the highest 30 pre-epidemic weekly values of all seasons.

It is important to note that the MEM threshold largely depends on the settings that are used. To select optimal settings, we assessed the performance of the MEM epidemic threshold for detecting epidemics by calculating its sensitivity, specificity and positive and negative predictive values using a cross-validation (leave-one season out) procedure in the MEM web application. For example, for a particular target season, the MEM algorithm determines pre-, post, and epidemic weeks as described above. This is considered the "correct" classification. With the remaining seasons, the epidemic threshold is calculated and weekly counts in the target season are classified as above or below the threshold. Weekly counts above the threshold are considered observed epidemic weeks and weeks below the threshold are considered observed non-epidemic weeks. Thus, each week has a correct and an observed outcome so that sensitivity, specificity, and positive and negative predictive values can be calculated. We chose settings that balanced sensitivity and specificity and that included a high percentage of all detections within the epidemic period. Using this cross-validation procedure, the MEM threshold correctly identified 91% of epidemic weeks and 79% of non-epidemic weeks and captured a high percentage (85%) of RSV detections within what was considered the RSV season according to the MEM algorithm.

Differences in RSV seasonality in Santa Rosa and Quetzaltenango were assessed by running the MEM algorithm on site-stratified data to calculate the season parameters described above. Data for the 2008-2009 season were available for Santa Rosa but not for Quetzaltenango. We also calculated Spearman's rank order correlation coefficient to test whether weekly RSV counts in the two sites were correlated.

WHO method

To explore whether a simpler averaging method would yield results similar to the MEM, we adapted a method for establishing average epidemic curves and epidemic thresholds for influenza that is

described in the *WHO Global Epidemiological Surveillance Standards for Influenza* (**Fig. 3.2**).[94] First, we drew smoothed weekly epidemic curves for the 2009/2012 to 2017/2018 seasons using a 4-week moving average comprised of the current weekly count and the three preceding weekly counts. Next, we aligned the peak weeks of all seasons on the median week of peak occurrence. The average epidemic curve was drawn by calculating the arithmetic mean for each week over all years. The epidemic threshold was defined as the annual median weekly number of cases for the average epidemic curve, which assumes that approximately half of the surveillance year is in-season and the other half is off-season. Based on this epidemic threshold, the onset and offset weeks of each season were identified as the first of two consecutive weeks above and below the epidemic threshold respectively.

Assessing threshold performance

To assess how well the MEM and WHO epidemic thresholds would perform prospectively, we calculated the proportion of annual RSV detections that would have occurred above thresholds calculated with previous seasons' data. For example, the 2012/13 epidemic threshold was calculated using data from 2009/10 through 2011/12, and the 2013/14 threshold was calculated using data from 2009/10 through 2012/13, and so on. We did not calculate thresholds for the 2009/10 to 2011/12 seasons as there were too few previous seasons of data available. We considered the onset and offset weeks of each season to be the second of two consecutive weeks above and below the epidemic threshold respectively. Assuming one RSV epidemic wave per surveillance season, we defined the epidemic period as the longest period of weekly counts above the epidemic threshold. False alerts were defined as weeks outside the epidemic period that had RSV counts above the epidemic threshold.

Assessing differences in seasonality by age

Variability of age distributions among RSV cases over the nine consecutive seasons was assessed visually using proportional stacked bar graphs and using a chi-square test for differences in age groups across seasons. Age was categorized into 6 groups: 0-6 months, 6-12 months, 1-2 years, >2-5 years, >5-65 years, and ≥65 years.

Results

From week 17 of 2009 through week 16 of 2018, 8,222 ARI cases were enrolled and tested for RSV at the Cuilapa National Hospital in Santa Rosa (n=3,883) and Western Regional Hospital in Quetzaltenango (n=4,339) (**Fig. 3.3**). At Cuilapa and Western Regional Hospitals, 1,043 (27%) and 1,278 (29%) ARI cases respectively tested positive for RSV by PCR. The median number of RSV detections per season was 223 (IQR: 198, 330). The majority (70%, n=1,618) of cases were in children less than 1 year of age. Adults aged ≥65 years comprised 2% (n=54) of cases.

RSV seasons varied substantially by year such that two differential patterns of seasonality were observed: an early season starting in May/June, and a late season starting in October/November. Epidemic onsets ranged from week 20 in the 2011/2012 season to week 44 in the 2012/2013 season (**Table 2**). We did not observe a consistent pattern of late seasons oscillating with early seasons or small seasons oscillating with larger ones. The median peak week of the epidemic occurred at week 33 (range: 28–9). Offset weeks ranged from week 38 in the 2011/2012 season to week 16 of 2013 in the 2012/2013 season. The epidemic period ranged from 17 to 24 weeks and included 70% to 95% of annual RSV detections.

The WHO method produced similar seasonal parameters to the MEM method. Median season onsets (week 24, MEM range: 20–44, WHO range: 17–44), offsets (week 46, range: 38–16), epidemic durations (21 weeks, MEM range: 17–24, WHO range: 13–29), and times from onset to peak weeks

(8 weeks, range: 5–17) were the same with both methods (**Table 3.2**). Both covered a substantial proportion of annual RSV cases (on average, 84% with MEM, vs. 87% with WHO method).

Using data from all surveillance seasons, the number of RSV detections that marks the start of the seasonal epidemic (epidemic threshold) was 4.1 cases/week under the MEM compared to 3.7 cases/week using the WHO method. MEM (WHO) epidemic thresholds calculated prospectively using previous seasons' data ranged from 2.1 (1.3) to 4.0 (3.3) RSV-associated ARI cases/week and captured between 70% (73%) and 96% (98%) of annual RSV detections (**Table 3.3**). The number of weeks in a given surveillance season that were above the epidemic threshold ranged from 15 (21) to 37 (38). The MEM thresholds produced a false positive alert in three seasons whereas the WHO thresholds did not produce any false positive alerts.

Although there was a strong positive relationship between weekly counts in the two sites (Spearman's rank order correlation coefficient for the weekly counts was 0.63, $p<0.01$), the seasons in Santa Rosa and Quetzaltenango diverged somewhat (**Fig. 3.4**). Across the seasons, onset weeks in the two sites differed by 2 to 10 weeks each season and offset weeks differed by 2 to 16 weeks (**Table 4**). In most seasons epidemic onsets occurred earlier (7 of 9 seasons) and ended earlier (6 of 9 seasons) in Quetzaltenango. However, there were exceptions to this pattern. For example, the 2013/14 season started, peaked, and ended about two months earlier in Santa Rosa than in Quetzaltenango.

Although inter-season differences in the proportions of RSV cases in each age group were statistically significant ($p<0.02$), the highest proportion of RSV cases was consistently in the ages <6 months (range: 45%–57%) and the lowest proportions of RSV cases were in the age groups ≥65 years (range: 1%–4%) and >2–5 years (range: 2%–4%). (**Fig. 3.5**) There was no correlation between the age distribution of RSV-associated ARI cases and the size or onset of seasonal epidemics.

Discussion

Our analysis of data obtained from prospective surveillance of hospitalized ARI over nine consecutive seasons provides a detailed overview of seasonal patterns in RSV infections in Guatemala. Our results demonstrate substantial variability in the timing of seasonal RSV epidemics. Season onsets varied by as much as 5 to 6 months such that two differential patterns of RSV seasonality were observed: an early season starting in May-July and finishing in September-November, and a late season starting in October-November and finishing in March-April, with both seasons having similar durations ranging from 4 to 6 months. This variability speaks to the difficulty in precisely predicting the timing of seasonal RSV epidemics based on onset weeks from past seasons and the value of epidemic thresholds as indicators of epidemic onset.

To be useful, thresholds should capture a high proportion of annual RSV detections in a short period of time. MEM epidemic thresholds calculated separately for each season from previous seasons' data gave an indication of how well the MEM epidemic threshold would perform prospectively. Thresholds calculated in this way performed well in terms of the proportion of annual RSV detections in the epidemic period. In each season, at least 70% of annual RSV detections occurred in weeks above the epidemic threshold. Typically, about 5 to 6 months were above the epidemic threshold but the epidemic period was particularly long (38 weeks) in the 2013/14 season because it was a large season preceded by seasons with much lower RSV activity. Because thresholds are calculated with previous seasons' data, seasons with above average RSV activity tend to have longer epidemic periods. The WHO epidemic thresholds were lower than the MEM thresholds so they produced longer epidemic periods that included higher proportions of annual RSV detections in some seasons. Unlike the MEM thresholds, the WHO thresholds did not produce any false alerts.

The choice of threshold depends on balancing the epidemic percentage and number of false alerts with shorter epidemic periods.

The WHO method produced similar seasonal parameters to the MEM during most seasons with the exception of the 2013/2014 season, which was the season with the greatest RSV activity. The general agreement between season parameters calculated using the MEM and WHO method supports the validity of the results but may also reflect that the two methods are based on similar assumptions and approaches. For example, both are averaging methods in that the WHO method determines the onset as the median number of weekly cases in the average epidemic curve, and the MEM settings we used determine the onset based on the mean slope of the epidemic curve. Other methods for estimating season parameters exist and there is no standard approach. Elsewhere, the different approaches used have resulted in varying and sometimes contradictory outcomes.[99]

We hypothesized that climatic differences between Santa Rosa (coastal lowlands) and Quetzaltenango (western highlands) might result in differing seasonal patterns. Elsewhere, including the United States, there is substantial subnational variation in RSV seasonality, with the RSV season starting in the South East and then moving to the North West.[99] Although seasonal patterns diverged to some extent between Santa Rosa and Quetzaltenango, there were no regular patterns in these differences. However, stratifying the analysis by site revealed important patterns that were masked in the pooled data. For example, the 2012/13 and 2013/14 seasons appear to be one prolonged season in the pooled data but stratification shows a distinct late epidemic followed by an early epidemic in Quetzaltenango that is masked by a delayed 2012/13 season that peaks in early 2013 into in Santa Rosa. As has been noted elsewhere, this demonstrates that national pooling of data may misrepresent RSV activity and the need for local data to precisely define RSV outbreaks in a given community.[138]

We had hypothesized that the age distribution of RSV cases might differ across seasons, with cases being older on average in seasons with later onsets. Elsewhere, delayed or off-season RSV epidemics have been associated with increased age of RSV cases due to the accumulation of RSV-naïve children, notably during the COVID-19 pandemic.[139] There was little inter-season variability in age distribution of RSV cases, and we did not find differences in the median age of children with RSV in delayed seasons and early seasons. This may be because delays in season onset were not of sufficient magnitude to result in shifted age distributions.

The occurrence of clear seasonal epidemics with a stable duration in Guatemala was consistent with RSV patterns in other areas.[34] Although epidemics in most countries are consistent over time with year-to-year variations of 1-4 weeks in the start, end, and/or peak of RSV activity, the multiyear periodicity we observed has been reported in a few countries.[80,140] For example, in Mexico, a 2-season year is followed by a milder year, where the outbreak starts in spring and activity is maintained almost all year round with no clear peaks.[80]

The mechanisms that shape these seasonal RSV patterns are unclear, but include contact rates between susceptible and infected individuals and host immunity. Although meteorological factors have been found to predict RSV incidence, the correlations between RSV incidence, temperature and relative humidity are particularly variable and inconsistent in tropical regions.[141] Moreover, mechanistic models have shown that undetectable seasonal changes in transmission can combine with population immunity to produce large oscillations in disease incidence.[142] Although there is limited research on the impact of RSV subtype on RSV seasonality, the predominant circulating antigenic group might also play a role in shaping seasonal patterns. In Finland, RSV antigenic groups A and B alternate in two-year cycles, and in Korea different genotypes dominate the circulation in consecutive epidemics.[143,144] A study in Beijing found that longer and earlier epidemics occurred during RSV A dominant seasons[145]

The observed variability in RSV seasonality has important implications for vaccine trials and eventual vaccine programs. If trials occur during seasons with relatively low RSV activity, they may not achieve adequate power to detect hypothesized effect sizes. Notably, a recent trial fell short of its projected number of RSV-associated, medically attended lower respiratory tract infections and failed to meet the prespecified criterion for success.[146] RSV vaccine trials should span several seasons to ensure that they achieve the targeted number of endpoints. In terms of an eventual vaccine strategy, the variability in seasonality in Guatemala suggests that maximal reduction in the burden of RSV disease would be best achieved through a year-round maternal RSV vaccination program.

Our study has several limitations and simplifying assumptions that bear noting. Although our analysis included nine seasons of consistently collected surveillance data – more than previous reports of RSV activity in Guatemala – the study period was not of optimal duration for assessing longer term trends and multiyear seasonality. We did not have data on RSV subtypes, which have been shown to correlate with different seasonality patterns in some settings.[145] Finally, we assumed that the population under surveillance remained stable over the study period; unknown violations of this assumption would have affected our estimates.

Despite these limitations, our study provided a detailed description of RSV seasonality in Guatemala that can guide the timing of eventual prevention strategies such as vaccination and immunoprophylaxis. Characterizing RSV seasonality is important for decision-making about vaccine timing and strategy, whereas setting epidemic thresholds that can be used prospectively to signal the start of a seasonal RSV epidemic can improve the accuracy of clinical diagnosis and the timely use of costly immunoprophylaxis. The inclusion of nine seasons of consistently collected surveillance data allowed for the identification of two differential patterns of seasonality, which was not possible in previous studies that covered shorter time periods. Ahead of the potential availability of new RSV

vaccines for pregnant women and young children, our findings can provide baseline information for immunization advisory groups to assess future RSV vaccine effects.

Tables and Figures

Table 3.1. Surveillance case definitions for hospitalized acute respiratory infection in Guatemala

Evidence of acute infection **All ages** ≥ 1 of the following:	**Signs or symptoms of respiratory disease**	
	All ages ≥ 1 of the following:	**Additional criteria for children <2 years** ≥ 1 of the following:
fever (≥38°C) or history of fever	tachypnea	repeated pauses in breathing while breastfeeding or drinking
hypothermia (<35.5°C)	cough	intercostal retractions
abnormal white blood cell count	expectoration (sputum production)	nasal flaring
abnormal white blood cell differential	chest pain	grunting
	hemoptysis	not drinking or breastfeeding
	dyspnea (difficulty breathing)	
	shortness of breath	
	sore throat	
	abnormal lung examination	

Notes. tachypnea was defined as: age <2 months: ≥ 60 breaths/minute, age 2-12 months: ≥ 50 breaths/ minute, age> 12 months-5 years: ≥ 40 breaths/ minute, age> 5 years: ≥ 20 breaths/minute). Abnormal white blood cell count (in children <5 years: <5,500 / cm^3 or> 15,000 / cm^3, in people ≥5 years: <3.000 / cm^3 or> 11,000 / cm^3). Any white blood cell differential abnormality as defined by the automated blood cell analyzer at each surveillance site.

Table 3.2. Characteristics of Respiratory Syncytial Virus (RSV) Seasons using RSV Counts from Hospital Surveillance of Acute Respiratory Infections in Guatemala, April 2009–April 2018

Surveillance season	peak week	onset week		offset week		epidemic duration, weeks		percentage of RSV detections in epidemic period		Time from onset to peak, weeks	
		WHO*	MEM‡	WHO	MEM	WHO	MEM	WHO	MEM	WHO	MEM
2009/2010	35	21	22	45	43	24	21	95%	90%	14	13
2010/2011	41	33	31	49	48	16	17	82%	84%	8	10
2011/2012	30	25	20	38	38	13	18	80%	90%	5	10
2012/2013	9	44	44	16	16	24	24	92%	92%	17	17
2013/2014	28	17	23	46	44	29	21	87%	74%	11	5
2014/2015	48	43	43	5	11	15	21	70%	83%	5	5
2015/2016	29	24	24	45	46	21	22	75%	76%	5	5
2016/2017	33	20	25	47	47	27	22	92%	83%	13	8
2017/2018	30	24	24	45	45	21	21	95%	95%	6	6
Median	33	24	24	46	46	21	21	87%	84%	8	8

Notes. *WHO: WHO average curve method (adapted from WHO Global Epidemiological Surveillance Standards for Influenza) ǂMEM: moving epidemic method

Table 3.3. Epidemic thresholds for Respiratory Syncytial Virus, Guatemala, April 2012–April 2018

Target season	Epidemic threshold, RSV detections/week		Percentage of RSV detections above threshold		Weeks above epidemic threshold, no.		False alerts, no.	
	WHO*	MEM‡	WHO	MEM	WHO	MEM	WHO	MEM
2012/2013	1.25	2.16	98%	90%	31	24	0	1
2013/2014	1.94	2.36	96%	96%	38	37	0	0
2014/2015	3.06	3.51	82%	70%	21	15	0	1
2015/2016	3.10	3.47	73%	73%	22	22	0	1
2016/2017	3.31	3.78	92%	92%	28	28	0	0
2017/2018	3.31	3.98	93%	93%	22	22	0	0

Notes. *WHO: WHO average curve method (adapted from WHO Global Epidemiological Surveillance Standards for Influenza) ‡MEM: moving epidemic method. Epidemic thresholds for each target season were calculated using data from previous seasons, starting with the 2009/2010 season.

Table 3.4. Characteristics of Seasons of Respiratory Syncytial Virus, by Surveillance Site, using the Moving Epidemic Method, Guatemala, April 2008–April 2018

	onset week		peak week		offset week		epidemic duration, weeks		percentage of RSV detections occurring within epidemic period		Time from onset to peak, weeks	
	SR	QU	SR	QU	SR	QU	SR	QU	SR	QU	SR	QU
2008/2009	33	-	42	-	50	-	17	-	72%	-	9	-
2009/2010	23	21	30	35	42	44	19	23	93%	89%	7	14
2010/2011	34	31	40	41	50	48	16	17	90%	82%	6	10
2011/2012	24	17	30	27	40	37	16	20	93%	91%	6	10
2012/2013	51	42	11	9	16	13	17	23	69%	89%	12	19
2013/2014	19	28	28	45	36	52	17	24	85%	86%	9	17
2014/2015	42	40	49	48	7	11	18	24	88%	84%	7	8
2015/2016	20	24	29	39	36	46	16	22	76%	75%	9	15
2016/2017	27	17	34	25	46	43	19	26	94%	88%	7	8
2017/2018	27	24	36	27	45	42	18	18	93%	88%	9	3

Abbreviations. SR: Santa Rosa, QU: Quetzaltenango.

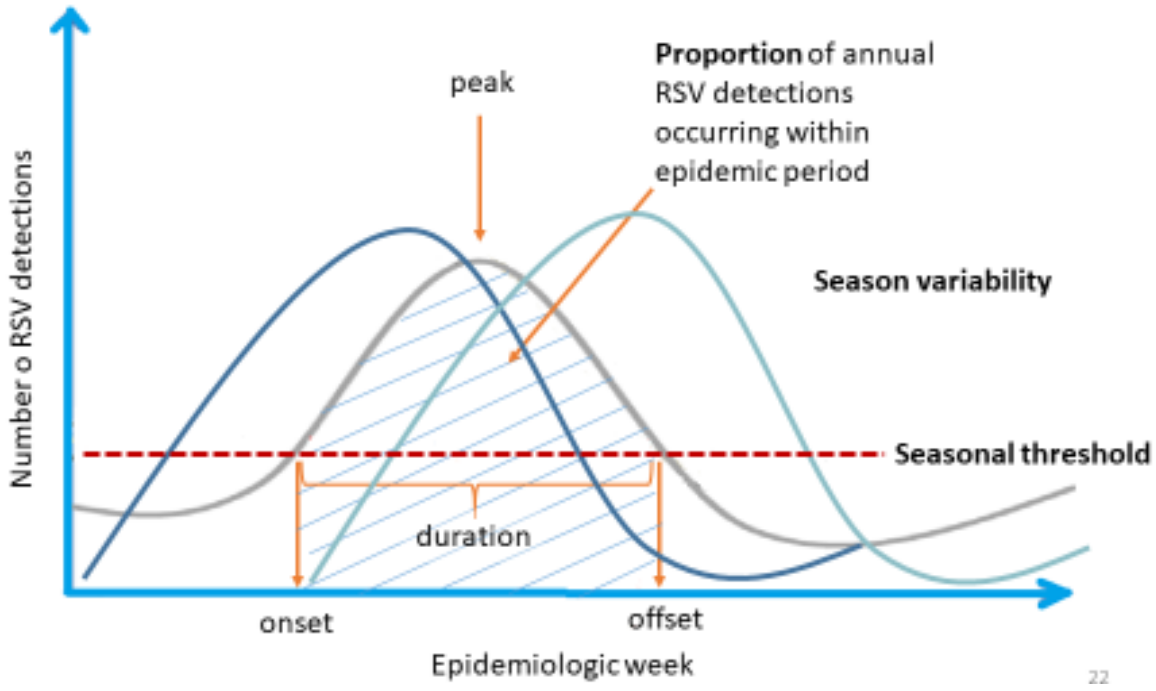
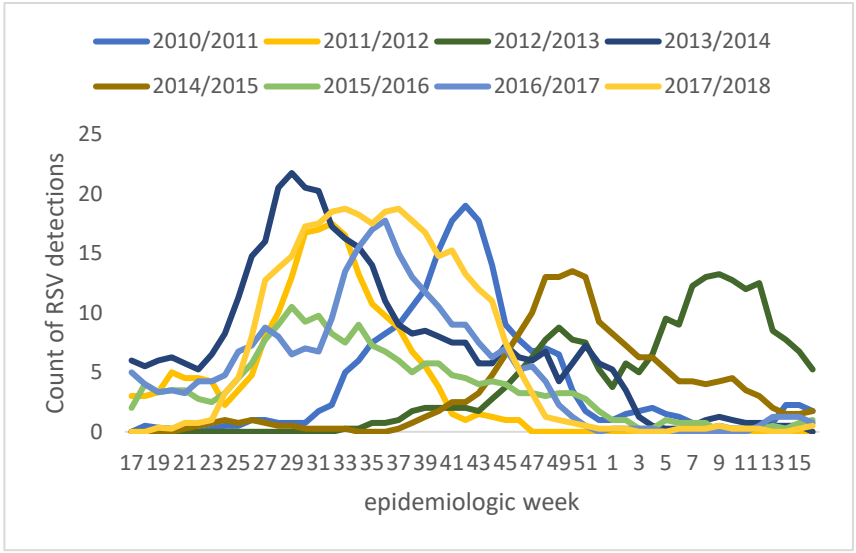
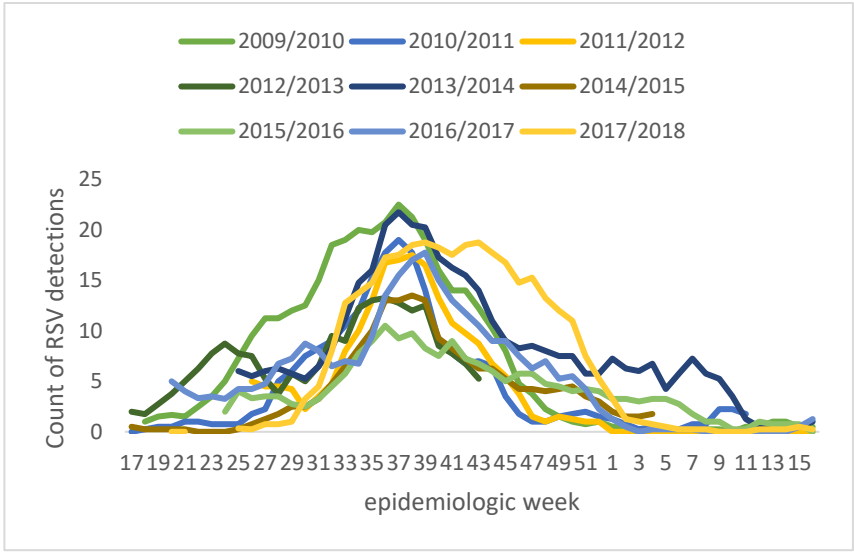


Figure 3.1. Illustration of season attributes

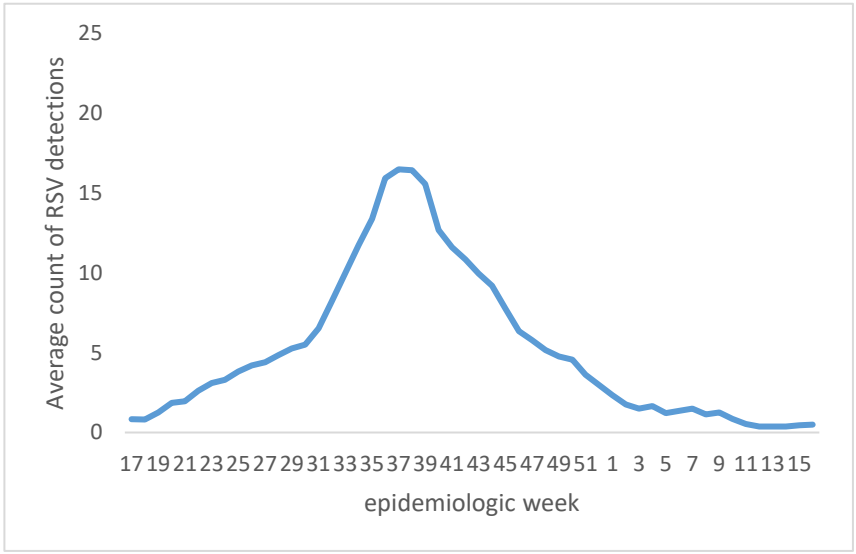
A


2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
2016/2017
2017/2018
Count of RSV detections
0
5
10
15
20
25
17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 1 3 5 7 9 11 13 15
epidemiologic week

B


2009/2010
2010/2011
2011/2012
2012/2013
2013/2014
2014/2015
2015/2016
2016/2017
2017/2018
Count of RSV detections
0
5
10
15
20
25
17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 1 3 5 7 9 11 13 15
epidemiologic week

C


Average count of RSV detections
0
5
10
15
20
25
17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 1 3 5 7 9 11 13 15
epidemiologic week

D

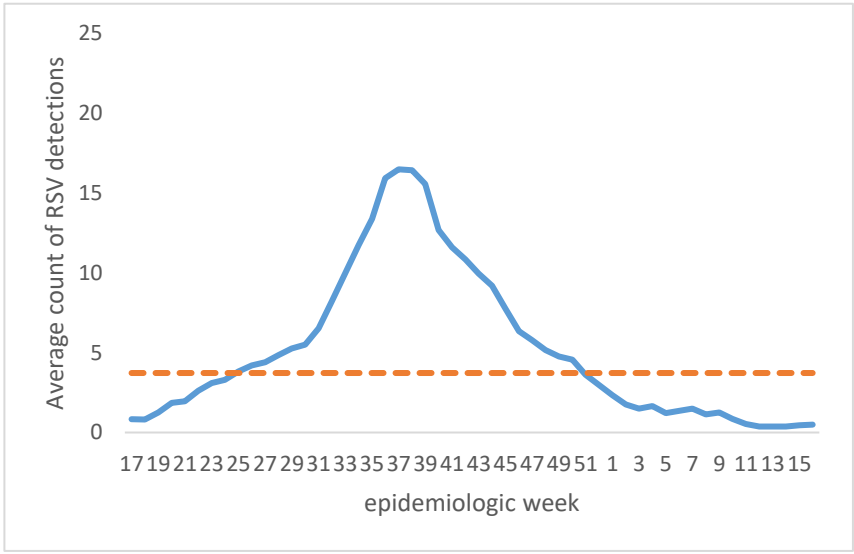

Average count of RSV detections
0
5
10
15
20
25
17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 1 3 5 7 9 11 13 15
epidemiologic week

Figure 3.2. Illustration of the WHO method to establish an epidemic threshold using respiratory syncytial virus detections from surveillance in Guatemala, April 2009–April 2018 (adapted from WHO Global Epidemiological Surveillance Standards for Influenza)

A. Draw smoothed (4-week moving average) epidemic curves with calendar week on the x-axis and the number of RSV detections on the y-axis.

B. Shift data to the point where peak weeks align on the median peak week

C. Calculate the average epidemic curve

D. Define epidemic threshold as the median weekly number of RSV detection

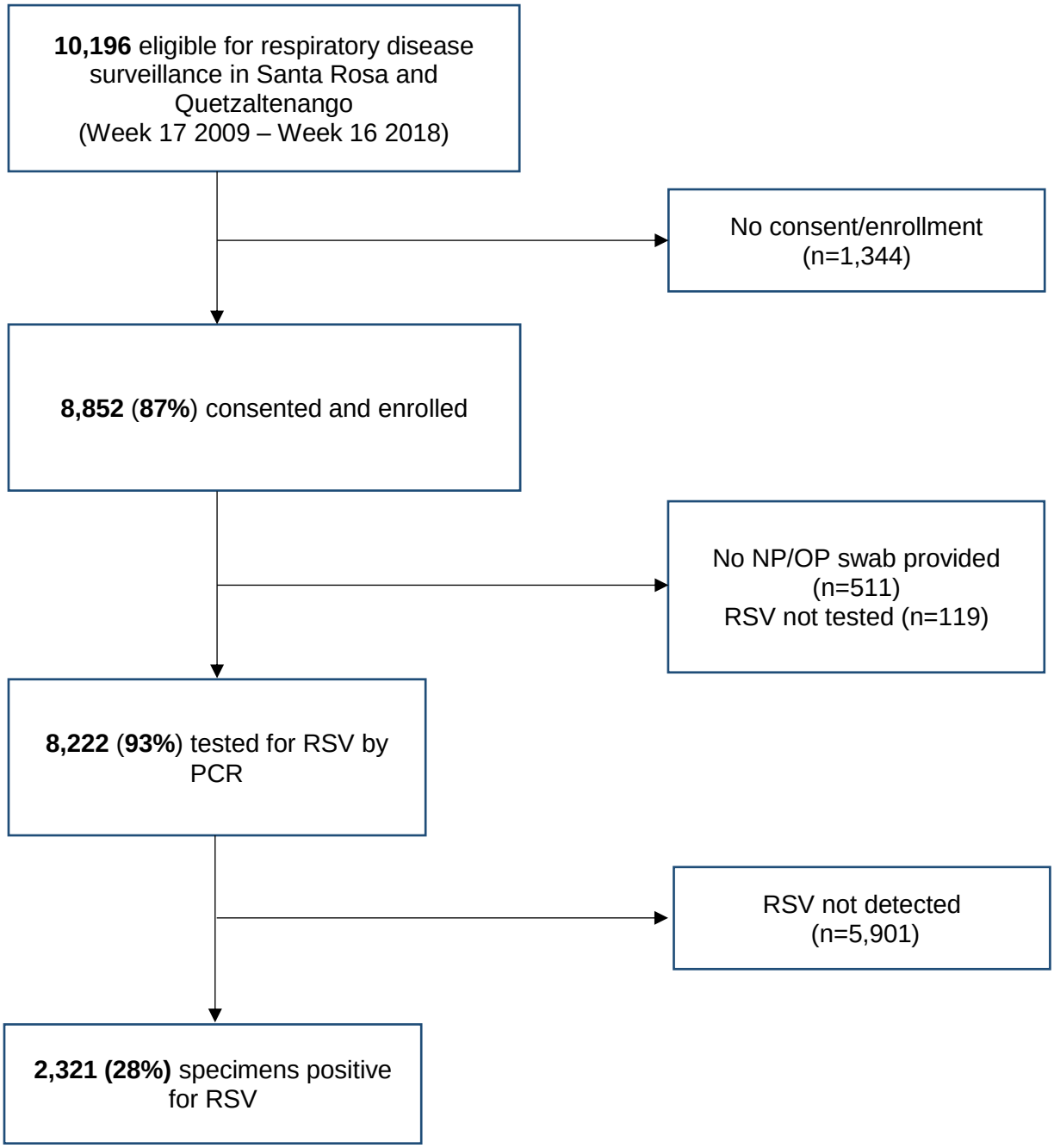


Figure 3.3. Selection of RSV-associated acute respiratory infection cases for analytic data set from hospitalized surveillance in Guatemala, 2009-2018

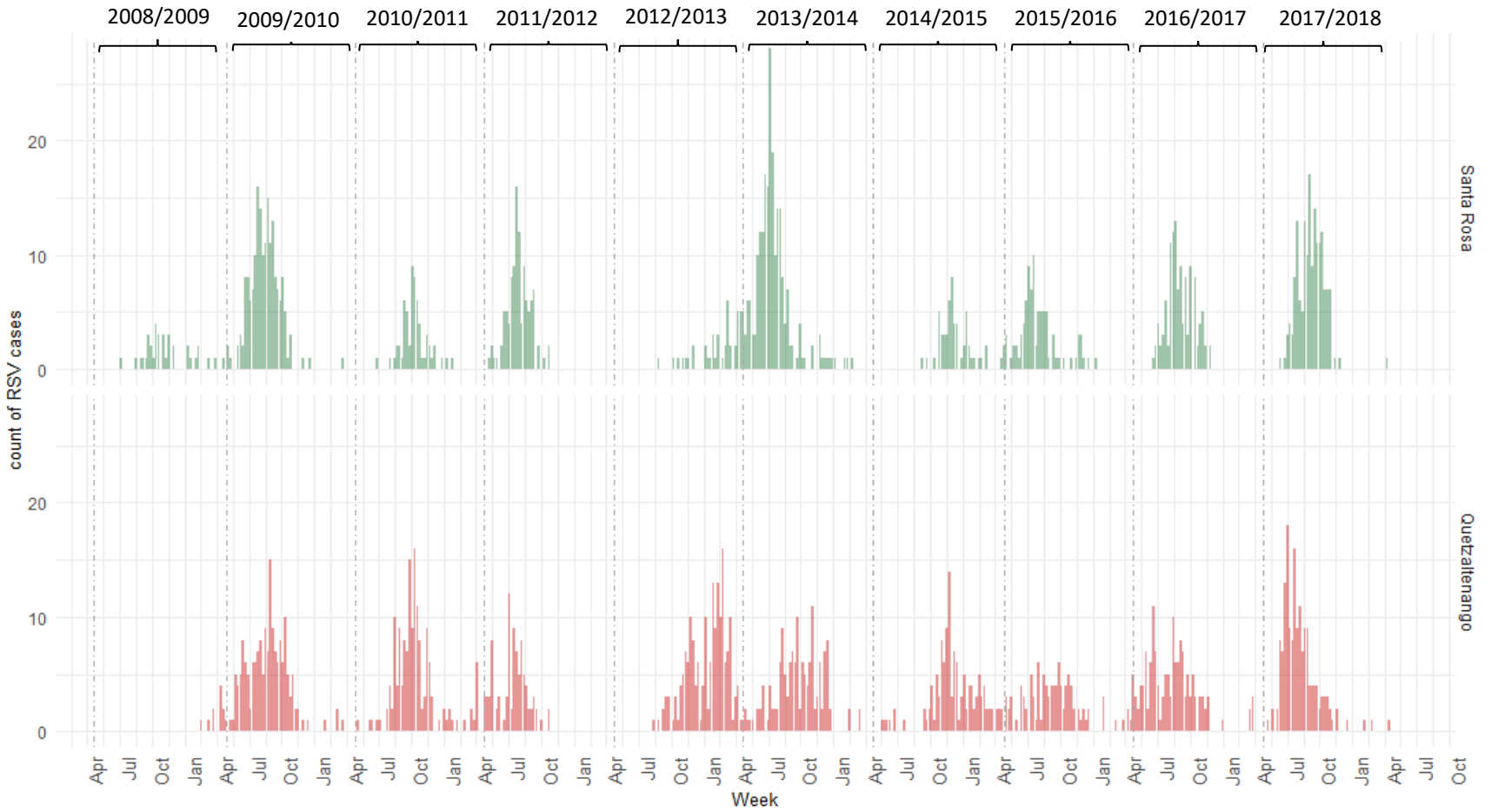


Figure 3.4. Epidemic curve of RSV cases reported through by surveillance site, Guatemala, 2007/08–2017/18 (n = 2,367 patients)

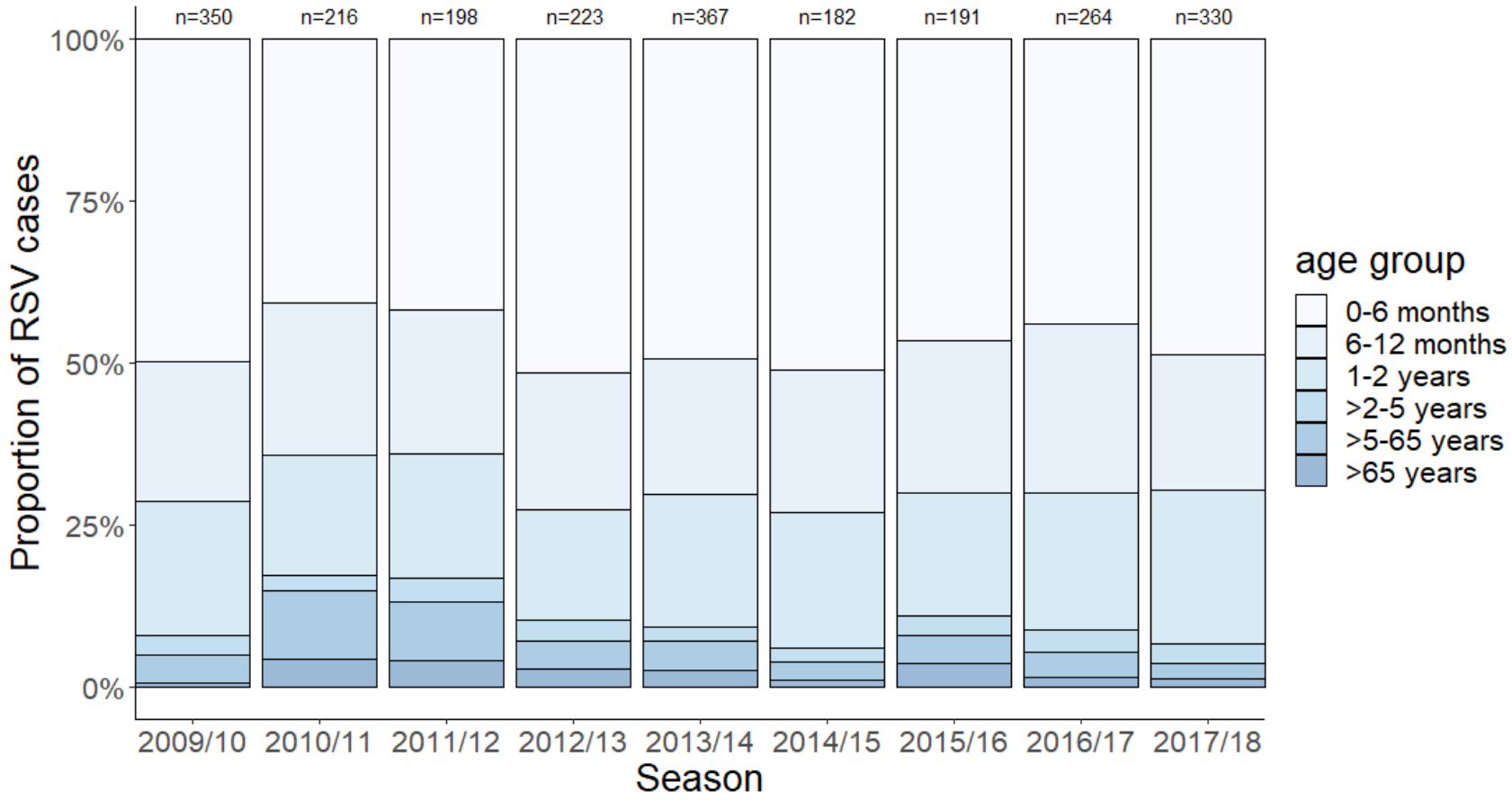


Figure 3.5. Distribution of patients with respiratory syncytial virus infections in 9 consecutive seasons, according to age group, Guatemala, 2009/10–2017/18 (n = 2321 patients)

Chapter 4. Effects of a 500-day liquefied petroleum gas stove intervention during gestation and infancy on respiratory illness in 2-year-old children

Introduction

Acute respiratory infections (ARI) exert a considerable burden and cost on health care systems.[1,2] Most of these infections are limited to the upper respiratory tract (i.e. the nose and throat). Globally in 2019, there were 237 million (95% UI: 212–265) prevalent cases of upper respiratory infections, contributing to 9,460 deaths (95% UI 5540–14 900) and 6·39 million (95% UI: 3·96–9·72) DALYs.[147] Findings from observational studies have established exposure to household air pollution (HAP) from solid fuels as an important risk factor for ARI.[1,148] Worldwide, it is estimated that 3 billion people rely on solid fuels (such as wood, charcoal, animal dung and agricultural residue) for cooking and heating.[103,149] Exposure to HAP from cooking with solid fuels is responsible for an estimated 2.3 million deaths and additional morbidity burden every year.[150] As the primary cooks, women bear the brunt of this exposure, and for their children, exposure begins in the prenatal period and often continues throughout the life course.

Many diseases have been linked to exposures experienced during gestation and early life. Nevertheless, relatively few studies have examined the effect of prenatal exposure to air pollution on infant lung function and early life respiratory infections, and most of these studies focused on ambient air pollution or maternal smoking.[123-125,151,152] Only two previous studies, the Drakenstein Child Health Study and the Ghana Randomized Air Pollution and Health Study (GRAPHS) have reported associations between prenatal HAP exposure and childhood respiratory outcomes.[125,153,154]

Although the mechanisms involved in the long-term effect of prenatal HAP exposure on respiratory illness have not been fully examined, there are critical periods of vulnerability to numerous adverse

outcomes in early development. Fetal lung development occurs across gestation and $PM_{2.5}$ can cross the placenta and may disrupt biological mechanisms that regulate fetal growth, maturation and development.[119-122] Findings from GRAPHS showed that prenatal HAP exposure impairs lung function in infants, potentially increasing subsequent risk of respiratory infections.[153] Prenatal HAP exposure is also associated with low birthweight, which in turn increases the risk of respiratory infections.[155-157]

Despite biologic plausibility for the effect of HAP exposure on pediatric respiratory health and consistent evidence from observational studies supporting the association, interventions aimed at reducing HAP exposure have struggled to demonstrate an impact on respiratory outcomes.[28,106,158,159] For example, one of the first randomized trials of a clean cook stove intervention (RESPIRE) estimated a 22% reduction in the primary outcome of physician-diagnosed pneumonia incidence, but this effect was not statistically significant.[114] However, RESPIRE found meaningful and statistically significant reductions for secondary outcomes of severe pneumonia. In Nepal, an improved biomass stove intervention showed weak evidence for a modest decline in the incidence of acute lower respiratory tract infections in children.[158] A subsequent intervention in Malawi, the Cooking And Pneumonia Study (CAPS), showed no effect of a cleaner burning biomass-fueled cookstove intervention on WHO Integrated Management of Childhood Illness–defined pneumonia incidence.[115,116] Similarly, GRAPHS saw no effect of a liquefied petroleum gas (LPG) intervention (compared to traditional stoves) on the incidence of child pneumonia.[117]

Findings of exposure-response analyses from these trials varied. RESPIRE exposure-response analyses demonstrated a protective effect of reductions in infant carbon monoxide (CO) exposure on physician-diagnosed pneumonia and severe pneumonia, whereas CAPS exposure-response analyses found no evidence of an association between CO exposure and pneumonia.[114,116] Both

analyses focused on the postnatal period. The GRAPHS exposure-response analysis focused on prenatal exposure and found that the risk for pneumonia and severe pneumonia in the first year of life increased by 10% and 15% respectively, per 1-part per million (ppm) increase in average prenatal CO exposure.[125]

A possible reason for the inability of past interventions to detect an effect on respiratory outcomes is that they did not achieve sufficient exposure contrast between intervention and control arms due to the continued use of traditional stoves and elevated ambient air pollution concentrations. For example, although RESPIRE reduced exposure by 50% on average (from 2.2 to 1.1 ppm carbon monoxide), exposure distributions for the intervention and control groups overlapped substantially.[114] In Nepal, the mean 20-hour kitchen $PM_{2.5}$ concentration was reduced from 1380 mg/m^3 to 936 mg/m^3, but this still exceeded WHO standards of 35 $\mu g/m^3$.[160,161] Similarly, although mean maternal $PM_{2.5}$ exposure in GRAPHS was 32% lower in the LPG arm compared to the control arm (52 ± 29 $\mu g/m^3$ vs 77 ± 44 $\mu g/m^3$), post-intervention exposures exceeded health-relevant targets.[118] In Malawi, median exposure to CO in both the intervention and control groups were below the WHO air quality guideline for CO (4 ppm).

The Household Air Pollution Intervention Network (HAPIN) trial has overcome this limitation, achieving a substantial reduction in personal exposure to $PM_{2.5}$, CO, and black carbon. HAPIN randomized households with pregnant women in four countries to receive an LPG stove and a continuous supply of fuel at no cost (ClinicalTrials.gov NCT02944682).[29] The intervention continued after pregnancy until the child reached 1 year of age, at which point fuel distribution ended. Post-randomization $PM_{2.5}$ exposures in the intervention arm were at the lower end of what has been reported for LPG and other clean fuel interventions, with 69% of $PM_{2.5}$ samples falling below the WHO Annual Interim Target of 35 $\mu g/m_3$.[162] Post-randomization CO levels were below

the WHO air quality standard in both arms (2.2 ppm in the intervention arm and 0.7 ppm in the control arm), and CO reductions exceeded those of RESPIRE (68% reduction vs. 50%).[114,162] For black carbon, post-randomization measurements were 11.1 ppm in the intervention arm and 4.0-4.3 ppm in the control arm.[162] The trial's primary health outcomes in children are birth weight, stunting at 12 months of age, and severe pneumonia in the first 12 months of life.

In primary HAPIN analyses, pneumonia was defined as (1) the presence of cough and/or difficult breathing and at least one general danger sign as defined by WHO, and primary endpoint pneumonia on a lung ultrasound or chest x-ray, or (2) the presence of cough and/or difficult breathing and hypoxemia (measured via pulse oximetry, SpO_2), or (3) death as a result of pneumonia as determined by verbal autopsy. This highly specific case definition for pneumonia was selected to minimize the possibility of nondifferential misclassification of pneumonia, which generally biases results towards the null hypothesis. For comparability with previous studies, the WHO definition of severe pneumonia is being assessed as a secondary outcome. A follow up period of 1 year was chosen because children younger than 1 year have the highest incidence of severe pneumonia. However, young children remain highly vulnerable to respiratory illness after the first year of life, and reducing morbidity and mortality in children under 5 years of age is a global health priority (SDG 3.2.1).

To assess the longer-term health impacts of the HAPIN intervention, three of the original four study sites are participating in a follow-up cohort study to follow HAPIN children through age 2 years and measure a range of exposures and outcomes including respiratory illness. Given the lower rate of respiratory illness in this age group, a more sensitive case definition is being used in this follow-up study than in the primary HAPIN analysis: care-giver reported illness with a cough at 24 months of life. Cough is a key symptom included in ARI case definitions that are commonly used

for surveillance purposes.[37,38] Few intervention studies have examined the effect of HAP on the mild-to-moderate spectrum of disease.[163,164] However, a recent water filter and improved cook stove intervention trial in Rwanda reduced the prevalence of reported ARI in children <5 years by 25% (PR 0.75, 95% CI 0.60–0.93). Although ARIs are not as serious as acute lower respiratory infections, they pose a burden for children, their families, and healthcare providers.[1] According to the 2014-15 Demographic Health Surveys, the 2-week period prevalence of ARIs in children under five years was 11% in Guatemala, 3% in India, and 6% in Rwanda, with higher rates among children age 6-23 months.[165-167] More than half of children with ARI sought advice or treatment from a health facility or provider.

Here we report on the effect of the original HAPIN intervention on the prevalence of illness with a cough in HAPIN children at age 2 years through an intent-to treat analysis, and examine exposure-response associations between prenatal HAP exposure and the prevalence of illness with a cough at 24 months of life.

Methods

Study population

Participants were from HAPIN, a multi-country randomized controlled trial of an LPG stove and fuel distribution intervention in 3,200 households in four low- and middle-income countries (India, Guatemala, Peru, and Rwanda) that has been described elsewhere.[29] The trial enrolled 800 pregnant women (aged 18-<35 years, 9 to <20 weeks gestation confirmed by ultrasound) in each study site and randomly assigned half to receive LPG stoves and an 18-month supply of fuel (ending when the child reached age 1 year). Participants in the control group continued using the stove and fuel of their choice, generally biomass stoves. To be eligible for enrollment, pregnant women had to be non-smokers and to cook with biomass stoves predominantly. To ensure balance between arms,

households were randomly allocated to intervention or control arms as and when they consented to participate; randomization was implemented at the site level after baseline exposure measurements were taken. In India, enrollment took place at two study sites (Nagapattinam and Villupuram). Guatemala and Rwanda had one study site each.

Upon exiting the study, participants in Guatemala, India, and Rwanda were invited to participate in a cohort study to follow children until age 2 years and evaluate whether the intervention had benefits beyond pregnancy and the child's first year of life. In Guatemala and Rwanda, LPG use was expected to be low after exiting the study due to lack of access and expense. In India, all control households received a free LGP stove and vouchers for 18 months of LPG fuel after exiting the study. Anticipated completion of follow up in the cohort study is June 2022.

Outcome ascertainment

Trained field workers conducted home visits at 24 months to assess child health status. The outcome was defined as caregiver-reported illness with a cough in the past 7 days. This case definition captures the mild-to-moderate spectrum of disease severity, predominantly upper respiratory tract infections, which nevertheless impair quality of life and productivity.[116]

Exposure assessment

Personal monitoring equipment was used to assess exposures to $PM_{2.5}$, CO, and BC over a 24-hour period in intervention and control participants. Full details on the method of assessment have been reported elsewhere.[162] Twenty-four-hour personal exposure for pregnant women was measured at baseline (prior to randomization), 24-28 weeks gestation, and 32-36 weeks gestation. At each monitoring period, pregnant participants were asked to wear a garment with instrumentation situated in the breathing zone, and asked to keep it nearby when sleeping, bathing, or conducting other activities for which the equipment could not be safely worn.[168] Prenatal exposure was defined as the time-weighted average of exposure measurements taken at baseline and at first and second

follow up visits during pregnancy. For participants in the control group, an average prenatal exposure was calculated from all available measurements. For participants in the intervention group, prenatal days prior to LPG installation were assigned the baseline measurement, and prenatal days following LPG installation were assigned to the average of all post-randomization measurements before birth. Future analyses will include exposure data collected at 6 months, 12 months, and 24 months. At 6 months and 12 months, BC and $PM_{2.5}$ in the home were measured. At 24 months, direct exposure measurements (monitor on child) are taken in Guatemala and Rwanda, and indirect measurements (mother wore a censor and child wore a beacon) are taken in India.

Covariates

At enrollment, a baseline survey was administered by trained field workers to obtain a range of information including the number of people who sleep in the home (a proxy for crowding), maternal health insurance status, passive smoking, and primary fuels used for cooking and heating. Household food insecurity was determined using a dietary diversity questionnaire and a food insecurity experience scale following methodology developed by FAO.[169] Breastfeeding through the first 6 months of life was determined through an infant feeding questionnaire. Child malnutrition (defined based on WHO simplified field table weight-for-length charts) was determined through anthropometry measurements and taken at 24 months of age. Child sex, birth weight, weight for gestational age z-score, and date of delivery were recorded at birth. Vaccination status was assessed at 9 and 12 months of age both by querying the mother and inspection of the child's vaccination card. Up-to-date vaccination at one-year of life was defined as receipt of 3 doses of pneumococcal conjugate vaccine, 3 doses of *Haemophilus influenzae* type b vaccine, and one dose of measles vaccine.

Statistical analysis

To identify imbalances in covariates between study arms, we compared characteristics of children in intervention and control households using means for continuous variables and proportions for

categorical variables. We also compared baseline characteristics of those who enrolled in the follow-up cohort and those who did not, using chi-square tests for categorical variables and the Wilcoxon rank sum test for continuous variables.

The outcome variable in all analyses was the prevalence of caregiver-reported illness with a cough (a key component of ARI case definitions) at 24 months of life. In an intent-to-treat analysis, we aimed to demonstrate the effect of the 500-day intervention over the most important period of early childhood development. We hypothesized that children born in intervention households have lower prevalence of illness with a cough at age 2 years relative to those born in control households. In an exposure-response analysis, we aimed to provide generalizable information transferable to other settings and interventions. We hypothesized that exposure to HAP pollutants during gestation (a critical developmental period) is positively associated with the prevalence of illness with a cough at age 2 years.

In the intent-to-treat analysis, we fit a log binomial regression model to estimate the prevalence ratio of illness with a cough in the intervention compared to the control group adjusting for randomization strata. The effect estimate from this model can be interpreted as the impact of the LPG intervention including any longer-term effects it may have on fuel use after the intervention ended. We did a sub-analysis with country-specific models because study participants in India received LPG stoves after the intervention period, resulting in bias towards the null.

We hypothesized that the effect of the intervention may be modified by country (India, Guatemala, and Rwanda), child sex, and gestational age at intervention (above vs. below the median of 18 weeks). To detect interaction by country and sex, we used a likelihood ratio test comparing the full model with interaction terms to the reduced model without the interaction terms. Because gestational age at intervention is measured in the intervention arm only, we ran two models: one

restricted to intervention participants below the median gestational age and the other restricted to intervention children above the median.

In the exposure-response analysis, we used log binomial regression to model the association between the prevalence of illness with a cough at 24 months and exposure to HAP pollutants ($PM_{2.5}$, BC, and CO) during gestation. We explored various transformations (e.g. linear, quadratic, cubic, natural log, or quintiles) of household air pollutants during gestation and selected the quadratic transformation because it gave the best model fit based on both the likelihood ratio test and Akaike Information Criterion. We estimated crude associations and associations adjusted for a priori confounders identified based on the literature and using the directed acyclic graphs (**Figure 4.1**). Hypothesized confounders included primary fuel used at 24 months, food insecurity score, vaccination status, breastfeeding in the first 6 months of life, number of people who sleep in the house, and season of birth (respiratory vs. non-respiratory season). We used likelihood ratio tests to assess interaction of prenatal exposure with country and sex. We also ran models stratified by gestational age at intervention.

Log binomial models were run in SAS version 9.4 (Cary, NC). All other models and analyses were run in R version 4.0.4 software (R Foundation for Statistical Computing).

Results

Study population

Of the 2,317 children enrolled in the HAPIN trial in Guatemala, India, and Rwanda, 2,171 (96%) enrolled in the follow-up cohort (**Figure 4.2**). Of these, 1783 (82%) had completed the 24-month child health status follow up visit (during which the prevalence of illness with a cough was ascertained) by February 2022. Of the 623 who did not have a 24-month follow up visit, 388 (62%) had not yet reached the follow up window (i.e. they were not yet 24 months of age) and 235 (38%) had reached 24 months of age but missed the health status follow up visit. Thus, 1548 (71%)

children enrolled in the follow-up cohort were included in the intent-to-treat analysis. Of these, 1440 (89%, or 66% of the follow-up cohort) had valid maternal prenatal $PM_{2.5}$ exposure assessments, 1478 (95%, or 68% of the follow-up cohort) had valid maternal prenatal CO assessments, and 1384 (89%, or 64% of the follow-up cohort) had valid maternal prenatal black carbon exposure assessments. Exposure measurements were more often missing in intervention arm children compared to control arm children (**Figure 4.2**).

In some countries, there were differences in those who did and did not enroll in the follow-up cohort in terms of study arm assignment, weeks during pregnancy with the intervention, maternal prenatal carbon monoxide exposure, food insecurity index, and stove use at baseline (**Table 4.1**). In Rwanda, a greater proportion of children in the intervention arm (9%) vs. the control arm (5%) did not enroll in the follow-up cohort, and those not enrolled in the follow-up cohort had lower average maternal prenatal CO exposure (1.2 vs 2.0 ppm). In India, those not enrolled in the follow-up cohort also had lower average maternal prenatal CO exposure (0.7 vs 1.6 ppm). In Guatemala and Rwanda, the number of weeks during pregnancy with the intervention was on average 2 weeks longer in those who did versus those who did not enroll in the follow-up cohort. In India, those not enrolled had a higher proportion of moderate/severe food insecurity (12% vs. 3.5%) and were less likely to use traditional cookstove for heating at baseline (0% vs 8.2%).

Among the 1548 children included in the intent-to-treat analysis, baseline characteristics were well balanced between intervention and control arms in all countries except for Rwanda, where a higher proportion of control households reported having a smoker (3.9% vs. 0.5%), moderate/severe food insecurity (39% vs. 24%), and wood as the primary fuel type at baseline (78% vs. 66%) (**Table 4.2**). In Guatemala and India, nearly all participating households used wood for cooking at baseline (n=1102, 99%). Half of the children included in the analyses were girls (n=727, 47%). *Haemophilus influenzae* type b vaccination rates were high, with 562 (82%), 414 (83%), and 421 (96%) children

receiving 3 doses by the first year of life in Guatemala, India, and Rwanda respectively. Pneumococcal vaccination coverage was high in Rwanda, where 386 (88%) children received 3 doses in the first year of life, but low in Guatemala (n=88, 14%) and India (n=2, 0.4%). Receipt of at least one dose of measles vaccine was highest in Rwanda (n=408, 93%), followed by India (n=380, 76%), and Guatemala (n=80, 13%).

During child health status visits at 24 months, 233 caretakers reported that their child had illness with a cough in the past 7 days. The prevalence of illness with a cough at 24 months was much higher in Rwanda (n=165, 38%) than in Guatemala (n=49, 8%) and India (n=18, 4%). There was no association between report of illness with a cough at 24 months and the type of fuel used for cooking at 24 months (**Table 4.3**).

Although maternal prenatal pollutant exposure distributions overlapped, the trial achieved strong contrasts in maternal prenatal pollutant exposure between intervention and control arms (**Figure 4.3**). In the intervention arm, the time-weighted average CO level was 1.3 (SD 1.8) ppm in the intervention arm compared to 2.0 (SD 2.6) ppm in the control arm. For black carbon, the time-weighted average level was 8.8 (SD 6.8) $\mu g/m^3$ in intervention arm compared to 12.0 (SD 7.0) in the control arm. For $PM_{2.5}$, the time-weighted average level was 82.9 (SD 67.0) $\mu g/m^3$ in intervention arm compared to 116.6 (SD 85.1) in the control arm. These strong contrasts were observed across study sites (**Table 4.2**).

Intent-to-treat analysis

The intention-to-treat analysis used randomization allocation as the indicator of exposure. In both crude analyses and analyses adjusted for randomization strata, passive smoking, primary fuel used at baseline, and food insecurity at baseline, we did not find evidence that the intervention had an effect on the prevalence of illness with a cough at 24 months of age (**Table 4.4**). We did not find statistical evidence of interaction between study arm and child sex or country. Moreover, we did not find

meaningful differences in prevalence ratios by sex using sex-specific models (PR for males: 1.04 [95%CI: 0.76, 1.41], PR for females: 1.05 [95%CI; 0.78, 1.41]). Country-specific prevalence ratios were heterogeneous and were not statistically significant (**Table 4.4**). The estimated prevalence ratio from the model restricted to intervention participants below the median gestational age at intervention initiation (PR: 1 .07 [95%CI: 0.82, 1.38]) was similar to that from the model restricted intervention children above the median gestational age at intervention initiation (PR: 1.02 [95%CI: 0.79, 1.33]).

Exposure-response analyses

In exposure-response analyses, we did not detect an association between any pollutant and the prevalence of illness with a cough at 24 months of life (**Figure 4.4**). Confidence intervals were wide and included the null at most exposure levels. Models adjusted for hypothesized confounders (sex, food insecurity index at 24 months [ordinal], exclusive breast feeding until 6 months of life, the number of people who sleep in the house, season of birth, and randomization strata) had little impact on prevalence estimates **(Figure 4.4**). We did not find statistical evidence of interaction between pollutant levels and sex or country. However, country-specific models suggested some heterogeneity in the association between pollutant levels and the outcome across countries (**Figure 4.5**). Prevalence estimates diverged somewhat in sub-analyses stratified by intervention arm participants who received the intervention prior to 18 weeks gestation (the median gestational age at intervention initiation) and those who received the intervention after 18 weeks gestation. However, all estimates had wide confidence intervals that included the null (**Figure 4.6**).

Discussion

Upper respiratory infections, including cough, are one of the most common diseases. Exposure to household air pollution from cooking with solid fuels is recognized as a key risk factor for upper

respiratory infection. Few previous studies have examined the effect of prenatal HAP exposure on early childhood respiratory infections.

Neither intent-to-treat nor exposure-response analyses of children enrolled in the HAPIN follow-up cohort provided evidence of an association between prenatal HAP exposure and the prevalence of illness with a cough in children at 24 months of life. Although our study was restricted to low-income households in low- and middle-income countries, we found considerable variability in the prevalence of illness with a cough across countries.

We had hypothesized that prenatal HAP exposure would affect the prevalence of illness with a cough in young children through various pathways. In utero exposure to HAP may alter fetal lung development, predisposing infants to future respiratory infections. This hypothesis was supported by findings from the GRAPHS study, which demonstrated that increased prenatal HAP exposure was associated with impaired infant lung function, which in turn, increased the risk of child pneumonia.[125] HAP exposure is also a risk factor for low birthweight, which is associated with respiratory illness in children.[129]

Despite biological plausibility for the association between prenatal HAP exposure and illness with a cough in early childhood, there are several reasons why we may have failed to detect an effect. Firstly, our outcome definition was highly sensitive but non-specific. Evidence from previous studies suggests that HAP exposure has a greater effect on the severe spectrum of respiratory disease.[28] Secondly, we had low power to detect an effect, as demonstrated by the wide confidence intervals around our estimates. Thirdly, prenatal exposure may be too distal to respiratory outcomes in children at 24 months of age. The few previous studies that reported the association between prenatal HAP exposure and respiratory outcomes in children looked at outcomes during the first year of life, when the incidence of respiratory infections is highest.[125,154]Finally, while the HAPIN

trial achieved a strong exposure contrast between study arms, our time-weighted measure of prenatal exposure included baseline exposure measurements taken before the intervention was initiated, dampening the exposure reduction (and contrast) resulting from the LPG intervention.

Few studies have examined the influence of prenatal household air pollution exposure on early childhood respiratory infections, and differences in outcome definitions across studies make direct comparisons difficult. To our knowledge no previous study used our outcome (illness with a cough at 24 months of life). The GRAPHS study examined the association of prenatal and postal HAP exposure on pneumonia and severe pneumonia in the first year of life. That study found a linear exposure-response relationship between prenatal CO exposure and the risk of pneumonia and severe pneumonia. In South Africa, the Drakenstein Child Health Study also reported associations between prenatal household air pollution (specifically, PM_{10} of more than the ambient standard) and childhood lower respiratory tract infections during the first year of life.[154] A cohort study in Poland found that prenatal exposure to $PM_{2.5}$ increased odds of recurrent (five or more) broncho-pulmonary infections by 7 years of age."[124]

Underlying reasons for the substantially higher prevalence (38%) of illness with a cough in Rwanda relative to India and Guatemala are unclear and warrant further research. However, this finding is consistent with other reports. For example, a large-scale, combined water filter and cookstove intervention trial in Western Province, Rwanda that used a more specific case definition (cough accompanied by reported rapid breathing or difficulty breathing) found an ARI prevalence of 9.9% in the intervention arm and 14.3% in the control arm.[170] Furthermore, the main HAPIN trial found that Rwanda had a considerably higher pneumonia rate than other trial sites (to be reported in the as-yet-unpublished main HAPIN analysis). It is also possible that differences in the interpretation of

study questionnaires (which are translated into local languages in each site) accounts for some of the difference in the prevalence of caregiver-reported illness with a cough across countries.

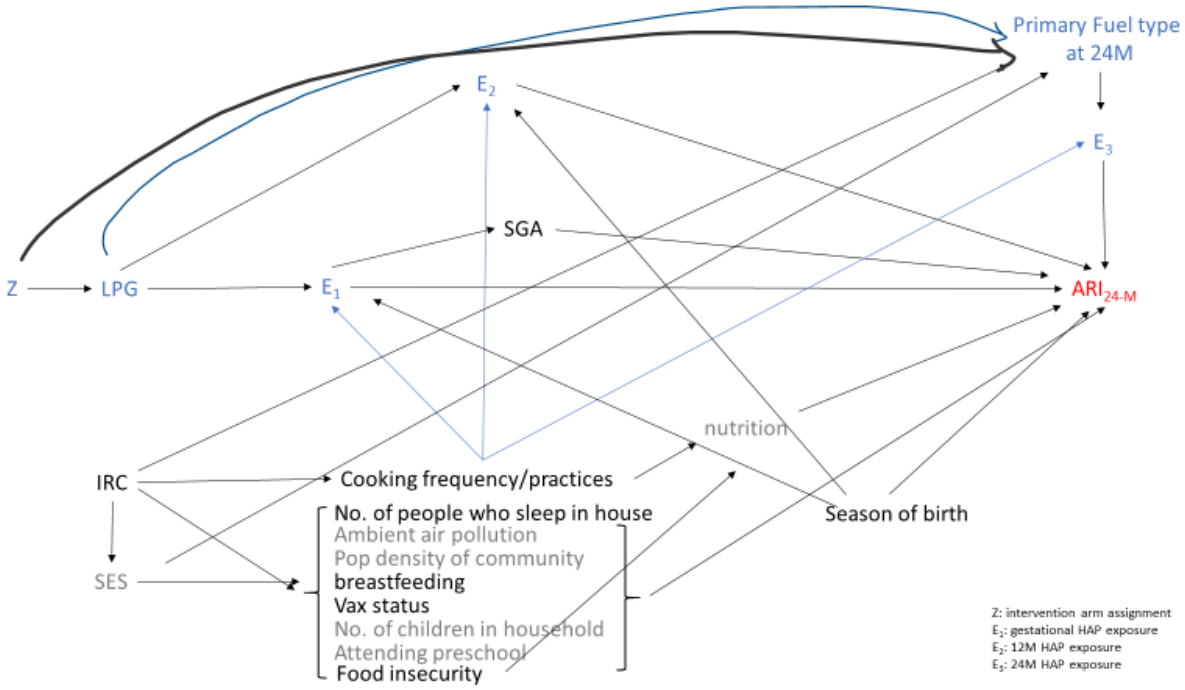
We note several strengths of our study. The experimental design with good balance of baseline covariates between intervention and control arms minimized the potential for confounding in intent-to-treat analyses. We prospectively measured exposure to three maternal household air pollutants at three time points prenatally, and achieved a good exposure contrast between intervention and control arms. The restriction of our study population to pregnant women from low socioeconomic communities in three low- and middle-income countries mitigated the potential for confounding by socioeconomic status in exposure-response analyses, and ensured that our study was representative of the populations most affected by HAP. Moreover, our study cohort was well-characterized with both individual- and household-level measures of potential confounders and mediators.

This study had several limitations that bear noting. Firstly, our outcome definition (caregiver-reported illness with a cough) was non-specific and captured mild respiratory illnesses, on which HAP exposure is less likely to have an effect. The impact of HAP exposure on severe pneumonia in the first year of life will be reported elsewhere. Secondly, although the study enrolled a high-risk population, exposures in the control arm were not as high as those reported in some earlier studies, such as RESPIRE; larger effects may be seen in children born to mothers with higher exposures.[114] Thirdly, although our exposure measurement strategy exceeds that of previous HAP studies, we did not measure maternal prenatal exposure continuously over the entire study period; 24-hour exposure measurements might not be representative of exposures over longer periods of time. Finally, at this writing, follow up in the HAPIN cohort is ongoing, so this analysis was based on partial data, which

did not include postnatal HAP exposure measurements. Nevertheless, exposure-response analyses adjusted for primary fuel type at 24 months (a strong predictor of exposure).

In conclusion, upper respiratory infections are among the most common diseases, and symptoms can significantly impair quality of life and productivity. HAP is considered a key risk factor for upper respiratory tract infections. However, our results do not suggest that reducing HAP exposures in the prenatal period can reduce the prevalence of early childhood illness with a cough. The non-specificity or our outcome definition was an important limitation, and our findings support previous evidence suggesting that the measurable effects of HAP are restricted to more severe outcomes. Future work will examine the role of postnatal HAP on prevalence of illness with a cough.

Tables and Figures



- The intervention arm assignment (Z) is an instrumental variable when examining the effect of the LPG stove on prevalence of illness with cough. After the intervention ends, we expect that the intervention might influence the primary fuel type used by households (LPG→Primary fuel type at 24M). Participants who received the intervention might be more likely to continue to use LPG in Rwanda and Guatemala. Conversely, in India, LPG fuel was provided to control households after the intervention, hence the arrow from Z to primary fuel type at 24M.
- HAP exposure during gestation is a hypothesized risk factor for small for gestational age (SGA) and prematurity (gestational age),[171,172] which are risk factors for acute respiratory infections[173,174] (ARI, proxied by illness with a cough in this analysis).
- IRC (i.e. study country) might modify the effect of the exposure on the outcome but it can also be conceptualized as a confounder. Cooking practices differ across countries (IRCs) and this could affect exposure to household air pollution. Cooking practices may also affect nutrition, which is a risk factor for ARI. There are many other risk factors for ARI that we would expect to correlate with IRC. They include the number of people who sleep in the house (i.e. crowding), ambient air pollution exposure, population density of the community of residence (a proxy for contact rate), breastfeeding, vaccination status, number of children living in a household, preschool attendance, and food insecurity.
- SES might affect the type of fuel used for cooking[175] after the intervention ends and is also a risk factor for ARI. However, within the narrow SES stratum enrolled in the trial, variation in assets/wealth might not have an impact on ARI.
- Season or month of birth might affect exposure to HAP pollutants because fuel use and cooking patterns can differ by season. For example, in colder months families might burn biomass fuels for heating. Respiratory pathogens (a necessary but not sufficient cause of ARI) also tend to circulate in season patterns.

Figure 4.1. Directed acyclic graph depicting hypothesized causal relationships for the effect of prenatal household air pollution exposure on the prevalence of illness with a cough at 24 months in the HAPIN follow-up cohort study.

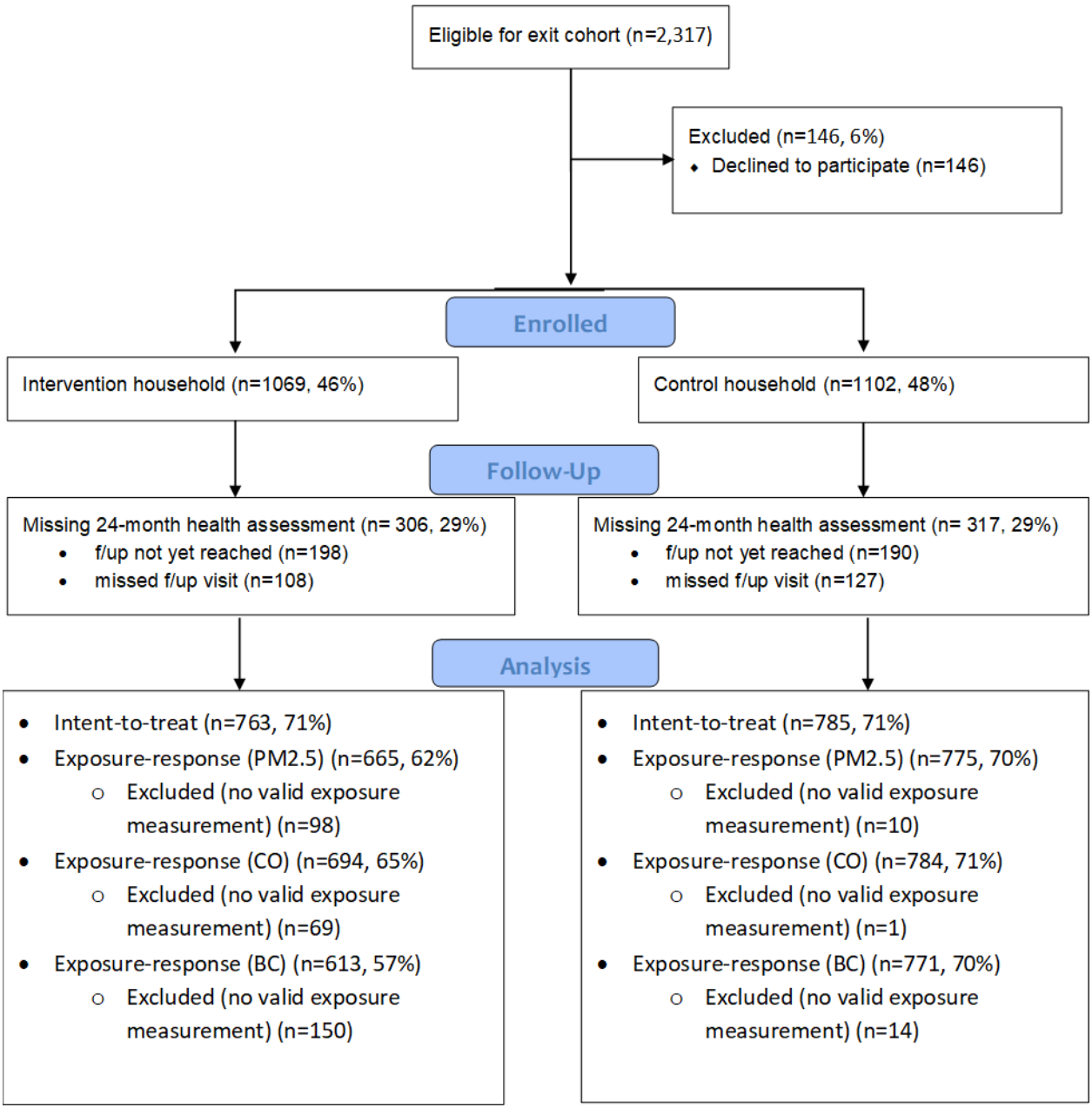


Figure 4.2. Flow diagram of participation in the HAPIN follow-up cohort study

Table 4.1. Characteristics of eligible households and children who did and did not enroll in the HAPIN follow-up cohort study

Characteristic	Guatemala			India			Rwanda			Pooled		
	not enrolled, N = 26[1]	enrolled, N = 744[1]	p-value[2]	not enrolled, N = 65[1]	enrolled, N = 709[1]	p-value[2]	not enrolled, N = 55[1]	enrolled, N = 718[1]	p-value[2]	not enrolled, N = 146[1]	enrolled, N = 2,171[1]	p-value[2]
Sex			0.9			0.3			0.2			>0.9
Male	13 (50%)	384 (52%)		31 (48%)	385 (54%)		33 (60%)	370 (52%)		77 (53%)	1,139 (52%)	
Female	13 (50%)	360 (48%)		34 (52%)	324 (46%)		22 (40%)	348 (48%)		69 (47%)	1,032 (48%)	
Study arm			>0.9			0.7			0.025			0.11
Control	13 (50%)	373 (50%)		31 (48%)	355 (50%)		20 (36%)	374 (52%)		64 (44%)	1,102 (51%)	
Intervention	13 (50%)	371 (50%)		34 (52%)	354 (50%)		35 (64%)	344 (48%)		82 (56%)	1,069 (49%)	
Gestational age at birth (weeks)	37.4 (4.4)	39.2 (1.4)	0.041	38.9 (1.5)	38.9 (1.5)	0.9	38.8 (3.7)	40.0 (1.6)	0.2	38.6 (3.1)	39.3 (1.6)	0.2
No. of weeks during pregnancy with intervention	19.1 (4.1)	21.3 (3.5)	0.038	19.6 (3.1)	19.9 (3.6)	0.6	20.0 (4.2)	21.9 (3.4)	0.022	19.7 (3.7)	21.0 (3.6)	0.003
N/A (control arm)	13	373		31	355		20	374		64	1102	
Gestational age at intervention	19.3 (4.1)	17.9 (3.2)	0.3	19.4 (2.6)	19.0 (3.5)	0.4	18.7 (2.9)	18.1 (3.2)	0.2	19.1 (3.0)	18.3 (3.3)	0.031
N/A (control arm)	13	373		31	355		20	374		64	1102	
Carbon Monoxide (ppm)	1.8 (1.3)	1.5 (1.6)	0.076	0.7 (0.7)	1.6 (2.3)	<0.001	1.2 (1.3)	2.0 (2.7)	0.016	1.1 (1.1)	1.7 (2.2)	<0.001
Unknown	1	24		4	34		3	37		8	95	
Black Carbon (µg/m3)	11.6 (4.3)	10.9 (6.4)	0.3	9.5 (7.7)	10.6 (8.6)	0.2	9.4 (4.5)	10.5 (6.5)	0.4	9.8 (6.2)	10.7 (7.2)	0.11
Unknown	3	72		5	70		11	129		19	271	
PM2.5 (µg/m3)	134.5 (84.8)	112.1 (83.4)	0.2	75.0 (48.7)	96.4 (93.9)	0.3	86.4 (65.2)	94.4 (72.6)	0.2	89.5 (65.5)	101.2 (84.0)	0.11
Unknown	3	46		5	58		3	54		11	158	
No. of people who sleep in home	5.7 (3.3)	5.2 (2.6)	0.6	3.9 (1.5)	3.8 (1.5)	0.5	3.2 (1.3)	3.5 (1.5)	0.13	4.0 (2.1)	4.2 (2.1)	0.2
Mother's health insurance at baseline			>0.9			0.4			0.2			0.9
None	26 (100%)	722 (97%)		65 (100%)	679 (96%)		5 (9.1%)	46 (6.4%)		96 (66%)	1,447 (67%)	
Public	0 (0%)	12 (1.6%)		0 (0%)	21 (3.0%)		48 (87%)	660 (92%)		48 (33%)	693 (32%)	
Private	0 (0%)	8 (1.1%)		0 (0%)	8 (1.1%)		2 (3.6%)	12 (1.7%)		2 (1.4%)	28 (1.3%)	
Unknown	0	2		0	1					0	3	
Someone in house smokes at baseline			0.2			0.9			0.4			0.069
No	23 (88%)	706 (95%)		44 (68%)	486 (69%)		52 (95%)	691 (97%)		119 (82%)	1,883 (87%)	
Yes	3 (12%)	38 (5.1%)		21 (32%)	223 (31%)		3 (5.5%)	25 (3.5%)		27 (18%)	286 (13%)	
Unknown							0	2		0	2	
Low birth weight (<2500g)			<0.001			0.082			0.008			0.004
No	9 (38%)	617 (85%)		46 (71%)	423 (60%)		40 (80%)	651 (91%)		95 (68%)	1,691 (79%)	
Yes	15 (62%)	110 (15%)		19 (29%)	285 (40%)		10 (20%)	62 (8.7%)		44 (32%)	457 (21%)	
Unknown	2	17		0	1		5	5		7	23	
Household food insecurity at baseline			>0.9			0.007			0.4			0.5
None	15 (58%)	409 (56%)		46 (72%)	582 (82%)		24 (45%)	261 (37%)		85 (59%)	1,252 (58%)	
Mild	9 (35%)	237 (32%)		10 (16%)	99 (14%)		12 (23%)	206 (29%)		31 (22%)	542 (25%)	

Moderate/Severe	2 (7.7%)	89 (12%)		8 (12%)	25 (3.5%)		17 (32%)	234 (33%)		27 (19%)	348 (16%)	
Unknown	0	9		1	3		2	17		3	29	
Primary fuel type at baseline			>0.9						0.5			0.3
Wood	26 (100%)	737 (100%)		65 (100%)	709 (100%)		39 (71%)	522 (73%)		130 (89%)	1,968 (91%)	
Other	0 (0%)	3 (0.4%)					0 (0%)	18 (2.5%)		16 (11%)	176 (8.1%)	
Unknown	0	4					0	2		0 (0%)	21 (1.0%)	
Charcoal							16 (29%)	176 (25%)		0	6	
Primary heating source at baseline			0.7			0.024			0.077			0.003
None	18 (69%)	493 (66%)		65 (100%)	644 (91%)		54 (98%)	710 (99%)		137 (94%)	1,847 (85%)	
Traditional cookstove	6 (23%)	211 (28%)		0 (0%)	58 (8.2%)		0 (0%)	6 (0.8%)		6 (4.1%)	275 (13%)	
Other	2 (7.7%)	40 (5.4%)		0 (0%)	7 (1.0%)		1 (1.8%)	0 (0%)		3 (2.1%)	47 (2.2%)	
Unknown							0	2		0	2	

[1] n (%); Mean (SD)
[2] Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test

Table 4.2. Characteristics of participants in the HAPIN follow-up cohort by study arm and country

	Guatemala		India		Rwanda		Pooled	
	Control	Intervention	Control	Intervention	Control	Intervention	Control	Intervention
Characteristic	N = 298[1]	N = 311[1]	N = 256[1]	N = 243[1]	N = 231[1]	N = 209[1]	N = 785[1]	N = 763[1]
Sex								
Male	151 (51%)	169 (54%)	141 (55%)	129 (53%)	123 (53%)	108 (52%)	415 (53%)	406 (53%)
Female	147 (49%)	142 (46%)	115 (45%)	114 (47%)	108 (47%)	101 (48%)	370 (47%)	357 (47%)
Gestational age at birth (weeks)	39.1 (1.4)	39.2 (1.5)	38.9 (1.4)	38.8 (1.6)	40.0 (1.5)	39.9 (1.5)	39.3 (1.5)	39.3 (1.6)
Weight for Gestational age z-score (median, IQR)	-0.9 (-1.5, -0.2)	-0.8 (-1.4, -0.2)	-1.5 (-2.0, -0.9)	-1.5 (-2.1, -1.0)	-0.8 (-1.3, -0.2)	-0.7 (-1.3, -0.0)	-1.1 (-1.7, -0.4)	-1.0 (-1.6, -0.4)
Unknown	4	5			6	3	10	8
No. of weeks during pregnancy with intervention	NA (NA)	21.1 (3.5)	NA (NA)	19.3 (3.5)	NA (NA)	21.7 (3.4)	NA (NA)	20.7 (3.6)
Unknown	298	0	256	0	231	0	785	0
Gestational age at intervention	NA (NA)	18.1 (3.3)	NA (NA)	19.5 (3.4)	NA (NA)	18.3 (3.3)	NA (NA)	18.6 (3.4)
Unknown	298	0	256	0	231	0	785	0
Carbon Monoxide	1.9 (1.9)	1.2 (1.2)	2.1 (3.0)	1.1 (1.6)	2.2 (3.0)	1.8 (2.6)	2.0 (2.6)	1.3 (1.8)
Unknown	0	24	1	20	0	25	1	69
Black Carbon	12.6 (5.4)	9.0 (7.5)	11.7 (9.5)	8.9 (7.4)	11.5 (5.2)	8.2 (4.1)	12.0 (7.0)	8.8 (6.8)
Unknown	5	56	4	38	5	56	14	150
PM2.5	136.2 (91.7)	88.4 (62.8)	106.5 (93.7)	79.3 (82.8)	102.8 (57.6)	78.3 (49.9)	116.6 (85.1)	82.9 (67.0)
Unknown	5	32	3	32	2	34	10	98
No. of people who sleep in home	5.2 (2.7)	5.2 (2.5)	3.7 (1.5)	3.7 (1.5)	3.4 (1.4)	3.5 (1.5)	4.2 (2.1)	4.3 (2.1)
Mother's health insurance at baseline								
None	292 (98%)	299 (97%)	246 (96%)	229 (94%)	14 (6.1%)	15 (7.2%)	552 (70%)	543 (71%)
Public	2 (0.7%)	7 (2.3%)	8 (3.1%)	9 (3.7%)	216 (94%)	192 (92%)	226 (29%)	208 (27%)
Private	4 (1.3%)	3 (1.0%)	2 (0.8%)	5 (2.1%)	1 (0.4%)	2 (1.0%)	7 (0.9%)	10 (1.3%)
Unknown	0	2					0	2
Mother's health insurance at 24M								
None	287 (97%)	296 (95%)	253 (100%)	243 (100%)	35 (16%)	18 (9.6%)	575 (75%)	557 (75%)
Public	4 (1.4%)	8 (2.6%)			176 (82%)	165 (88%)	180 (24%)	173 (23%)
Private	3 (1.0%)	5 (1.6%)	1 (0.4%)	0 (0%)	3 (1.4%)	4 (2.1%)	7 (0.9%)	9 (1.2%)
Other	1 (0.3%)	1 (0.3%)					1 (0.1%)	1 (0.1%)
Unknown	3	1	2	0	17	22	22	23
Malnutrition at 24M								
No	294 (99%)	306 (98%)	251 (98%)	243 (100%)	105 (100%)	82 (100%)	650 (99%)	631 (99%)
Yes	4 (1.3%)	5 (1.6%)	5 (2.0%)	0 (0%)			9 (1.4%)	5 (0.8%)
Unknown					126	127	126	127

Someone in house smokes at baseline								
No	284 (95%)	295 (95%)	173 (68%)	159 (65%)	221 (96%)	207 (100%)	678 (86%)	661 (87%)
Yes	14 (4.7%)	16 (5.1%)	83 (32%)	84 (35%)	9 (3.9%)	1 (0.5%)	106 (14%)	101 (13%)
Unknown					1	1	1	1
Someone in house smokes at 24M								
No	275 (99%)	282 (99%)	223 (88%)	213 (89%)	222 (96%)	204 (98%)	720 (94%)	699 (95%)
Yes	4 (1.4%)	3 (1.1%)	30 (12%)	27 (11%)	9 (3.9%)	5 (2.4%)	43 (5.6%)	35 (4.8%)
Unknown	19	26	3	3			22	29
Month of birth								
Jan	39 (13%)	42 (14%)	38 (15%)	25 (10%)	21 (9.1%)	20 (9.6%)	98 (12%)	87 (11%)
Feb	28 (9.4%)	33 (11%)	18 (7.0%)	18 (7.4%)	7 (3.0%)	6 (2.9%)	53 (6.8%)	57 (7.5%)
Mar	14 (4.7%)	21 (6.8%)	9 (3.5%)	12 (4.9%)	18 (7.8%)	16 (7.7%)	41 (5.2%)	49 (6.4%)
Apr	19 (6.4%)	20 (6.4%)	15 (5.9%)	12 (4.9%)	24 (10%)	33 (16%)	58 (7.4%)	65 (8.5%)
May	21 (7.0%)	26 (8.4%)	13 (5.1%)	15 (6.2%)	26 (11%)	24 (11%)	60 (7.6%)	65 (8.5%)
Jun	19 (6.4%)	20 (6.4%)	15 (5.9%)	18 (7.4%)	19 (8.2%)	13 (6.2%)	53 (6.8%)	51 (6.7%)
Jul	18 (6.0%)	14 (4.5%)	23 (9.0%)	19 (7.8%)	21 (9.1%)	17 (8.1%)	62 (7.9%)	50 (6.6%)
Aug	33 (11%)	23 (7.4%)	29 (11%)	19 (7.8%)	23 (10.0%)	21 (10%)	85 (11%)	63 (8.3%)
Sep	27 (9.1%)	36 (12%)	21 (8.2%)	29 (12%)	23 (10.0%)	22 (11%)	71 (9.0%)	87 (11%)
Oct	31 (10%)	28 (9.0%)	31 (12%)	30 (12%)	19 (8.2%)	13 (6.2%)	81 (10%)	71 (9.3%)
Nov	26 (8.7%)	24 (7.7%)	25 (9.8%)	25 (10%)	13 (5.6%)	8 (3.8%)	64 (8.2%)	57 (7.5%)
Dec	23 (7.7%)	24 (7.7%)	19 (7.4%)	21 (8.6%)	17 (7.4%)	16 (7.7%)	59 (7.5%)	61 (8.0%)
Low birth weight (<2500g)								
No	244 (83%)	258 (84%)	155 (61%)	140 (58%)	209 (91%)	192 (92%)	608 (78%)	590 (78%)
Yes	50 (17%)	48 (16%)	101 (39%)	103 (42%)	20 (8.7%)	17 (8.1%)	171 (22%)	168 (22%)
Unknown	4	5			2	0	6	5
Household food insecurity at baseline								
None	161 (55%)	174 (57%)	209 (82%)	203 (84%)	71 (31%)	94 (46%)	441 (57%)	471 (63%)
Mild	98 (33%)	97 (32%)	38 (15%)	30 (12%)	66 (29%)	60 (30%)	202 (26%)	187 (25%)
Moderate/Severe	36 (12%)	36 (12%)	8 (3.1%)	8 (3.3%)	89 (39%)	49 (24%)	133 (17%)	93 (12%)
Unknown	3	4	1	2	5	6	9	12
Household food insecurity at 24M								
None	249 (85%)	253 (83%)	206 (80%)	202 (84%)	64 (28%)	73 (35%)	519 (67%)	528 (70%)
Mild	31 (11%)	30 (9.9%)	41 (16%)	31 (13%)	49 (21%)	52 (25%)	121 (16%)	113 (15%)
Moderate/Severe	14 (4.8%)	20 (6.6%)	9 (3.5%)	8 (3.3%)	117 (51%)	83 (40%)	140 (18%)	111 (15%)
Unknown	4	8	0	2	1	1	5	11
Primary fuel type at baseline								
Wood	293 (99%)	310 (100%)	256 (100%)	243 (100%)	179 (78%)	137 (66%)	728 (93%)	690 (91%)
LPG/electric	2 (0.7%)	1 (0.3%)			2 (0.9%)	3 (1.4%)	4 (0.5%)	4 (0.5%)
Unknown	3	0			1	1	4	1
Charcoal					49 (21%)	68 (33%)	49 (6.3%)	68 (8.9%)
Primary fuel type at 24M								

Wood	265 (95%)	245 (86%)	18 (7.1%)	70 (29%)	182 (79%)	140 (67%)	465 (61%)	455 (62%)
LPG/electric	14 (5.0%)	40 (14%)	235 (93%)	170 (71%)	9 (3.9%)	14 (6.7%)	258 (34%)	224 (30%)
Unknown	19	26	3	2			22	28
Cow dung			0 (0%)	1 (0.4%)			0 (0%)	1 (0.1%)
Charcoal					40 (17%)	55 (26%)	40 (5.2%)	55 (7.5%)
Primary heating source at baseline								
None	199 (67%)	215 (69%)	239 (93%)	218 (90%)	228 (99%)	207 (100%)	666 (85%)	640 (84%)
Traditional cookstove	90 (30%)	90 (29%)	14 (5.5%)	23 (9.5%)	2 (0.9%)	1 (0.5%)	106 (14%)	114 (15%)
Other	9 (3.0%)	6 (1.9%)	3 (1.2%)	2 (0.8%)			12 (1.5%)	8 (1.0%)
Unknown					1	1	1	1
Primary heating source at 24M								
None	260 (93%)	266 (93%)	253 (100%)	241 (100%)	231 (100%)	209 (100%)	744 (98%)	716 (97%)
Traditional cookstove	19 (6.8%)	18 (6.3%)					19 (2.5%)	18 (2.4%)
Other	0 (0%)	1 (0.4%)					0 (0%)	1 (0.1%)
Unknown	19	26	3	2			22	28
Exclusive Breastfeeding until 6 months								
No	259 (100%)	263 (100%)	212 (85%)	217 (90%)	189 (94%)	174 (94%)	660 (93%)	654 (95%)
Unknown	39	48	6	3	31	24	76	75
Yes			38 (15%)	23 (9.6%)	11 (5.5%)	11 (5.9%)	49 (6.9%)	34 (4.9%)
Haemophilus influenzae type b vaccine								
No	20 (6.7%)	27 (8.7%)	37 (14%)	48 (20%)	9 (3.9%)	10 (4.8%)	66 (8.4%)	85 (11%)
Yes	278 (93%)	284 (91%)	219 (86%)	195 (80%)	222 (96%)	199 (95%)	719 (92%)	678 (89%)
Pneumococcal vaccine								
No	248 (83%)	273 (88%)	254 (99%)	243 (100%)	25 (11%)	29 (14%)	527 (67%)	545 (71%)
Yes	50 (17%)	38 (12%)	2 (0.8%)	0 (0%)	206 (89%)	180 (86%)	258 (33%)	218 (29%)
Measles								
No	253 (85%)	276 (89%)	55 (21%)	64 (26%)	13 (5.6%)	19 (9.1%)	321 (41%)	359 (47%)
Yes	45 (15%)	35 (11%)	201 (79%)	179 (74%)	218 (94%)	190 (91%)	464 (59%)	404 (53%)
Illness w/cough at 24M								
No	277 (93%)	283 (91%)	247 (96%)	233 (96%)	144 (62%)	131 (63%)	668 (85%)	647 (85%)
Yes	21 (7.0%)	28 (9.0%)	9 (3.5%)	10 (4.1%)	87 (38%)	78 (37%)	117 (15%)	116 (15%)

[1] Mean (SD); n (%)

Table 4.3. Association between the prevalence of illness with a cough and type of fuel used for cooking at 24 months among children in the HAPIN follow-up cohort

Primary fuel type at 24M	Guatemala Illness with a cough			India Illness with a cough			Rwanda Illness with a cough			Pooled Illness with a cough		
	No N = 560[1]	Yes N = 49[1]	p-value[2]	No N = 480[1]	Yes N = 19[1]	p-value[2]	No N = 275[1]	Yes N = 165[1]	p-value[3]	No N = 1,315[1]	Yes N = 233[1]	p-value[2]
			>0.9			>0.9			0.7			<0.001
Wood	470 (90%)	40 (91%)		85 (18%)	3 (16%)		200 (73%)	122 (74%)		755 (59%)	165 (72%)	
Other	50 (9.6%)	4 (9.1%)		389 (82%)	16 (84%)		13 (4.7%)	10 (6.1%)		452 (36%)	30 (13%)	
LPG/electric	40	5		5	0					45	5	
Cow dung				1 (0.2%)	0 (0%)					1 (<0.1%)	0 (0%)	
Charcoal							62 (23%)	33 (20%)		62 (4.9%)	33 (14%)	

[1] *n (%)*
[2] Fisher's exact test
[3] Pearson's Chi-squared test

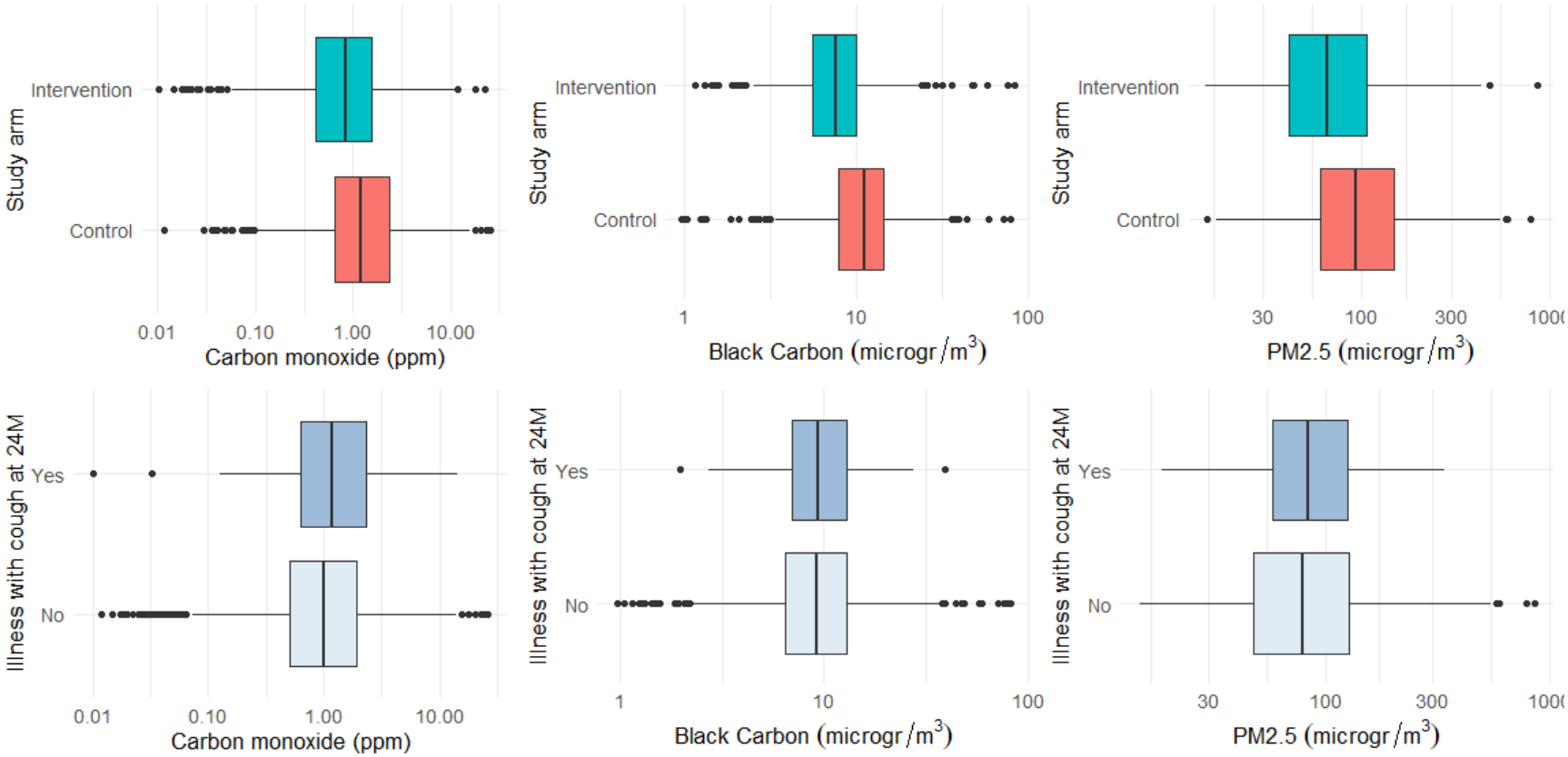


Figure 4.3. Boxplots displaying distributions of maternal prenatal pollutant levels in each study arm and according to presence of illness with a cough in HAPIN children at 24 months of age.

Table 4.4. Prevalence ratios estimated from intent-to-treat analyses for the effect of a liquefied petroleum gas stove intervention during pregnancy and the first year of life on illness with a cough at 2 years of age

Model	Prevalence ratio (95% CI)			
	pooled	Guatemala	India	Rwanda
unadjusted*	1.04 (0.84, 1.29)	1.28 (0.74, 2.20)	0.90 (0.37, 2.17)	0.99 (0.78, 1.26)
adjusted**	0.99 (0.80, 1.23)	1.18 (0.68, 2.04)	1.19 (0.49, 2.87)	0.95 (0.74, 1.22)

*adjusts for randomization strata in pooled analysis and in the analysis restricted to India
**adjusts for randomization strata, tobacco use in household at baseline, primary fuel at baseline, and food insecurity at baseline

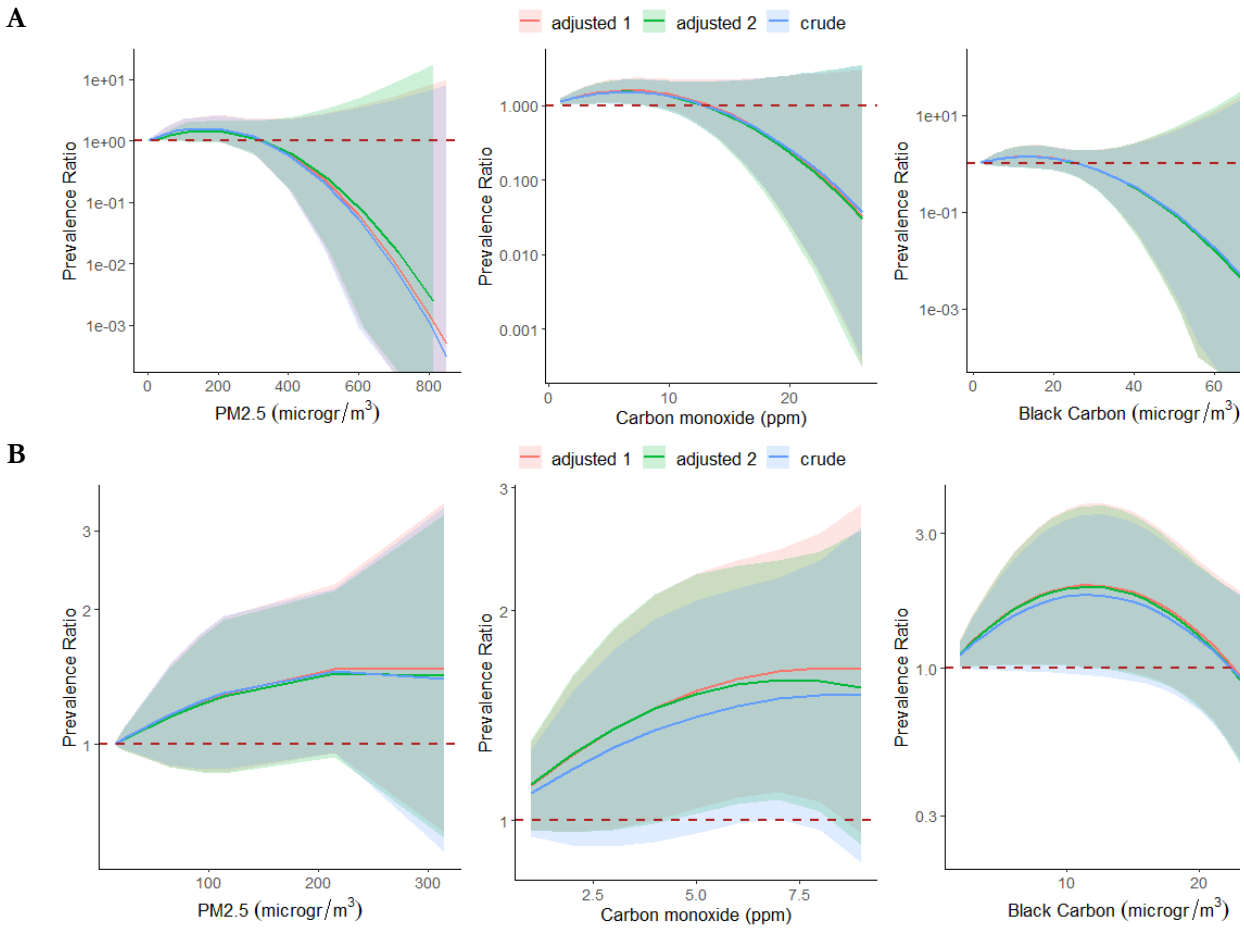


Figure 4.4. Graphs showing dose-response relationships between prenatal $PM_{2.5}$, CO, and black carbon exposure and care-giver reported illness with a cough in HAPIN children at 24 months. Crude models include the randomization strata (design element). Adjusted model 1 additionally includes season of birth, and type of fuel (solid vs. LPG/electric) used for cooking at 24 months. Adjusted model 2 additionally controls for sex, food insecurity index at 24 months (ordinal), exclusive breast feeding until 6 months of life, the number of people who sleep in the house. Panel A are results of models including all exposure data and panel B are results models with outlying observations (≥10ppm for CO, >25 μg/m3 for BC, and >300 μg/m3 for PM2.5) removed.

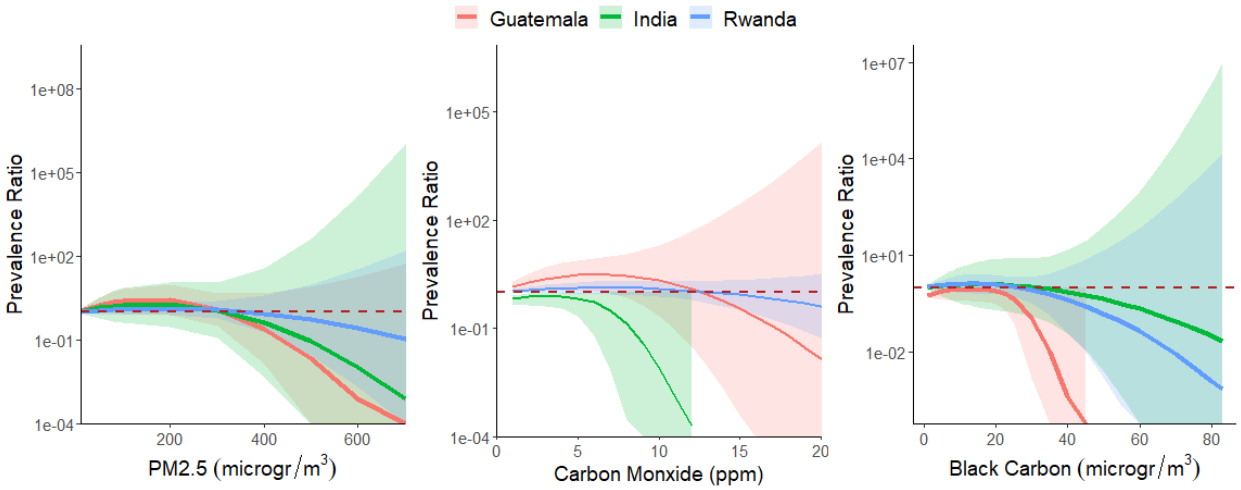


Figure 4.5. Exposure-response association between pollutant levels and the prevalence of illness with a cough at 24 months of life estimated with crude, country-specific models.
N.B. Estimates for Guatemala for black carbon and for Rwanda for carbon monoxide were generated with logistic regression because log binomial models did not converge.

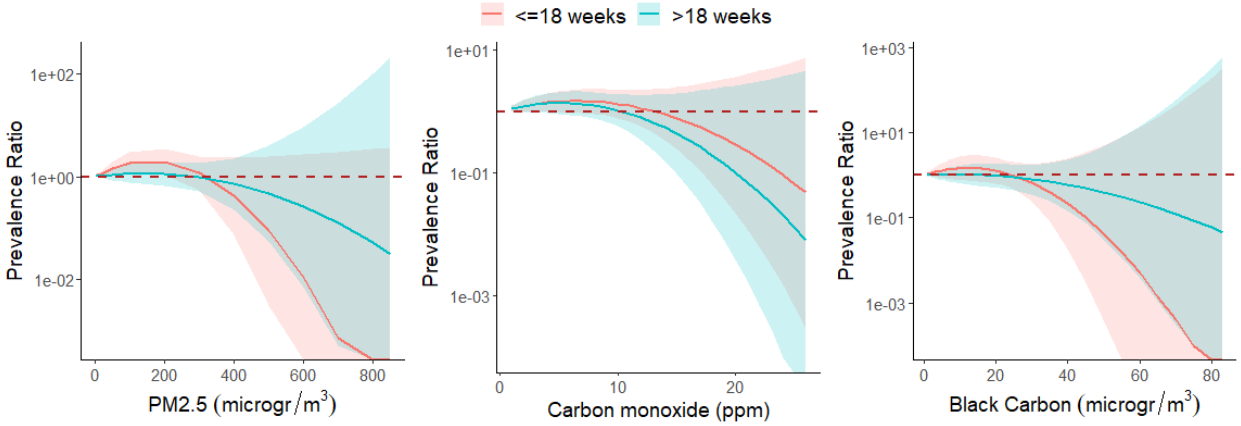


Figure 4.6. Exposure-response association between pollutant levels and the prevalence of illness with a cough at 24 months of life from models stratified by gestational age at interventions (above. vs. below median gestational age of 18 weeks as the start of intervention)

Chapter 5. Real-Time PCR quantification cycle values to estimate the population attributable fraction of *Streptococcus pneumoniae* in severe acute respiratory infections in adults in six countries: a prospective case-control study

Abstract

Background. Qualitative interpretation of quantitative RT-PCR (qPCR) results from upper respiratory tract specimens has limited utility for understanding the etiologic role of *S. pneumoniae* in WHO-defined severe acute respiratory infections (SARI) because pneumococcal carriage in the respiratory tract is common. Previous studies have shown an association between PCR quantification cycle (Cq) values (a proxy for bacterial load) and respiratory disease. We aimed to use the strength of this association to estimate the population attributable fraction (PAF) of SARI due to *S. pneumoniae* in adults.

Methods. Between 2013 and 2015, we enrolled 2,388 adults hospitalized with SARI and 1,135 frequency matched asymptomatic adults (from clinics, non-infectious disease wards, and the community) in a case-control study of SARI etiology among adults in 6 countries: Bangladesh, China, Egypt, Guatemala, Kenya, and Thailand. Nasopharyngeal and oropharyngeal specimens from study subjects were tested for 29 pathogens using the Taqman Array Card. We used logistic regression adjusting for matching factors and hypothesized confounders (age, sex, enrollment month, HIV, smoking status, underlying medical conditions, antibiotic use, and pathogen co-detections) to model the association between *S. pneumoniae* Cq values and SARI, and estimated the PAF of *S. pneumoniae*.

Results. In multivariable analyses, we found a non-linear association between Cq values and SARI that increased with increasing bacterial load (i.e. lower Cq values). The proportion of SARI cases attributed to *S. pneumoniae* varied across countries, ranging from 0.1% (95% Confidence Interval (CI): 0-1.3%) in Egypt to 18.5% (95% CI: 4.8%-35.2%) in Kenya. In Thailand, Guatemala, Bangladesh, and China, the PAFs were 12.3% (95% CI: 6.6%-19.9%), 11.3% (95% CI: 3.5%-18.1%), 10.7% (95% CI: 3.3%-19.0%), and 3.8% (95% CI: 0.1%-8.4%) respectively.

Conclusions. Compared to qualitative PCR, qPCR Cq values improved understanding of the etiologic role of *S. pneumoniae* in SARI at a population level. When more comprehensive diagnostics are not available, using qPCR can help understand the changing etiology of respiratory infections and the potential benefits gained from prevention measures.

Introduction

Globally, lower respiratory tract infections (LRI), defined as pneumonia or bronchiolitis, resulted in an estimated 2.49 million (2.27–2.74) million deaths in 2019, making them the fourth leading cause of mortality for all ages.[2] Rapid progress has been made in reducing the burden of LRI in children but parallel improvements among the oldest adults have not been observed.[2,36] Low- and middle-income countries bear a disproportionate burden of LRI, which strongly correlates with poverty.[2] Understanding the etiology of LRI and the disease burden attributed to specific pathogens (the population attributable fraction [PAF]) helps guide policymakers in planning public health interventions against respiratory pathogens.[176]

Streptococcus pneumoniae is recognized as an important cause of community-acquired pneumonia, and is the most frequently detected pathogen in upper respiratory tract specimens among adults with WHO-defined severe acute respiratory infection (SARI).[7,177] However, identifying pneumococci as

the specific etiologic agent of pneumonia is challenging because pneumococcal carriage in the respiratory tract is common and because it is difficult to sample the lung. Obtaining lower respiratory tract samples such as lung aspirates and pleural fluid is invasive and infeasible in most research settings. Instead, researchers rely on upper respiratory tract samples as a proxy for the site of infection.[45] These samples are of limited utility for understanding pneumonia etiology because bacteria frequently colonize the upper respiratory tract and commonly used diagnostics such as PCR can detect small amounts of nucleic acid. Thus, it is not possible to distinguish infection from colonization based on a qualitative (the presence or absence of a target sequence) PCR result.

An approach proposed for differentiating pathogen carriage from clinically significant infection is the quantification of pathogen load.[8,9] Studies of diarrheal disease etiology, for which the high prevalence of pathogen carriage also poses a methodological challenge, have used the association between Real-Time PCR (qPCR) quantification cycle (Cq) values (semi-quantitative measures of the amount of pathogen in a clinical specimen) and disease status to distinguish clinically relevant infection from colonization, and to estimate the attributable fractions for specific pathogens.[17-23] A similar approach could contribute to our understanding of the etiology of respiratory infections as several studies have found that higher pathogen load in the upper respiratory tract is associated with pneumonia, and for some pathogens, including *S. pneumoniae*, it is associated with more severe outcomes.[10-16,61-69] Many of these studies attempted to identify diagnostic cutoffs to distinguish colonization from clinically relevant infection in individual cases; we are not aware of any studies that used the association between bacterial load and disease to estimate population-level parameters such as the population attributable fraction.

The "Study of the etiology of community-acquired pneumonia in adults: Use of TAC multiple pathogen detection platforms in the International Emerging Infections Program sites" (hereafter

referred to as the TAC study) was conducted to understand the etiology of community-acquired pneumonia among hospitalized adults across six low- and middle-income countries in Africa, Asia, and the Americas.[7] In this study, the accurate determination of the specific causes of disease was challenging because of the high prevalence of asymptomatic carriage of some pathogens, particularly bacteria. While *S. pneumoniae* was the most frequently detected pathogen among cases of SARI, it was detected with similar frequency in asymptomatic adults, leading to the conclusion that it was not an important cause of SARI.[7] The main study analysis applied a non-parametric Bayesian regression extension of a partially latent class model approach to estimate proportions of SARI caused by specific pathogens.[7]

We aimed to determine whether an alternative approach using bacterial load (as measured by Cq values) could improve estimates of the fraction of SARI attributable to *S. pneumoniae*. Specifically, we sought to assess the association between *S. pneumoniae* Cq values and SARI and to use the strength of this association to estimate the PAF of S. *pneumoniae a*mong adults with SARI in six low- and middle-income countries.

Methods

Study population and setting

The TAC study was a prospective case-control study that used a real-time PCR-based multiple pathogen detection platform to understand the etiology of severe respiratory disease among hospitalized adults in Bangladesh, China, Egypt, Guatemala, Kenya, and Thailand. The study methods have been described previously.[7,178]

Briefly, over a 12- to 24-month period in each country, the study enrolled adults at least 18 years of age who were hospitalized with SARI, defined as an acute respiratory infection with history/measured fever ≥ 38°C and cough with onset within the last 10 days. As a comparison

group, adults without symptoms of respiratory infection in the prior week were enrolled and frequency matched to cases by age group, month of enrollment, catchment area, and HIV infection status (Kenya only). Depending on the participating country, asymptomatic adults were enrolled from the community or from non-infectious disease departments in the same facilities in which cases were enrolled.[7] Demographics, clinical data, and nasopharyngeal and oropharyngeal specimens were collected from both SARI cases and asymptomatic adults.

Laboratory procedures

Specimens were tested for 29 pathogens using Taqman® Array Card (TAC, Thermo Fisher Scientific, Carlsbad, CA) molecular assay at field laboratories in each participating study site. The Taqman Array Card is a microfluidic, multiple pathogen detection platform that uses solid-phase real-time PCR technology.[179] Bacterial load was measured by qPCR Cq values. The Cq value is the number of qPCR cycles of amplification needed for the fluorescence of a PCR target to be detected. Cq levels are inversely proportional to the amount of target nucleic acid in the sample (i.e. the lower the Cq value the greater the amount of target nucleic acid in the sample). Every 3.3 increase in the Cq value correlates with approximately 10-fold less bacterial load in the clinical specimen. Cq values greater than 45 were considered negative or non-detects. Valid results required proper functioning of controls.

Statistical analysis

We compared the frequencies of basic demographics, underlying medical conditions, antibiotic use, and pathogen detections among SARI patients and asymptomatic adults. We used chi-square and Fisher's exact tests for categorical variables and calculated odds ratios to assess the association of each pathogen detection with SARI in each country. For pathogens that were frequently detected (≥5%), including *S. pneumoniae,* we assessed whether Cq value distributions differed between SARI cases and asymptomatic adults using density plots and the Wilcoxon rank sum test. This preliminary

analysis of Cq value distributions guided subsequent modeling decisions about whether to specify co-detected pathogens as binary (detect/no detect) or continuous (Cq value).

The Cq value is undefined when the PCR test is negative. Thus, for regression analyses, we transformed Cq values such that samples in which *S. pneumoniae* was not detected were assigned a transformed Cq value of zero, the sample with the highest Cq value (i.e., smallest pathogen load) among those with detectable *S. pneumoniae* was assigned a transformed Cq value of 1, and all other samples with detectable *S. pneumoniae* were assigned a Cq value relative to the maximum Cq value using the following formula: transformed Cq value = - (Cq value – maximum Cq value – 1). Thus, the transformed Cq values represent relative measures of pathogen load, with a maximum transformed Cq value of 25.

To estimate the strength of association between *S. pneumoniae* Cq values and SARI, we fitted a multivariable logistic regression model where the outcome was disease status (SARI vs. asymptomatic) and the exposure was the transformed *S. pneumoniae* Cq value. We used a quadratic transformation of Cq values because it improved model fit as assessed by the Akaike information criterion. We also considered log and exponential transformations of Cq values but they did not improve model fit. We adjusted for matching factors[180] (age, enrollment month, and HIV), and hypothesized confounders including sex, smoking status, underlying medical conditions, antibiotic use in the 24 hours before admission, and the co-detection of viral pathogens (a composite term denoting the detection vs. no detection of any one of adenovirus; influenza A, B and C; human metapneumovirus; enterovirus/rhinovirus; human coronavirus 229E, NL64, OC43, and HKU1; parainfluenza viruses 1-4, and frequently detected bacterial pathogens (i.e. bacteria detected in ≥5% SARI cases and asymptomatic adults, specifically *Haemophilus influenzae* all types*, Klebsiella pneumoniae, Moraxella catarrhalis, Staphylococcus aureus*).[181-183] We included interaction terms between Cq value and country because it improved model fit based on the likelihood ratio test, and

country stratified models suggested heterogeneity in the association across countries. The odds ratios (OR) from this model depend on the Cq value used as the reference level and can be interpreted as the relative odds of being a SARI case vs. an asymptomatic adult for a 1 cycle increase in transformed Cq value, adjusting for all covariates.

The PAF was estimated by calculating the attributable fraction among those exposed to a given Cq value (1 – 1/OR) and summing the attributions across each of j cases with the following equation:

$$\sum_i^j AFi \qquad \text{where } AF_i = 1/j \text{ x } (1 - 1/OR_i).^{17,184}$$

The lower bound of Cq-specific odds ratios was set to one so that attributable fractions could not be negative. Bootstrapping with 1000 iterations was used to estimate 95% confidence intervals.

We performed three sensitivity analyses to understand how our modeling assumptions impacted estimates. In the first sensitivity analysis, we aimed to assess how the estimated PAFs of a conventional analysis using binary qualitative PCR results (*S. pneumoniae* positive vs. negative) compare to those from our main analysis, holding all other modeling assumptions constant. In the second sensitivity analysis, we aimed to examine how the assumption of a quadratic relationship between Cq values and SARI affected PAF estimates. In this analysis, we ran the multivariable model with a categorical exposure variable representing quartiles of Cq values, holding all other modeling assumptions constant. In the third sensitivity analysis, we aimed to examine whether a Ct value cutoff of 31 cycles would yield similar results to our main analysis. We ran the same multivariable model with categorical exposure variable representing Cq values ≤31 cycles, Cq values >31 cycles, and non-detects. A cutoff of 31 cycles was chosen because it was the quantity for which the odds ratio exceeded 1 in the multivariable model that excluded the interactions between *S. pneumoniae* Cq values and country.

Results

Between October 2013 and October 2015, 2,388 SARI patients and 1,135 asymptomatic adults were enrolled in the TAC study. Compared to asymptomatic adults, those with SARI were more likely to have underlying medical conditions including asthma (12.5% vs. 3.2%, p<0.01), chronic obstructive pulmonary disorder (6.1% vs. 0.5%, p<0.01), and diabetes (8.4% vs. 5.4%, p<0.01) (**Table 5.1**). In all countries except Kenya, SARI cases more frequently had immunosuppressive conditions (4.6% vs. 1.9%, p<0.01) and took antibiotics prior to enrollment (31.7% vs. 4.4%, p<0.01). Adults with SARI self-reported smoking less frequently than asymptomatic adults in China (24.9% vs. 45.1, p<0.01) and in Guatemala (6.2% vs. 19.0%, p<0.01), whereas smoking was higher among adults with SARI than among asymptomatic adults in Thailand (20.5% vs. 11.1%, p<0.01).

At least one pathogen was detected in specimens from 1,821 (76.3%) SARI cases and 738 (65.0%) asymptomatic adults. Codetections were frequent; two or more pathogens were detected in 944 (39.5%) and 349 (30.7%) SARI cases and asymptomatic adults, respectively. Viral detections were more frequent in SARI patients (n=1182, 49.5%) than in asymptomatic adults (n=153, 13.2%). Pathogens more frequently detected among SARI patients included influenza A (18.6% of SARI cases vs. 1.5% of asymptomatic adults); influenza B (7.7% vs. 0.3%), and respiratory syncytial virus (4.0% vs. 0.4%) (**Table 5.2**). *S. pneumoniae* and *H. influenzae* were the most commonly detected pathogens and were detected at similar frequency in SARI cases and asymptomatic adults (23.7% vs. 25.2% for *S. pneumoniae*, and 22.7% vs. 26.3% for *H. influenzae*).

The median *S. pneumoniae* Cq value was lower in SARI cases compared to asymptomatic adults (29.7 vs. 32.4, p<0.01) across all sites combined but there was heterogeneity in Cq values across countries (**Figure 5.1**). Differences of ≥2 cycles in median Cq values between SARI cases and asymptomatic adults were found in Bangladesh (28.5 vs. 30.2, p<0.01), Guatemala (30.8 vs. 35.6, p<0.01), Kenya

(28.6 vs. 32.6, p<0.01), and Thailand (29.7 vs. 32.7, p<0.01), but not in China (32.6 vs. 33.1, p=0.74) or Egypt (31.8 vs. 32.4, p=0.16). Median rhinovirus Cq values also differed between SARI cases and asymptomatic adults (26.9 vs 29.0, p<0.01). However, Cq value distributions of other frequently detected bacteria (*H. influenzae*, *K. pneumoniae*, *M. catarrhalis*, and S. *aureus*) did not differ between SARI cases and asymptomatic adults (**Figure 5.2**).

In multivariable analyses we found a quantity-dependent association between Cq values and SARI. The strength of the association increased with increasing pathogen load (i.e. lower Cq values) but there was heterogeneity across study sites (**Figure 5.3**). Associations were strongest in Thailand followed by Kenya, Guatemala, China, Bangladesh, and Egypt. However, confidence intervals around adjusted odds ratios were wide and included the null in China and Egypt and at high Cq values in other countries. The proportion of SARI cases attributed to *S. pneumoniae* varied across countries, ranging from 0.1% (95% CI: 0-1.3%) in Egypt to 18.5% (95% CI: 4.8%-35.2%) in Kenya (**Table 5.3**). In Thailand, Guatemala, Bangladesh, and China, the PAFs were 12.3% (95% CI: 6.6%-19.9%), 11.3% (95% CI: 3.5%-18.1%), 10.7% (95% CI: 3.3%-19.0%), and 3.8% (95% CI: 0.1%-8.4%), respectively.

The model that used qualitative PCR results estimated lower PAFs in all countries but China; confidence intervals were wide and included negative values. The model that used a 31-cycle cutoff for *S. pneumoniae* exposure produced similar PAFs to the main analysis (**Table 5.3**). Compared to non-detects, the odds of SARI in those with *S. pneumoniae* Cq values ≤ 31 were 1.7 (95%CI: 1.0, 2.7) times higher in Bangladesh, 2.1 (95%CI: 0.7, 6.4) times higher in China, 2.2 (95%CI: 1.1, 4.3) times higher in Kenya, 3.0 (95%CI: 1.4, 7.2) times higher in Guatemala, and 2.9 (95%CI: 1.6, 5.4) times higher in Thailand. In Egypt, the opposite association was observed but the odds ratio was not statistically significant (OR: 0.6, 95%CI: 0.3, 1.2). In all countries but China, those with *S. pneumoniae* Cq values >31 had lower odds of SARI compared to non-detects (**Figure 5.4**). The model that used

Cq quartiles as the exposure yielded PAFs that were consistent with but somewhat higher (<1.6%) than those generated using the quadratic transformation.

Discussion

In this analysis, we estimated the fraction of SARI attributable to *S. pneumoniae* in adults in six low- and middle-income countries. A conventional analysis using binary (detect vs. no detect) qualitative PCR results may underestimate the causal attribution of *S. pneumoniae* to SARI. Incorporating *S. pneumoniae* Cq values from NP/OP specimens in our modeling approach suggested a larger etiologic role of *S. pneumoniae* in SARI. The proportion of SARI that could have been prevented through the elimination of *S. pneumoniae* during the study period varied across countries, ranging from 0.1% to 18.5%.

These results demonstrate that higher *S. pneumoniae* load in the upper respiratory tract is associated with SARI, and this association is useful in understanding the etiologic role of *S. pneumoniae* in SARI on a population level. As the PAF is a function of both the prevalence of *S. pneumoniae* in SARI cases and the strength of the association between Cq values and SARI, inter-country heterogeneity in PAFs can be explained by variation in these parameters across populations. The prevalence of *S. pneumoniae* in cases ranged from 11% in China to 46% in Kenya, which had the highest estimated PAF. Heterogeneity in the odds ratios for the association between *S. pneumoniae* Cq values and SARI could be due to different characteristics of the study populations, as well as different severity thresholds for hospital admission, and different criteria for the selection of asymptomatic adults. For example, the Kenyan study site was an impoverished rural area with limited access to clean water, whereas the Chinese study site was in an urban setting with a much higher socioeconomic status. In China, asymptomatic adults were non-household members accompanying SARI patients whereas Kenya used community-based asymptomatic adults, and other sites recruited asymptomatic adults

from non-infectious disease departments. Carriage, as determined by S. *pneumoniae* prevalence in the upper respiratory tract of asymptomatic adults ranged from 7% in China to 41% in Kenya. Moreover, RSV was included in the testing panel in all sites except for China; elsewhere, high pneumococcal colonization density in the upper respiratory tract has been associated RSV coinfection.[10] Antibiotic use, which lowers bacterial density, also varied considerably across countries, and could explain some of the inter-country variability in odds ratios. In Egypt, where *S. pneumoniae* detection was associated with lower odds of SARI, 74% of SARI cases took antibiotics in the 24 hours before admission compared to 7%-30% of SARI cases in the other sites. Although we adjusted for self-reported antibiotic use prior to study enrollment, there may have been residual confounding as many cases received antibiotics after admission and self-reported antibiotic use has been shown to be an unreliable predictor of serum antimicrobial activity.[183]

The appearance of a protective effect at high Cq values might be a result of high antibiotic use in cases, which could reduce lower bacterial densities to undetectable levels. This hypothesis is supported by our finding that cases who took antibiotics prior to admission were half as likely to have a *S. pneumoniae* detection compared to cases who did not take antibiotics. Pathogen competition could also play a role in this finding. For example, production of hydrogen peroxide by pneumococci has been found to inhibit *H. influenzae* and S. *aureus* in vitro, and pneumocins have the potential to eliminate other pneumococcal strains. [185]

The analysis that used a 31-cycle *S. pneumoniae* cutoff produced similar PAFs to the main analysis. If we consider Cq values below the cutoff to be a proxy for infection, PAFs from this analysis can be interpreted as the fraction of SARI that could be eliminated by removing pneumococcal infection. While previous studies have found that cutoffs are not useful for diagnostic purposes,[10] our analysis demonstrates that they can be useful for modeling population-level parameters such as the PAF.

As has been seen in previous studies, higher density of *S. pneumoniae* in the upper respiratory tract, as indicated by lower Cq values, was associated with a higher likelihood of SARI.[10,11,62,70] While the difference in Cq values between cases and asymptomatic adults is not useful for diagnostic purposes in individual patients,[10] at a population level it can improve our understanding of the etiology of infection. Comparisons of PAFs from previous studies in other populations is challenging because estimates vary substantially due to differences in rates of asymptomatic carriage, outcome definitions, and diagnostics used. However, CDC estimates that pneumococci account for 10% to 30% of adult community-acquired pneumonia.[56] Our estimates ranged from 0.1% in Egypt to 18.5% in Kenya.

With the exception of China, PAF estimates from the main analysis using continuous Cq values were higher (and more precise) than those in an analysis that used binary, qualitative PCR results, holding all other modelling decisions constant. With the exception of Egypt, our estimates were also higher than those reported in the original analysis of these data that used a Bayesian latent class model without incorporating Cq values.[7] In that analysis, SARI cases attributed to *S. pneumoniae* ranged from 0.3% in Egypt to 4.3% in Guatemala.[7] A possible reason why the results of the original analysis were lower is that it estimated etiologic fractions for many pathogens simultaneously and restricted the etiologic fractions of all pathogens to sum to one, whereas our approach placed no such restriction on the output. Whether attributable fractions from a single model can sum to more than 100% is an area of debate. Due to pathogen interactions, we believe it is reasonable to assume that total PAFs can exceed 100%.[186,187]

The subset of data from Thailand was previously analyzed using different methods and additional assays, including a urine antigen test for *S. pneumoniae*.[178] That analysis found the prevalence of pneumococcal pneumonia to be 8% using urine antigen tests, 10% using binary PCR and urine antigen test results, and 11% by using a Bayesian latent class model with binary urine antigen test

results and Cq values. We could not include data on urine antigen test results in our analysis as it was not available in all countries. Nevertheless, we estimated a PAF of 12.3% in Thailand, which is consistent with the previous Thai estimate. In contrast, the original analysis of the TAC data estimated an etiologic fraction under 5% for Thailand.

Although this analysis includes data from six countries with systematic enrollment of SARI cases and asymptomatic adults using standardized enrollment criteria and procedures, it has several limitations. We were unable to calculate pathogen copy numbers from sample Cq values but instead had to rely on Cq values as relative measures of pathogen load. However, a study of diarrhea etiology found that analyses based on Cq values and pathogen copy numbers yield similar results.[17,23] Aspects of sample collection and specimen processing can affect Cq values, leading to non-differential exposure mismeasurement and an expectation of bias towards the null.[188] For example, all things being equal, more voluminous specimens will have lower Cq values than specimens of lower volume. The density of host DNA (as measured by human RNase P) gives some indication of specimen volume, and we found that RNase P Cq values were weakly correlated (ϱ=0.3) with *S. pneumoniae* Cq values. However, RNase P Cq values varied little between SARI cases (mean: 25.7, SD: 2.6) and asymptomatic adults (mean: 25.8, SD: 2.3). In addition to these limitations, detection of genomic material by PCR does not require viable organism and thus does not necessarily indicate the presence of infectious pathogens. Pathogen load also varies over the course of infection and the cross-sectional design of our study is a limitation in this regard.

While our data set included 3,523 subjects, this was insufficient to control for all pathogens included in the respiratory panel without introducing sparse data bias. It was also insufficient to investigate interactions between pathogens, which could play an important causal role in SARI. For example, the interaction between *S. pneumoniae* and influenza virus is well-recognized, with influenza virus infection being a risk factor for the subsequent development of pneumococcal disease.[62,189,190] Future

research is needed to address the sequence and role of multiple pathogens in SARI etiology and pathogenesis.

The results of this study should not be interpreted as the global fraction of adult SARI attributable to *S. pneumoniae* or as the etiologic fraction of *S. pneumoniae* at the national level for each participating country. The epidemiology and etiology of SARI changes over time with the introduction of vaccines and emergence of new pathogens, such as SARS-CoV-2. When this study was initiated in 2013, only Kenya and Guatemala had introduced pneumococcal conjugate vaccine (PCV), and coverage has varied over time. Bangladesh has since introduced PCV but China, Egypt, and Thailand have not.[191] As vaccine coverage increases, vaccine preventable burden declines and other pathogens not included in vaccines become more important. Moreover, the COVID-19 pandemic has dramatically altered the epidemiology and etiology of respiratory infections. Thus, population attributable fractions estimated in 2013-2015 may differ substantially from those in other time periods. Nevertheless, this approach could be applied to future studies to understand the evolving etiology of SARI.

Despite these limitations, we found that Cq values improve understanding of the etiologic role of *S. pneumoniae* in SARI. Binary results from qPCR may underestimate the causal role of *S. pneumoniae*, which would suggest that it be given lower priority for SARI prevention. However, studies using diagnostics and samples that are not typically used in routine surveillance (which often relies on qPCR) show that *S. pneumoniae* is an important contributor to SARI. When more comprehensive diagnostics are not available, estimating PAFs with PCR Cq values could be useful for translating surveillance data into numbers that can help policymakers understand the changing etiology of respiratory infections and the potential benefits gained from prevention measures such as vaccination.

Tables and Figures

Table 5.1. Basic Demographic and Underlying Medical Conditions for Asymptomatic Adults and Adults with Severe Acute Respiratory Infections (SARI) in Six Countries, 2013–2015

	Bangladesh		China		Egypt	
	SARI n=499 (no., %)	Asymptomatic n=198 (no., %)	SARI n=537 (no., %)	Asymptomatic n=216 (no., %)	SARI n=504 (no., %)	Asymptomatic n=209 (no., %)
Age (years)						
18-49	251 (50.3)	100 (50.5)	414 (77.1)	166 (76.9)	283 (56.2)	113 (54.1)
50-64	132 (26.5)	52 (26.3)	78 (14.5)	31 (14.4)	170 (33.7)	70 (33.5)
65+	116 (23.3)	46 (23.2)	45 (8.4)	19 (8.8)	51 (10.1)	26 (12.4)
Male	321 (64.3)	134 (67.7)	310 (57.8)	111 (51.9)	240 (47.6)	118 (56.5)
Medical History[1]						
Current smoker	97 (19.4)	45 (22.7)	131 (24.9)	97 (45.1)	111 (22.0)	53 (25.4)
HIV/AIDS	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Asthma	122 (24.5)	0 (0)	2 (0.4)	0 (0)	29 (5.8)	2 (1.0)
COPD	61 (12.2)	2 (1.0)	12 (2.2)	3 (1.4)	8 (1.6)	0 (0)
Diabetes	25 (5.0)	2 (1.0)	6 (1.1)	4 (1.9)	48 (9.5)	11 (5.3)
Immunosuppression[2]	65 (13.0)	0 (0)	0 (0)	0 (0)	2 (0)	0 (0)
Antibiotic use prior to hospital admission	114 (22.8)	3 (1.5)	160 (29.8)	0 (0)	372 (73.8)	0 (0)

	Guatemala		Kenya		Thailand	
	SARI n=304 (no., %)	Asymptomatic n=174 (no., %)	SARI n=187 (no., %)	Asymptomatic n=121 (no., %)	SARI n=357 (no., %)	Asymptomatic n=217 (no., %)
Age (years)						
18-49	118 (38.7)	65 (37.4)	111 (59.4)	84 (69.4)	79 (22.1)	53 (24.4)
50-64	84 (27.6)	49 (28.2)	40 (21.4)	20 (16.5)	108 (30.3)	69 (31.8)
65+	102 (33.4)	60 (34.5)	36 (19.3)	17 (14.1)	170 (47.6)	95 (43.8)
Male	98 (32.2)	90 (51.7)	60 (32.1)	22 (18.2)	161 (45.1)	80 (36.9)
Medical History[1]						
Current smoker	19 (6.2)	33 (19.0)	14 (7.5)	5 (4.1)	73 (20.5)	24 (11.1)
HIV/AIDS	6 (2.0)	0 (0)	74 (39.6)	70 (57.9)	6 (1.7)	0 (0)
Asthma	81 (26.6)	2 (1.2)	10 (5.4)	0 (0)	55 (15.4)	6 (2.8)
COPD	14 (4.6)	0 (0)	3 (1.6)	0 (0)	47 (13.2)	1 (0.5)
Diabetes	53 (17.4)	18 (10.3)	5 (2.7)	0 (0)	63 (17.7)	26 (12.0)
Immunosuppression[2]	19 (6.3)	0 (0)	5 (2.7)	19 (15.7)	18 (5.0)	3 (1.4)
Antibiotic use prior to hospital admission	54 (17.8)	8 (4.6)	32 (17.1)	37 (30.6)	26 (7.3)	2 (0.9)

Abbreviations: no., number

Self-reported medical history with the exception of HIV/AIDS in Kenya

[2] Immunosuppression includes receiving chemotherapy, having a documented autoimmune disease, and using oral or injection corticosteroids for at least 14 days within 2 weeks of enrollment.

NB. Percentage is out of those without missing data.

Table 5.2. Pathogen detections in Asymptomatic Adults and Adults with Severe Acute Respiratory Infections (SARI) in Six Countries, 2013–2015

	Bangladesh					China					Egypt				
	SARI (n=499)		Asymptomatic (n=198)		OR (95%CI)	SARI (n=537)		Asymptomatic (n=216)		OR (95%CI)	SARI (n=504)		Asymptomatic (n=209)		OR (95%CI)
	n	(%)	n	(%)		n	(%)	n	(%)		n	(%)	n	(%)	
Adenovirus	5	(1%)	1	(1%)	2.0 (0.2-94.7)	20	(4%)	9	(4%)	0.9 (0.4-2)	2	(0%)	2	(1%)	0.4 (0.03-5.7)
Bordetella pertussis	0	(0%)	0	(0%)		1	(0%)	1	(0%)	0.4 (0.01-31.8)	1	(0%)	0	(0%)	
Chlamydophila pneumoniae	1	(0%)	0	(0%)		13	(2%)	6	(3%)	0.9 (0.3-2.3)	1	(0%)	0	(0%)	
Influenza A	66	(13%)	8	(4%)	3.6 (1.7-7.7)	207	(39%)	7	(3%)	18.8 (8.7-40.8)	83	(16%)	0	(0%)	
Influenza B	27	(5%)	0	(0%)		34	(6%)	0	(0%)		74	(15%)	1	(0%)	35.9 (45.0-259.9)
Influenza C	2	(0%)	0	(0%)		2	(0%)	3	(1%)	0.3 (0.02-2.4)	4	(1%)	0	(0%)	
Group A *Streptococcus*	5	(1%)	3	(2%)	0.7 (0.16-2.8)	10	(2%)	2	(1%)	2.0 (0.4-9.4)	9	(2%)	5	(2%)	0.7 (0.3-2.2)
Coronavirus 229E	3	(1%)	0	(0%)		7	(1%)	1	(0%)	2.9 (0.4-23.4)	3	(1%)	1	(0%)	1.3 (0.1-66.6)
Coronavirus NL63	2	(0%)	1	(1%)	0.8 (0.04-47.0)	3	(1%)	4	(2%)	0.3 (0.1-1.4)	3	(1%)	1	(0%)	1.2 (0.1-65.8)
Coronavirus OC43	9	(2%)	1	(1%)	3.6 (0.5-28.8)	17	(3%)	5	(2%)	1.4 (0.5-3.8)	7	(1%)	4	(2%)	0.7 (0.2-2.5)
Coronavirus HKU1	3	(1%)	0	(0%)		2	(0%)	3	(1%)	0.3 (0.02-2.4)	1	(0%)	1	(0%)	0.4 (0.01-32.7)
Haemophilus influenzae	128	(26%)	44	(22%)	1.2 (0.8-1.8)	92	(17%)	46	(21%)	0.8 (0.5-1.1)	91	(18%)	62	(30%)	0.5 (0.4-0.8)
Human metapneumovirus	1	(0%)	1	(1%)	0.4 (0.01-31.2)	15	(3%)	2	(1%)	3.1 (0.7-13.6)	10	(2%)	1	(0%)	4.2 (0.5-33.2)
Klebsiella pneumoniae	67	(13%)	15	(8%)	1.9 (1.0-3.4)	19	(4%)	9	(4%)	0.9 (0.4-1.9)	30	(6%)	9	(4%)	1.4 (0.7-3.0)
Legionella species	0	(0%)	0	(0%)		1	(0%)	0	(0%)		0	(0%)	0	(0%)	
Moraxella catarrhalis	84	(17%)	36	(18%)	0.9 (0.6-1.4)	27	(5%)	13	(6%)	0.8 (0.4-1.6)	50	(10%)	48	(23%)	0.4 (0.2-0.6)
Mycoplasma pneumoniae	1	(0%)	1	(1%)	0.4 (0.01-31.3)	10	(2%)	1	(0%)	4.1 (0.5-32.3)	2	(0%)	0	(0%)	
Mycobacterium tuberculosis	3	(1%)	0	(0%)		0	(0%)	0	(0%)		0	(0%)	0	(0%)	
Human parainfluenza virus 1	5	(1%)	0	(0%)		2	(0%)	0	(0%)		0	(0%)	0	(0%)	
Human parainfluenza virus 2	0	(0%)	0	(0%)		5	(1%)	0	(0%)	2.0 (0.2-96.5)	0	(0%)	0	(0%)	
Human parainfluenza virus 3	12	(2%)	1	(1%)	4.9 (0.6-37.6)	3	(1%)	2	(1%)	0.6 (0.1-7.3)	9	(2%)	0	(0%)	
Human parainfluenza virus 4	6	(1%)	0	(0%)		0	(0%)	1	(0%)		1	(0%)	0	(0%)	
Pneumocystis jiroveci (PCP)	0	(0%)	0	(0%)		1	(0%)	0	(0%)		1	(0%)	0	(0%)	
Pseudomonas aeruginosa	24	(5%)	4	(2%)	2.5 (0.8-7.2)	8	(1%)	1	(0%)	3.3 (0.4-26.3)	0	(0%)	0	(0%)	
Respiratory syncytial virus	17	(3%)	2	(1%)	3.5 (0.8-15.1)	NA	NA	NA	NA	NA	8	(2%)	0	(0%)	
Staphylococcus aureus	62	(12%)	22	(11%)	1.1 (0.7-1.9)	19	(4%)	7	(3%)	1.1 (0.5-2.7)	44	(9%)	25	(12%)	0.7 (0.4-1.2)
Streptococcus pneumoniae	165	(33%)	57	(29%)	1.2 (0.9-1.8)	59	(11%)	16	(7%)	1.6 (0.9-2.8)	65	(13%)	61	(29%)	0.4 (0.2-0.5)
Rhinovirus/Enterovirus	82	(16%)	13	(7%)	2.8 (1.5-5.2)	71	(13%)	23	(11%)	1.3 (0.8-2.2)	47	(9%)	18	(9%)	1.1 (0.6-1.9)

	Guatemala					Kenya					Thailand				
	SARI (n=304)		Asymptomatic (n=174)		OR (95%CI)	SARI (n=187)		Asymptomatic (n=121)		OR (95%CI)	SARI (n=357)		Asymptomatic (n=217)		OR (95%CI)
	n	(%)	n	(%)		n	(%)	n	(%)		n	(%)	n	(%)	
Adenovirus	3	(1%)	0	(0%)		3	(2%)	5	(4%)	0.4 (0.1-2.0)	1	(0%)	5	(2%)	0.1 (0.01-1.0)
Bordetella pertussis	0	(0%)	1	(1%)		0	(0%)	0	(0%)		1	(0%)	1	(0%)	0.6 (0.01-47.8)
Chlamydophila pneumoniae	2	(1%)	0	(0%)		0	(0%)	0	(0%)		1	(0%)	1	(0%)	0.6 (0.01-47.8)
Influenza A	30	(10%)	1	(1%)	19.1 (2.6-141.2)	20	(11%)	1	(1%)	14.5 (1.9-109.2)	37	(10%)	0	(0%)	
Influenza B	5	(2%)	1	(1%)	2.9 (0.3-138.5)	0	(0%)	0	(0%)		44	(12%)	1	(0%)	30.4 (4.2-222.1)
Influenza C	0	(0%)	0	(0%)		0	(0%)	0	(0%)		0	(0%)	0	(0%)	
Group A *Streptococcus*	9	(3%)	6	(3%)	0.9 (0.3-2.5)	2	(1%)	0	(0%)		4	(1%)	2	(1%)	1.2 (0.2-13.6)
Coronavirus 229E	4	(1%)	0	(0%)		0	(0%)	0	(0%)		1	(0%)	1	(0%)	0.6 (0.01-47.8)
Coronavirus NL63	0	(0%)	0	(0%)		3	(2%)	1	(1%)	2.0 (0.2-104.1)	2	(1%)	0	(0%)	
Coronavirus OC43	2	(1%)	0	(0%)		5	(3%)	2	(2%)	1.6 (0.3-16.9)	0	(0%)	0	(0%)	
Coronavirus HKU1	3	(1%)	2	(1%)	0.9 (0.1-10.4)	0	(0%)	0	(0%)		2	(1%)	1	(0%)	1.2 (0.1-72.1)
Haemophilus influenzae	60	(20%)	49	(28%)	0.6 (0.4-0.99)	36	(19%)	20	(17%)	1.2 (0.7-2.2)	133	(37%)	77	(35%)	1.1 (0.8-1.5)
Human metapneumovirus	9	(3%)	0	(0%)		4	(2%)	0	(0%)		2	(1%)	0	(0%)	
Klebsiella pneumoniae	17	(6%)	18	(10%)	0.5 (0.3-1.0)	16	(9%)	6	(5%)	1.9 (0.7-4.9)	71	(20%)	48	(22%)	0.9 (0.6-1.3)
Legionella species	0	(0%)	1	(1%)		0	(0%)	0	(0%)		1	(0%)	0	(0%)	
Moraxella catarrhalis	38	(13%)	19	(11%)	1.2 (0.7-2.1)	25	(13%)	12	(10%)	1.4 (0.7-2.9)	45	(13%)	13	(6%)	2.3 (1.2-4.3)
Mycoplasma pneumoniae	4	(1%)	0	(0%)		0	(0%)	0	(0%)		2	(1%)	0	(0%)	
Mycobacterium tuberculosis	1	(0%)	0	(0%)		0	(0%)	0	(0%)		1	(0%)	0	(0%)	
Human parainfluenza virus 1	1	(0%)	1	(1%)	0.6 (0.01-45.4)	3	(2%)	0	(0%)		5	(1%)	1	(0%)	3.1 (0.3-145.7)
Human parainfluenza virus 2	1	(0%)	0	(0%)		2	(1%)	0	(0%)		3	(1%)	0	(0%)	
Human parainfluenza virus 3	8	(3%)	0	(0%)		1	(1%)	0	(0%)		10	(3%)	0	(0%)	
Human parainfluenza virus 4	3	(1%)	0	(0%)		1	(1%)	0	(0%)		2	(1%)	0	(0%)	
Pneumocystis jiroveci (PCP)	2	(1%)	0	(0%)		2	(1%)	0	(0%)		4	(1%)	0	(0%)	
Pseudomonas aeruginosa	13	(4%)	1	(1%)	7.8 (1.0-60.0)	10	(5%)	10	(8%)	0.7 (0.3-1.6)	19	(5%)	15	(7%)	0.8 (0.4-1.5)
Respiratory syncytial virus	13	(4%)	0	(0%)		3	(2%)	0	(0%)		32	(9%)	2	(1%)	10.6 (2.5-44.6)
Staphylococcus aureus	36	(12%)	24	(14%)	0.9 (0.5-1.5)	37	(20%)	17	(14%)	1.5 (0.8-2.8)	27	(8%)	25	(12%)	0.6 (0.4-1.1)
Streptococcus pneumoniae	75	(25%)	45	(26%)	1.0 (0.6-1.5)	86	(46%)	50	(41%)	1.2 (0.8-2.0)	114	(32%)	57	(26%)	1.3 (0.9-1.9)
Rhinovirus/Enterovirus	69	(23%)	4	(2%)	15.8 (5.6-44.1)	31	(17%)	11	(9%)	2.2 (1.0-4.5)	38	(11%)	5	(2%)	5.1 (2.0-13.0)

Abbreviations: OR, odds ratio; CI, confidence interval

Figure 5.1. PCR quantification cycle value distributions of *S. pneumoniae* in asymptomatic adults and adults with severe acute respiratory infection in six low- and middle-income countries, 2013-2015.

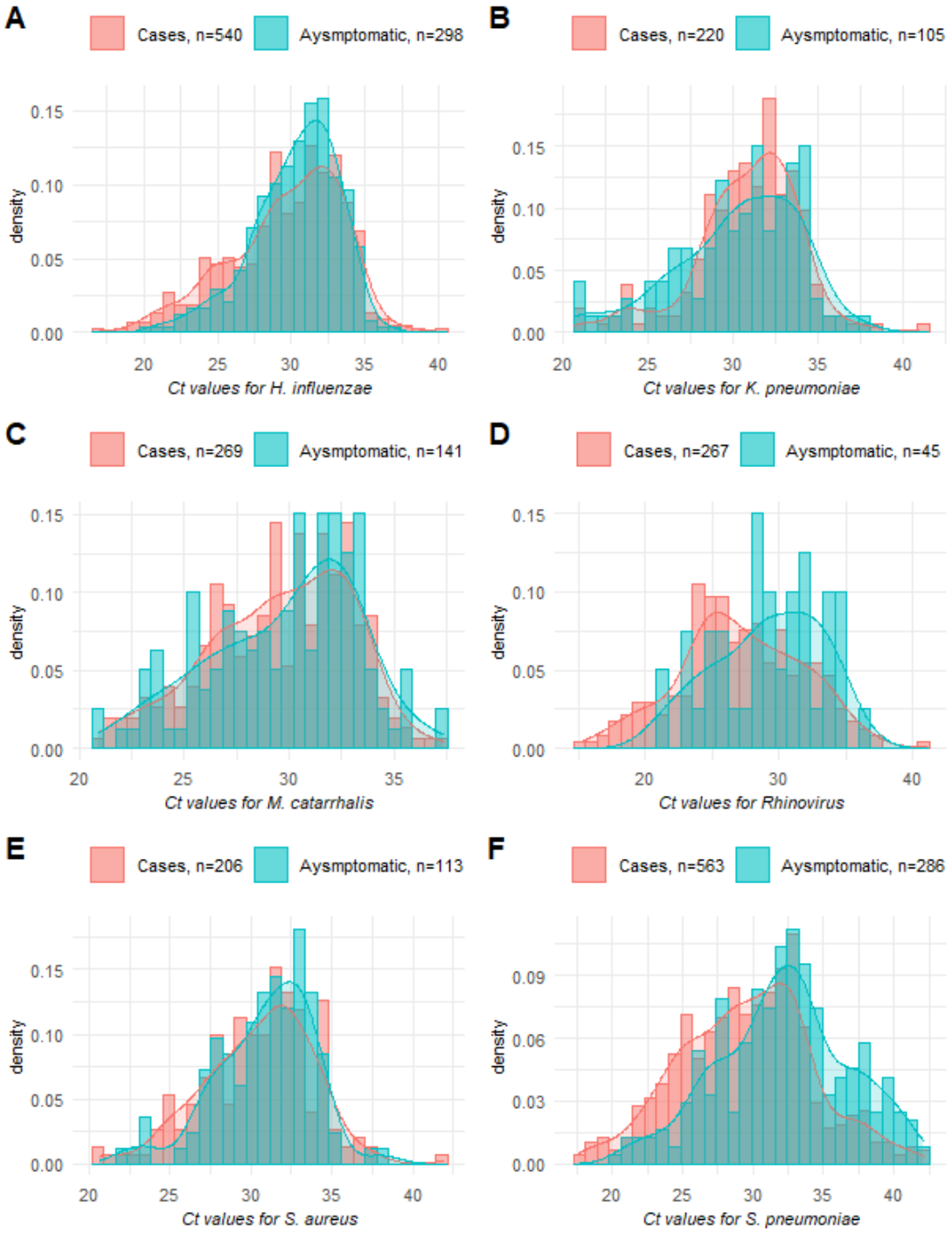


Figure 5.2. PCR quantification cycle value distributions of frequently detected pathogens in asymptomatic adults and adults with severe acute respiratory infection in six low- and middle-income countries, 2013-2015.

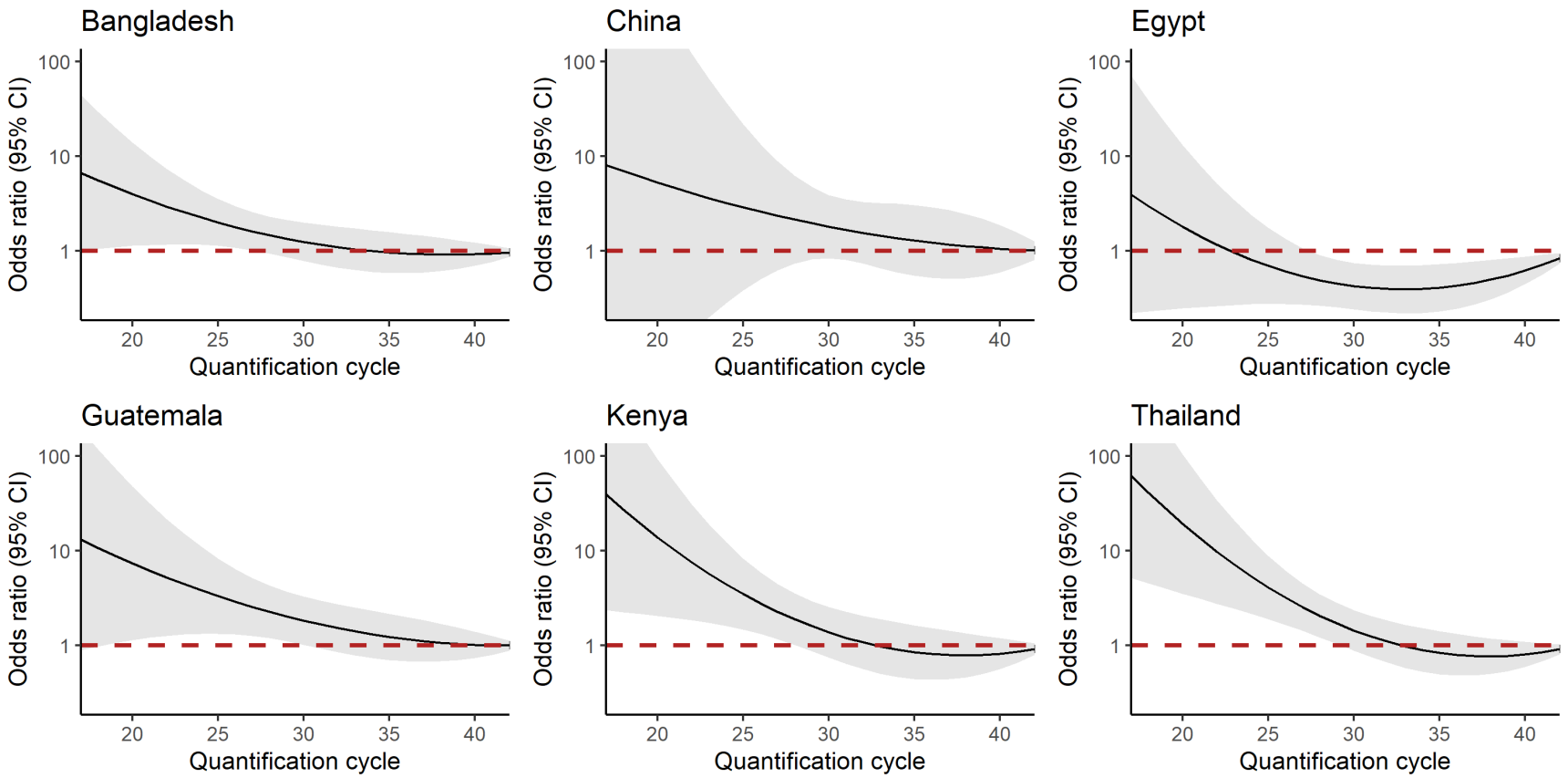


Figure 5.3. Relationship between pathogen quantity and severe acute respiratory infection in adults in six low- and middle-income countries, 2013-2015.

The plots display odds ratios (black lines) and 95% confidence intervals (gray bands) on $\log_{10}$ scale. The red dotted line represents an odds ratio of 1.

Table 5.3. Population Attributable Fraction of *S. Pneumoniae* to Severe Acute Respiratory Infections in Adults in Six Low- and Middle-Income Countries, 2013-2015

Country	PAF model 1 (95% CI)	PAF model 2 (95% CI)	PAF model 3 (95% CI)	PAF model 4 (95% CI)
Bangladesh	10.7% (3.3, 19.0)	9.7 % (1.0, 16.3)	8.9% (-2.4, 18.1)	11.2% (3.1, 19.1)
China	3.8% (0.1, 8.4)	2.1% (-2.1, 4.4)	4.5% (-1.8, 9.2)	5.0% (2.3, 10.7)
Egypt	0.1% (0.0, 1.3)	-4.9% (-12.8, 0.2)	-17.6% (-31.8, -7.6)	0.0% (0.0, 1.9)
Guatemala	11.3% (3.5, 18.1)	8.5% (2.2, 13.6)	4.5% (-7.5, 13.4)	12.4% (5.3, 18.7)
Kenya	18.5% (4.8, 35.2)	16.9 (-1.2, 29.4)	12.2% (-20.0, 32.1)	20.1% (5.0, 34.8)
Thailand	12.3% (6.6, 19.9)	13.0 (6.9, 18.7)	8.7% (-3.1, 18.5)	13.8% (8.2, 20.5)

PAF: population attributable fraction, CI: confidence interval
Model 1 is the primary model that defines *S. pneumoniae* exposure using Cq values as a continuous variable. Model 2 includes all subjects but defines the exposure as a categorical variable representing *S. pneumoniae* Cq values ≤31 vs. non-detects. Model 3 defines the exposure using binary, qualitative PCR results (*S. pneumoniae* positive vs. *S. pneumoniae* negative) and includes the same covariates as Model 1. Model 4 defines *S. pneumoniae* exposure using a categorical variable based on Cq value quartiles.

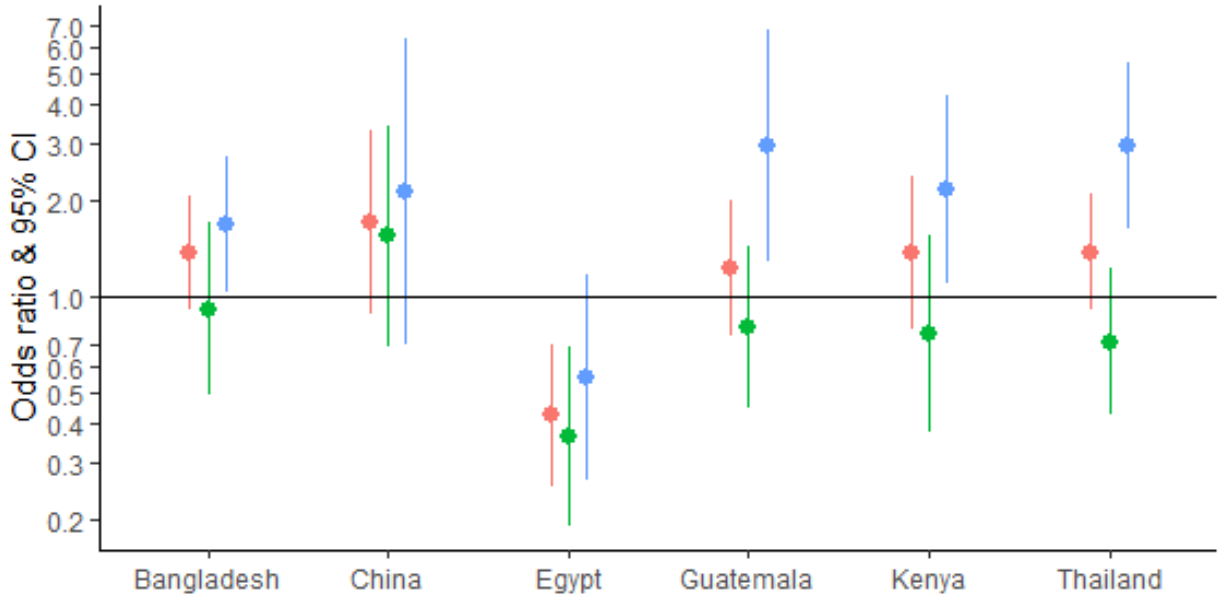


Figure 5.4. Odds ratios for the relationship between *S. pneumoniae* and severe acute respiratory infection in adults in six low- and middle-income countries, 2013-2015.
The plots display odds ratios (dots) and 95% confidence intervals (lines). The reference group for all ORs is adults with a negative PCR result for *S. pneumoniae*. Red represents is the odds of SARI in those with a positive *S. pneumoniae* PCR result, green is the odds of SARI in those with a Cq value >31 cycles, and blue is the odds of SARI in adults with Cq values ≤31 cycles.

Supplementary figures

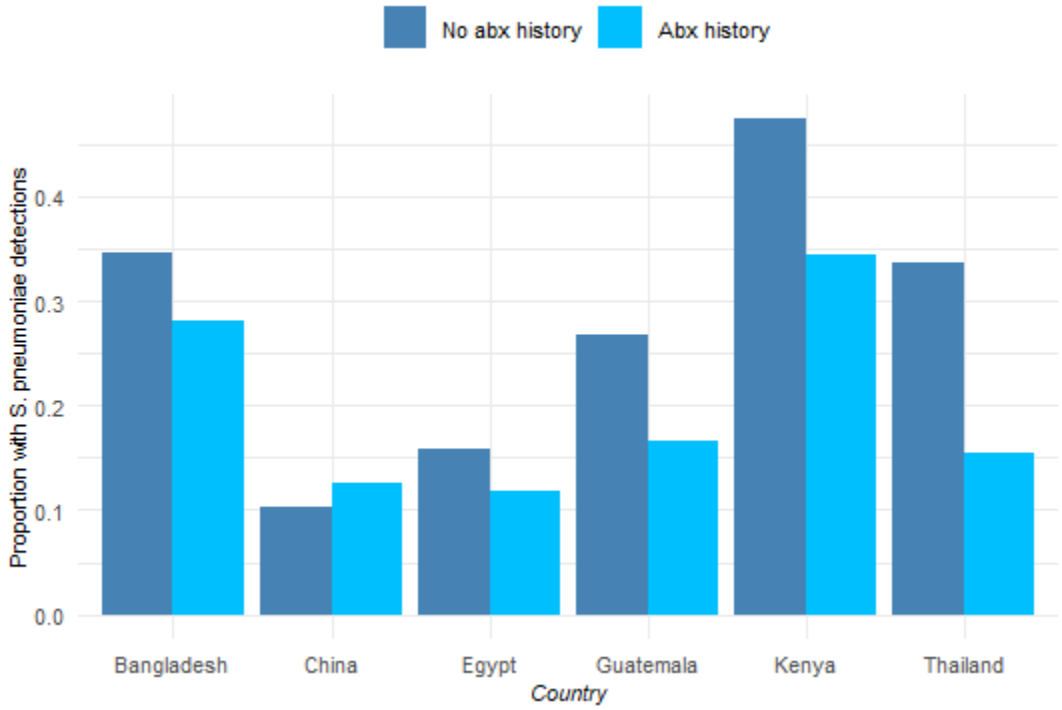


Figure 5.5. Proportion of SARI cases with S. pneumoniae detections by history of antibiotic use and country

The odds of *S. pneumoniae* detection in *SARI cases* who had taken antibiotics before hospitalization are 0.72 (95% CI: 0.56, 0.93) times that of cases who did not take antibiotics before admission, adjusted for country.

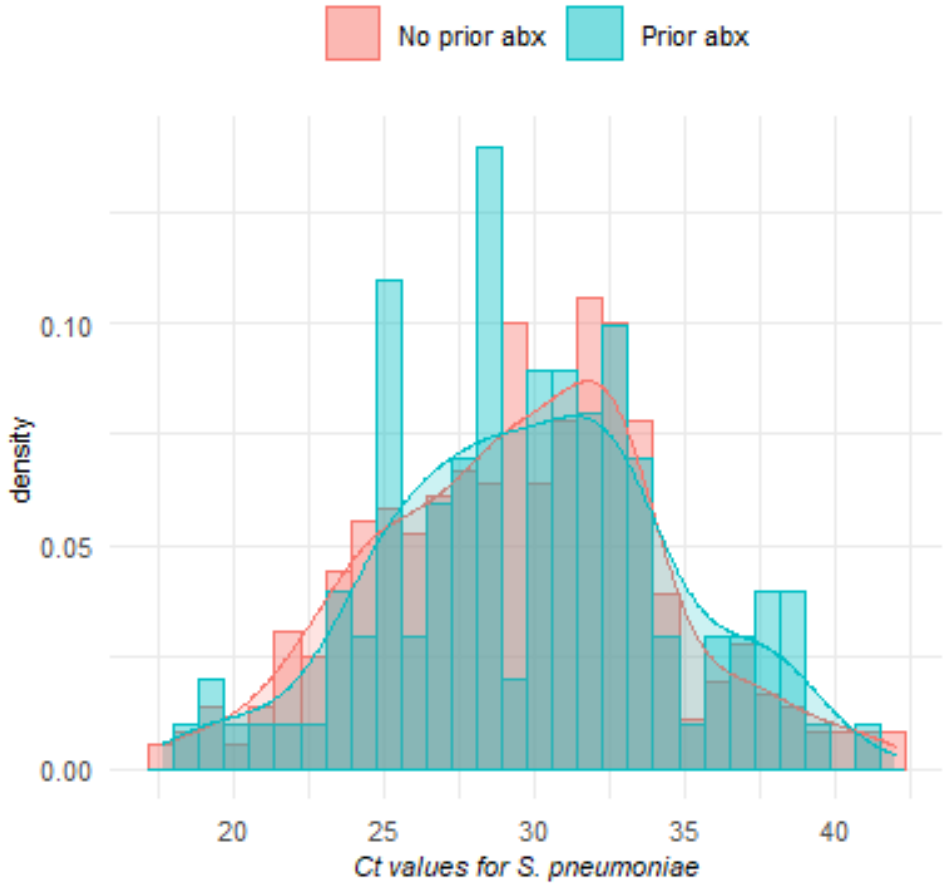


Figure 5.6. Distributions of *S. pneumoniae* Ct values in SARI cases who took antibiotics in the 24 hours prior to hospital admission and those who did not.

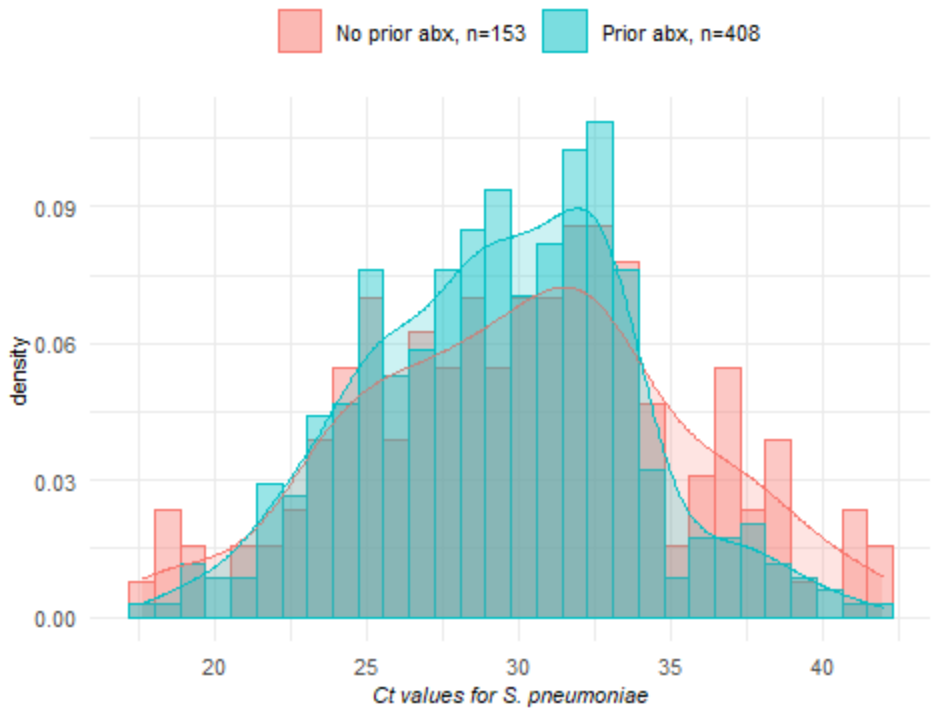


Figure 5.7. Distributions of *S. pneumoniae* Ct values in SARI cases who took antibiotics prior to specimen collection and those who did not.
N.B. This plot is for SARI cases only. Very few (n=50) controls took antibiotics prior to enrollment and the majority were in Kenya so the association between antibiotic use and Ct distribution in controls could not be assessed.

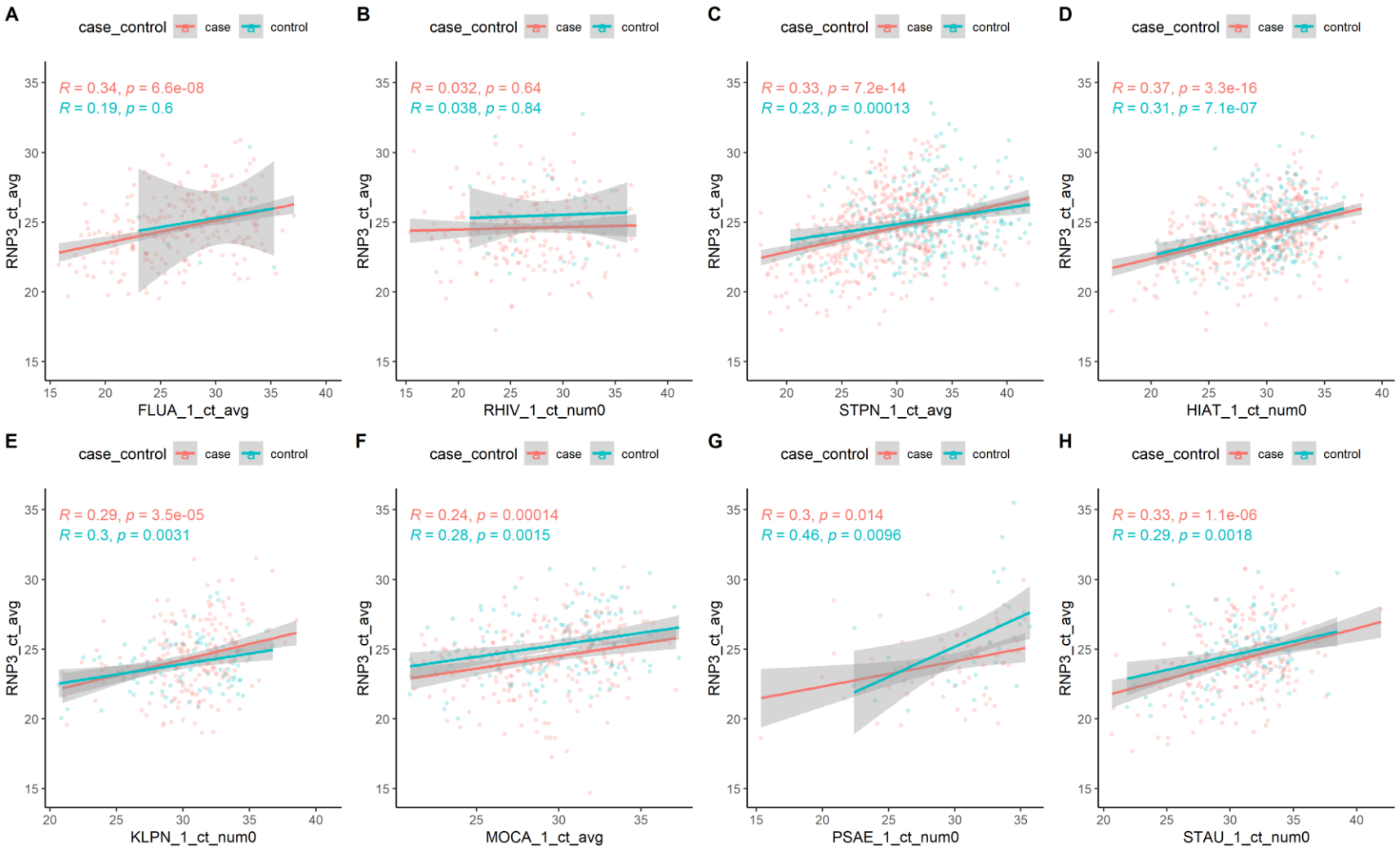


Figure 5.8. Correlation between pathogen-specific Ct values and RNase P Ct values.
Ct values are weakly to moderately correlated with Ct values of RNase P. In other words, lower Ct values are found in specimens of higher quality.

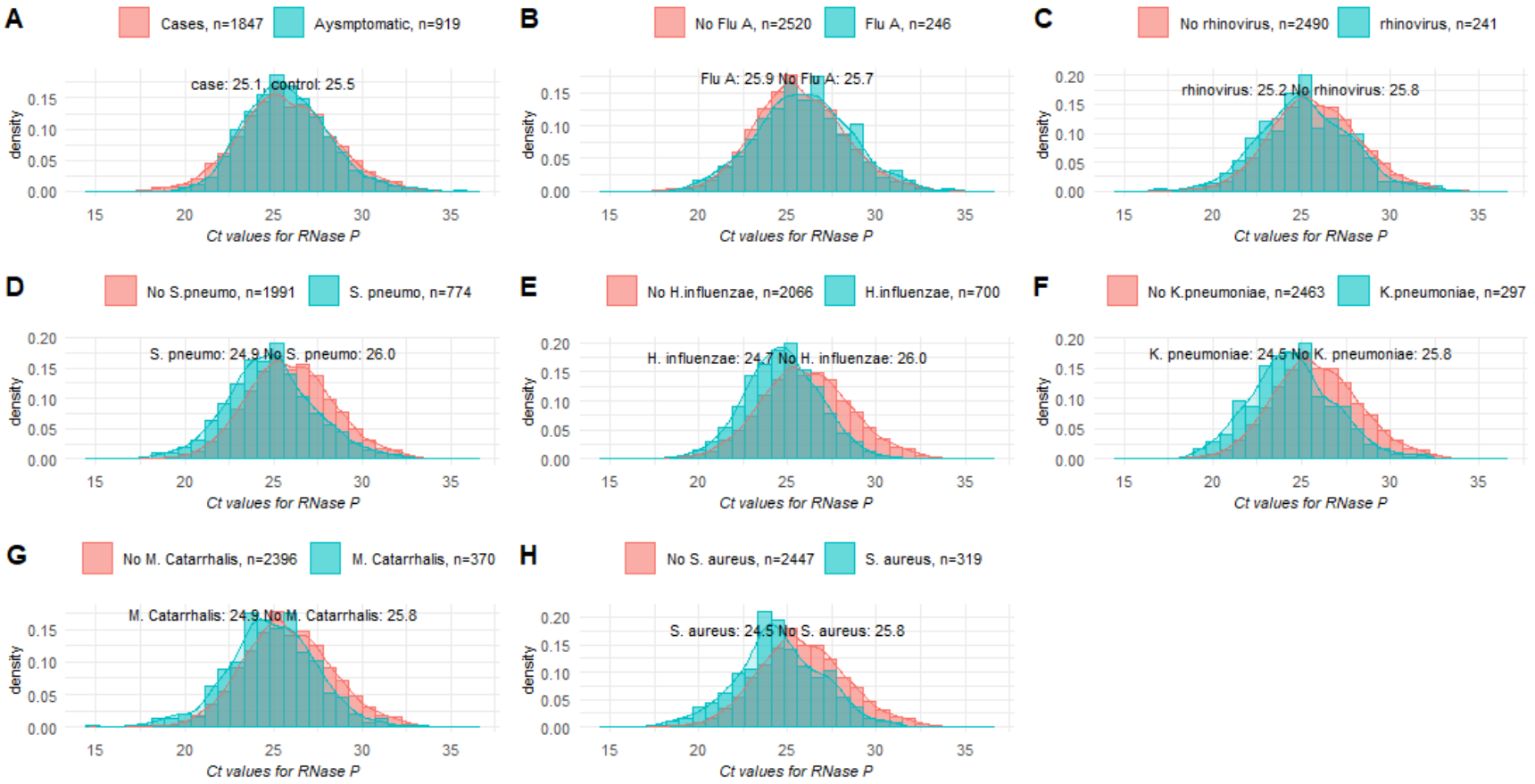


Figure 5.9. Distributions of Ct values for RNase P in cases and controls, and in specimens with and without detections of specific pathogens.

There is a statistically significant difference in the distributions of RNase P Ct values in SARI cases and asymptomatic adults but the difference in means between the two groups is small. RNase P Ct values are lower in subjects with bacterial pathogen detections than in those without bacterial pathogen detections, suggesting that specimen quality may have affected the probability of detecting pathogens in NP/OP swabs. The difference in RNase P Ct values in subjects with and without influenza A virus (which is the agent most strongly associated with SARI) is small and not statistically significant.

Chapter 6. Summary and conclusions

Summary of findings

Acute respiratory infections exert a considerable burden and cost on health care systems.[1,2] At the severe end of the disease severity pyramid, lower respiratory tract infections are the fourth leading cause of mortality for all ages and the second leading cause of death among children younger than 5 years worldwide.[2] At the mild end of the disease severity pyramid, upper respiratory infections occur with high frequency and significantly impair quality of life and productivity. In 2019, there were an estimated 17.2 billion incident cases of upper respiratory tract infections worldwide, contributing to 9,460 deaths and 6.39 million DALYS.[2] Low- and middle-income countries bear a disproportionate burden of acute respiratory infections, which strongly correlate with poverty.

Intervening to reduce this burden requires better understanding of the causes and risk factors for acute respiratory infections as well as identification of effective interventions. In order to prioritize pathogens for vaccine development and to guide national vaccine policy, it is important to estimate the proportion of disease attributable to specific etiologic agents. Once vaccines are developed, effectively deploying them and assessing their impact requires knowledge of pathogen seasonality. Reducing the burden of ARI also requires identifying effective interventions to reduce environmental factors, such as HAP, that affect the infectious agent and host defenses. This dissertation addressed knowledge gaps surrounding the etiology, environmental risk factors, and patterns of acute respiratory infections in low- and middle-income settings by (1) characterizing RSV seasonality in Guatemala, (2) examining the association between prenatal HAP exposure and illness with a cough in 24-month old children, and (3) estimating the fraction of SARI attributable to *S. pneumoniae.*

In chapter 3 (aim 1), we described RSV seasonality in Guatemala using nine consecutive years of surveillance data. To our knowledge, this is the longest time series of RSV data for which epidemiologic analyses have been reported for the country. We identified key attributes of RSV seasons including the onset week, offset week, epidemic duration, and epidemic threshold–the level of virus activity that signals the onset of a seasonal epidemic. Our results demonstrate considerable variability in the timing of seasonal RSV epidemics over seasons, and some variability across regions. Season onsets varied up to 5 months such that two differential patterns of RSV seasonality were observed: an early season starting in June-July and finishing in September-November, and a late season starting in October-November and finishing in March-April. This variability speaks to the importance of continually monitoring RSV seasonality and to the difficulty in precisely predicting the timing of seasonal RSV epidemics based on onset weeks from past seasons. These finding suggest that maximal reduction in RSV disease burden would be achieved through year-round maternal and infant vaccination programs.

In chapter 4, we examined the association of prenatal HAP exposure with illness with a cough in 24-month-old children enrolled in the HAPIN trial in three low- and middle-income countries. We did not find evidence that the HAPIN LPG stove and fuel intervention during pregnancy and the first year of life had an impact on the prevalence of illness with a cough in children at 24 months of life. Similarly, in exposure-response analyses, we did not find evidence of an association between three important HAP pollutants (carbon monoxide, $PM_{2.5}$, and black carbon) and the prevalence of illness with a cough at 24 months. These findings suggest that the reductions in prenatal HAP exposure that can be achieved through a clean fuel intervention may not have a measurable long-term impact on upper respiratory tract infections in young children.

In chapter 5, we used qPCR results from upper respiratory tract specimens to understand the etiologic role of *S. pneumoniae* in WHO-defined SARI in adults in six low- and middle-income countries. We found a non-linear association between Cq values and SARI that increased with increasing bacterial load (i.e. lower Cq values). The proportion of SARI cases attributed to *S. pneumoniae* varied across countries, ranging from 0.1% (95% CI: 0-1.3%) in Egypt to 18.5% (95% CI: 4.8%-35.2%) in Kenya. For most countries, population attributable fractions estimated using quantitative PCR Cq values were higher than those estimated using qualitative interpretation of PCR results. These results suggest that Cq values can improve understanding of the etiologic role of *S. pneumoniae* in SARI at a population level.

Limitations

As discussed in previous chapters, the dissertation studies had several limitations. Below, we elaborate on key limitations.

Aim 1

Although the descriptive analysis of RSV seasonality in chapter 3 included nine consecutive years of surveillance data, which exceeds that of previous RSV reports from Guatemala, a longer time series would have allowed us to more comprehensively assess changes in seasonality over time as well as drivers of epidemic onsets. For example, approaches to investigate the association of the timing of onset or peaks of epidemics with weather conditions require many epidemic episodes to achieve adequate statistical power, since each epidemic episode is effectively one data point.[192] We also lacked data on RSV antigenic groups, so we could not explore the potential the impact of RSV groups on seasonality.

The methods we used to describe RSV seasonality were developed for establishing alert thresholds for influenza. They assume one epidemic wave per season and some consistency in seasonal patterns

as is typical of influenza waves in temperate areas. They have limitations when seasonality is highly variable. For example, the methods require that surveillance seasons first be defined. We defined surveillance seasons in Guatemala as epidemiologic week 17 to week 16, following average troughs in RSV activity. However, the 2012-13 epidemic wave was delayed, and in Santa Rosa, it continued into the 2013/14 surveillance season. The MEM classified the epidemic offset of the 2012/13 season in Santa Rosa as week 16 (the last week of the surveillance season) even though visual inspection of the epidemic curves shows that the wave continued into the next surveillance season.

Aim 2

In chapter 4 (aim 2), we used a well-characterized study cohort with prospectively collected individual- and household-level data. Nevertheless, the study has important limitations. Perhaps the most important limitation is the outcome definition: caregiver-reported illness with a cough. Previous studies suggest that HAP exposure has the greatest effect on severe respiratory outcomes in children. However, we were not powered to detect severe outcomes in this cohort. A second limitation is that the exposure contrast in the study, though strong, might not have been sufficient to have an impact on respiratory health, particularly in the context of high ambient pollution. That said, exposures in the control arm were lower than those reported in some earlier studies, such as RESPIRE, which reported an average maternal carbon monoxide level of 4.8 ppm in the control arm compared to 2.0 ppm in control participants in the HAPIN exit cohort.[114]

Aim 3

In Chapter 5 (aim 3), we used the association between PCR Cq values and SARI to estimate the proportion of SARI attributable to *S. pneumoniae* in adults. Although this analysis used data from a case-control study that included 3,523 subjects from six countries with systematic enrollment of SARI cases and asymptomatic adults using standardized enrollment criteria and procedures, there

were several important limitations. A key limitation relates to the interpretation of the population attributable fraction (PAF).

Strictly speaking, the PAF can be interpreted as the proportion of disease cases over a specified time that would be prevented following elimination of the exposures, while distributions of other risk factors in the population remain unchanged, assuming the exposures are causal. It is of public health relevance when interpreted in terms of the impact of potential interventions. In the context of our study, in which the reference category for models was those with no detection of *S. pneumoniae* in upper respiratory tract samples, the estimated PAF is the proportion of SARI in adults that could be eliminated if *S. pneumoniae* were removed from the upper respiratory tract. However, an intervention that completely eliminates *S. pneumoniae* carriage in the upper respiratory tract may be unrealistic; thus, our PAF estimate may have more of a theoretical than a practical value. Furthermore, the PAF is a relative measure of disease burden; absolute measures may be of more relevance to public health and to policy makers who need to know the number of cases of disease that could be averted through interventions such as vaccination. Other study designs, such as vaccine probe studies, are better suited to estimate the absolute burden of disease incidence that could be prevented through vaccination.

Another important limitation of this study was the high prevalence of antibiotic use in some study countries, particularly Egypt. Antibiotics use potentially introduces reverse causality into the analysis because they are taken as a result of the illness (the outcome) and they lower bacterial load (the exposure). Adjustment for antibiotic use cannot address this issue and may be inappropriate because antibiotic use is a descendent of the outcome (SARI). Across the study countries, 15% to 98% of the SARI cases took antibiotics prior to specimen collection and 7% to 75% took antibiotics in the 24 hours before hospital admission. Restricting the analysis to those who did not take antibiotics would

have substantially reduced the sample size and would have dropped Egypt (where antibiotic usage was highest) from the analysis altogether.

As this was a secondary analysis, it was not powered to examine pathogen interactions or to perform sub-analyses by disease severity. Interactions between pathogens likely play an important role in SARI etiology, and it is recognized that viral infections predispose the host to *S. pneumoniae* infection such that the two jointly cause disease. More comprehensively studying the role of prior viral infections in bacterial pneumonia would require a longitudinal study design.

We hypothesized that the proportion of disease attributable to S. pneumoniae increases with disease severity, but we were not powered to perform sub-analyses by severity. In particular, had the sample size been larger with complete data on chest x-ray findings, we would have estimated the PAF for radiographically confirmed pneumonia. This would have allowed us to compare our estimates with those of vaccine probe studies, in which radiographically confirmed pneumonia is the endpoint typically used.

As is often the case in case-control studies, control selection might have introduced selection bias into our analysis. Each study country used different types of controls (e.g. hospital controls in some sites and community controls in others) and it was difficult to determine if these controls were representative of the source population that gave rise to the cases. Had there been different types of controls in each country, we may have been better able to assess the role of control selection in the PAF estimates.

Future directions

In examining the patterns, etiology, and environmental risk factors for acute respiratory infections, this dissertation highlighted future research areas and generated additional hypotheses surrounding the drivers of acute respiratory infections. Examining the causes and drivers of RSV seasonality was

beyond the scope of aim 1, but it remains an important question. There is a large body of previous research examining the association between weather patterns and respiratory virus seasonality. While climate correlates with respiratory virus seasonality, the COVID-19 pandemic has shown that human behavior and contact patterns may be more important, underlying drivers of seasonal patterns.

In aim 2, we examined the association of prenatal HAP exposure with illness with a cough in children at 24 months of life. In future analyses we plan to estimate the effect of post-natal HAP exposure on the prevalence of illness with cough in children at 24 months of age using additional exposure and outcome data from the HAPIN trial and exit cohort study. Future work might also explore the reasons why the prevalence of illness with a cough in Rwanda was so much higher than in Guatemala and India.

References

1. Jin X, Ren J, Li R, et al. Global burden of upper respiratory infections in 204 countries and territories, from 1990 to 2019. *EClinicalMedicine.* 2021;37:100986.
2. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet.* 2020;396(10258):1204-1222.
3. Mulholland KW, Martin W ; . *Pneumonia in Children: Epidemiology, Prevention and Treatment.* 1st ed: Pinter & Martin Ltd; 2016.
4. WHO/The United Nations Children's Fund. *End preventable deaths: Global Action Plan for Prevention and Control of Pneumonia and Diarrhoea.* Geneva: World Health Organization;2013.
5. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2018;392(10159):1736-1788.
6. O'Brien KL, Baggett HC, Brooks WA, et al. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *The Lancet.* 2019;394(10200):757-779.
7. Milucky J, Pondo T, Gregory CJ, et al. The epidemiology and estimated etiology of pathogens detected from the upper respiratory tract of adults with severe acute respiratory infections in multiple countries, 2014–2015. *PloS one.* 2020;15(10):e0240309.
8. Klugman KP, Madhi SA, Albrich WC. Novel approaches to the identification of Streptococcus pneumoniae as the cause of community-acquired pneumonia. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2008;47 Suppl 3:S202-206.

9. Hammitt LL, Feikin DR, Scott JAG, et al. Addressing the Analytic Challenges of Cross-Sectional Pediatric Pneumonia Etiology Data. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;64(suppl_3):S197-s204.

10. Baggett HC, Watson NL, Deloria Knoll M, et al. Density of Upper Respiratory Colonization With Streptococcus pneumoniae and Its Role in the Diagnosis of Pneumococcal Pneumonia Among Children Aged <5 Years in the PERCH Study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;64(suppl_3):S317-s327.

11. Albrich WC, Madhi SA, Adrian PV, et al. Use of a rapid test of pneumococcal colonization density to diagnose pneumococcal pneumonia. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2012;54(5):601-609.

12. Fuller JA, Njenga MK, Bigogo G, et al. Association of the CT values of real-time PCR of viral upper respiratory tract infection with clinical severity, Kenya. *Journal of medical virology.* 2013;85(5):924-932.

13. Utokaparch S, Marchant D, Gosselink JV, et al. The relationship between respiratory viral loads and diagnosis in children presenting to a pediatric hospital emergency department. *The Pediatric infectious disease journal.* 2011;30(2):e18-23.

14. Buckingham SC, Bush AJ, Devincenzo JP. Nasal quantity of respiratory syncytical virus correlates with disease severity in hospitalized infants. *The Pediatric infectious disease journal.* 2000;19(2):113-117.

15. DeVincenzo JP, El Saleeby CM, Bush AJ. Respiratory syncytial virus load predicts disease severity in previously healthy infants. *The Journal of infectious diseases.* 2005;191(11):1861-1868.

16. DeVincenzo JP, Wilkinson T, Vaishnaw A, et al. Viral load drives disease in humans experimentally infected with respiratory syncytial virus. *American journal of respiratory and critical care medicine.* 2010;182(10):1305-1314.

17. Liu J, Platts-Mills JA, Juma J, et al. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *Lancet.* 2016;388(10051):1291-1301.

18. Barletta F, Ochoa TJ, Mercado E, et al. Quantitative real-time polymerase chain reaction for enteropathogenic Escherichia coli: a tool for investigation of asymptomatic versus symptomatic infections. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2011;53(12):1223-1229.

19. Lindsay B, Ochieng JB, Ikumapayi UN, et al. Quantitative PCR for detection of Shigella improves ascertainment of Shigella burden in children with moderate-to-severe diarrhea in low-income countries. *Journal of clinical microbiology.* 2013;51(6):1740-1746.

20. Phillips G, Tam CC, Conti S, et al. Community incidence of norovirus-associated infectious intestinal disease in England: improved estimates using viral load for norovirus diagnosis. *American journal of epidemiology.* 2010;171(9):1014-1022.

21. Mukhopadhya I, Sarkar R, Menon VK, et al. Rotavirus shedding in symptomatic and asymptomatic children using reverse transcription-quantitative PCR. *Journal of medical virology.* 2013;85(9):1661-1668.

22. Kaplon J, Fremy C, Pillet S, et al. Diagnostic Accuracy of Seven Commercial Assays for Rapid Detection of Group A Rotavirus Antigens. *Journal of clinical microbiology.* 2015;53(11):3670-3673.

23. Platts-Mills JA, Liu J, Rogawski ET, et al. Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: a reanalysis of the MAL-ED cohort study. *The Lancet Global health.* 2018;6(12):e1309-e1318.

24. Smith KR, Bruce N, Balakrishnan K, et al. Millions dead: how do we know and what does it mean? Methods used in the comparative risk assessment of household air pollution. *Annual review of public health.* 2014;35:185-206.

25. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012;380(9859):2224-2260.

26. Berry CE, Billheimer D, Jenkins IC, et al. A Distinct Low Lung Function Trajectory from Childhood to the Fourth Decade of Life. *American journal of respiratory and critical care medicine.* 2016;194(5):607-612.

27. Dherani M, Pope D, Mascarenhas M, Smith KR, Weber M, Bruce N. Indoor air pollution from unprocessed solid fuel use and pneumonia risk in children aged under five years: a systematic review and meta-analysis. *Bull World Health Organ.* 2008;86(5):390-398c.

28. Goodman D, Crocker ME, Pervaiz F, et al. Challenges in the diagnosis of paediatric pneumonia in intervention field trials: recommendations from a pneumonia field trial working group. *The Lancet Respiratory medicine.* 2019;7(12):1068-1083.

29. Clasen T, Checkley W, Peel JL, et al. Design and Rationale of the HAPIN Study: A Multicountry Randomized Controlled Trial to Assess the Effect of Liquefied Petroleum Gas Stove and Continuous Fuel Distribution. *Environmental health perspectives.* 2020;128(4):47008.

30. Zhou Y, Yang G-D, Feng K, et al. [Clinical features and chest CT findings of coronavirus disease 2019 in infants and young children]. *Zhongguo Dang Dai Er Ke Za Zhi.* 2020.

31. Human S, Moore ML. How close are we to a respiratory syncytial virus vaccine? *Future Virology.* 2016;11(12):749-752.

32. World Health Organization. *WHO Product Development for Vaccines Advisory Committee (PDVAC) meeting - 2019.* World Health Organization; 2019-07-22 16:16:35 2019.

33. Janet S, Broad J, Snape MD. Respiratory syncytial virus seasonality and its implications on prevention strategies. *Human vaccines & immunotherapeutics.* 2018;14(1):234-244.

34. Li Y, Reeves RM, Wang X, et al. Global patterns in monthly activity of influenza virus, respiratory syncytial virus, parainfluenza virus, and metapneumovirus: a systematic analysis. *The Lancet Global health.* 2019;7(8):e1031-e1045.

35. World Health Organization. RSV Vaccine Research and Development Technology Roadmap. Priority activities for development, testing, licensure and global use of RSV vaccines, with a specific focus on the medical need for young children in low- and middle-income countries. 2017:12.

36. 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;18(11):1191-1210.

37. World Health Organization. RSV surveillance case definitions. 2022; https://www.who.int/teams/global-influenza-programme/global-respiratory-syncytial-virus-surveillance/case-definitions, 2022.

38. Breiman RF, Van Beneden CA, Farnon EC. Surveillance for respiratory infections in low- and middle-income countries: experience from the Centers for Disease Control and Prevention's Global Disease Detection International Emerging Infections Program. *The Journal of infectious diseases.* 2013;208 Suppl 3:S167-172.

39. Fitzner J, Qasmieh S, Mounts AW, et al. Revision of clinical case definitions: influenza-like illness and severe acute respiratory infection. *Bull World Health Organ.* 2018;96(2):122-128.

40. Institute for Health Metrics and Evaluation (IHME). GBD 2019 Cause and Risk Summary: Lower Respiratory infections. https://www.healthdata.org/results/gbd_summaries/2019, 2022.

41. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. 2020; http://vizhub.healthdata.org/gbd-compare, 2022.

42. Quantifying risks and interventions that have affected the burden of lower respiratory infections among children younger than 5 years: an analysis for the Global Burden of Disease Study 2017. *The Lancet Infectious diseases.* 2020;20(1):60-79.

43. Almirall J, Serra-Prat M, Bolíbar I, Balasso V. Risk Factors for Community-Acquired Pneumonia in Adults: A Systematic Review of Observational Studies. *Respiration.* 2017;94(3):299-311.

44. Klugman KP, Rodgers GL. PERCH in Perspective: What Can It Teach Us About Pneumonia Etiology in Children? *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;64(suppl_3):S185-s187.

45. O'Brien KL, Baggett HC, Brooks WA, et al. Introduction to the Epidemiologic Considerations, Analytic Methods, and Foundational Results From the Pneumonia Etiology Research for Child Health Study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;64(suppl_3):S179-s184.

46. Self WH, Williams DJ, Zhu Y, et al. Respiratory Viral Detection in Children and Adults: Comparing Asymptomatic Controls and Patients With Community-Acquired Pneumonia. *The Journal of infectious diseases.* 2015;213(4):584-591.

47. Johansson N, Kalin M, Tiveljung-Lindell A, Giske CG, Hedlund J. Etiology of community-acquired pneumonia: increased microbiological yield with new diagnostic methods. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2010;50(2):202-209.

48. Jennings LC, Anderson TP, Beynon KA, et al. Incidence and characteristics of viral community-acquired pneumonia in adults. *Thorax.* 2008;63(1):42-48.

49. Angeles Marcos M, Camps M, Pumarola T, et al. The role of viruses in the aetiology of community-acquired pneumonia in adults. *Antiviral therapy.* 2006;11(3):351-359.

50. McCracken JP, Smith KR, Diaz A, Mittleman MA, Schwartz J. Chimney stove intervention to reduce long-term wood smoke exposure lowers blood pressure among Guatemalan women. *Environmental health perspectives.* 2007;115(7):996-1001.

51. Deloria Knoll M, Fu W, Shi Q, et al. Bayesian Estimation of Pneumonia Etiology: Epidemiologic Considerations and Applications to the Pneumonia Etiology Research for Child Health Study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;64(suppl_3):S213-s227.

52. Wahl B, O'Brien KL, Greenbaum A, et al. Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000-15. *The Lancet Global health.* 2018;6(7):e744-e757.

53. Feikin DR, Scott JA, Gessner BD. Use of vaccines as probes to define disease burden. *Lancet.* 2014;383(9930):1762-1770.

54. Cutts FT, Zaman SM, Enwere G, et al. Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial. *Lancet.* 2005;365(9465):1139-1146.

55. Jain S, Self WH, Wunderink RG, et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N Engl J Med.* 2015;373(5):415-427.

56. Centers for Disease Control and Prevention. Pinkbook: Pneumococcal Disease. In: Hall E. WAP, Hamborsky J., et al.,, ed. *Epidemiology and Prevention of Vaccine-Preventable Diseases.* 14th ed. Washington, D.C.: Public Health Foundation; 2021.

57. Gessner BD, Isturiz R, Snow V, Grant LR, Theilacker C, Jodar L. The rationale for use of clinically defined outcomes in assessing the impact of pneumococcal conjugate vaccines against pneumonia. *Expert review of vaccines.* 2021;20(3):269-280.

58. Thorpe JE, Baughman RP, Frame PT, Wesseler TA, Staneck JL. Bronchoalveolar Lavage for Diagnosing Acute Bacterial Pneumonia. *The Journal of infectious diseases.* 1987;155(5):855-861.

59. Kralik P, Ricchi M. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. *Frontiers in microbiology.* 2017;8:108.

60. Njenga MK, Paweska J, Wanjala R, et al. Using a field quantitative real-time PCR test to rapidly identify highly viremic rift valley fever cases. *Journal of clinical microbiology.* 2009;47(4):1166-1171.

61. Stralin K, Herrmann B, Abdeldaim G, Olcen P, Holmberg H, Molling P. Comparison of sputum and nasopharyngeal aspirate samples and of the PCR gene targets lytA and Spn9802 for quantitative PCR for rapid detection of pneumococcal pneumonia. *Journal of clinical microbiology.* 2014;52(1):83-89.

62. Wolter N, Tempia S, Cohen C, et al. High nasopharyngeal pneumococcal density, increased by viral coinfection, is associated with invasive pneumococcal pneumonia. *The Journal of infectious diseases.* 2014;210(10):1649-1657.

63. Tenenbaum T, Franz A, Neuhausen N, et al. Clinical characteristics of children with lower respiratory tract infections are dependent on the carriage of specific pathogens in the nasopharynx. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology.* 2012;31(11):3173-3182.

64. Spuesens EB, Fraaij PL, Visser EG, et al. Carriage of Mycoplasma pneumoniae in the upper respiratory tract of symptomatic and asymptomatic children: an observational study. *PLoS medicine.* 2013;10(5):e1001444.

65. Anh DD, Huong Ple T, Watanabe K, et al. Increased rates of intense nasopharyngeal bacterial colonization of Vietnamese children with radiological pneumonia. *The Tohoku journal of experimental medicine.* 2007;213(2):167-172.

66. Vu HT, Yoshida LM, Suzuki M, et al. Association between nasopharyngeal load of Streptococcus pneumoniae, viral coinfection, and radiologically confirmed pneumonia in Vietnamese children. *The Pediatric infectious disease journal.* 2011;30(1):11-18.

67. Esposito S, Zampiero A, Terranova L, et al. Pneumococcal bacterial load colonization as a marker of mixed infection in children with alveolar community-acquired pneumonia and respiratory syncytial virus or rhinovirus infection. *The Pediatric infectious disease journal.* 2013;32(11):1199-1204.

68. Maillet M, Maubon D, Brion JP, et al. Pneumocystis jirovecii (Pj) quantitative PCR to differentiate Pj pneumonia from Pj colonization in immunocompromised patients. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology.* 2014;33(3):331-336.

69. Muhlethaler K, Bogli-Stuber K, Wasmer S, et al. Quantitative PCR to diagnose Pneumocystis pneumonia in immunocompromised non-HIV patients. *The European respiratory journal.* 2012;39(4):971-978.

70. Park DE, Baggett HC, Howie SRC, et al. Colonization Density of the Upper Respiratory Tract as a Predictor of Pneumonia-Haemophilus influenzae, Moraxella catarrhalis, Staphylococcus aureus, and Pneumocystis jirovecii. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;64(suppl_3):S328-s336.

71. Heymann DL. *Control of communicable diseases manual: An official report of the American Public Health Association.* 20 ed. Washington, D.C.2015.

72. Nair H, Nokes DJ, Gessner BD, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *Lancet.* 2010;375(9725):1545-1555.

73. Shi T, Denouel A, Tietjen AK, et al. Global Disease Burden Estimates of Respiratory Syncytial Virus-Associated Acute Respiratory Infection in Older Adults in 2015: A Systematic Review and Meta-Analysis. *The Journal of infectious diseases.* 2019.

74. Giersing BK, Modjarrad K, Kaslow DC, Moorthy VS, Committee WHOPDfVA, Committee WHOPDfVPDA. Report from the World Health Organization's Product Development for Vaccines Advisory Committee (PDVAC) meeting, Geneva, 7-9th Sep 2015. *Vaccine.* 2016;34(26):2865-2869.

75. Plotkin SA, Orenstein WA, Offit PA. *Plotkin's vaccines.* Seventh edition.. ed. Philadelphia, PA: Philadelphia, PA : Elsevier; 2018.

76. Midgley CM, Haynes AK, Baumgardner JL, et al. Determining the Seasonality of Respiratory Syncytial Virus in the United States: The Impact of Increased Molecular Testing. *The Journal of infectious diseases.* 2017;216(3):345-355.

77. Cromer D, van Hoek AJ, Newall AT, Pollard AJ, Jit M. Burden of paediatric respiratory syncytial virus disease and potential effect of different immunisation strategies: a modelling and cost-effectiveness analysis for England. *The Lancet Public health.* 2017;2(8):e367-e374.

78. Kim L, Rha B, Abramson JS, et al. Identifying Gaps in Respiratory Syncytial Virus Disease Epidemiology in the United States Prior to the Introduction of Vaccines. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;65(6):1020-1025.

79. Pitzer VE, Viboud C, Alonso WJ, et al. Environmental drivers of the spatiotemporal dynamics of respiratory syncytial virus in the United States. *PLoS pathogens.* 2015;11(1):e1004591.

80. Obando-Pacheco P, Justicia-Grande AJ, Rivero-Calle I, et al. Respiratory Syncytial Virus Seasonality: A Global Overview. *The Journal of infectious diseases.* 2018;217(9):1356-1364.

81. Freitas AR, Donalisio MR. Respiratory syncytial virus seasonality in Brazil: implications for the immunisation policy for at-risk populations. *Mem Inst Oswaldo Cruz.* 2016;111(5):294-301.

82. Donaldson GC. Climate change and the end of the respiratory syncytial virus season. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2006;42(5):677-679.

83. Paiva TM, Ishida MA, Benega MA, et al. Shift in the timing of respiratory syncytial virus circulation in a subtropical megalopolis: implications for immunoprophylaxis. *Journal of medical virology.* 2012;84(11):1825-1830.

84. Kovacs SD, Mullholland K, Bosch J, et al. Deconstructing the differences: a comparison of GBD 2010 and CHERG's approach to estimating the mortality burden of diarrhea, pneumonia, and their etiologies. *BMC infectious diseases.* 2015;15:16.

85. McAllister DA, Liu L, Shi T, et al. Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis. *The Lancet Global health.* 2019;7(1):e47-e57.

86. Rudan I, O'Brien KL, Nair H, et al. Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries. *J Glob Health.* 2013;3(1):010401.

87. Kirolos A, Christides A, Xian S, Reeves R, Nair H, Campbell H. A landscape review of the published research output relating to respiratory syncytial virus (RSV) in North & Central America and Europe between 2011-2015. *Journal of Global Health.* 2019;9(1):010425.

88. Bruce N, Weber M, Arana B, et al. Pneumonia case-finding in the RESPIRE Guatemala indoor air pollution trial: standardizing methods for resource-poor settings. *Bull World Health Organ.* 2007;85(7):535-544.

89. Laguna-Torres VA, Sanchez-Largaespada JF, Lorenzana I, et al. Influenza and other respiratory viruses in three Central American countries. *Influenza and other respiratory viruses.* 2011;5(2):123-134.

90. Schlaudecker EP, Heck JP, Macintyre ET, et al. Etiology and seasonality of viral respiratory infections in rural Honduran children. *The Pediatric infectious disease journal.* 2012;31(11):1113-1118.

91. Fischer Langley G, McCracken J Fau - Arvelo W, Arvelo W Fau - Estevez A, et al. The epidemiology and clinical characteristics of young children hospitalized with respiratory syncytial virus infections in Guatemala (2007-2010). (1532-0987 (Electronic)).

92. Verani JR, McCracken J, Arvelo W, et al. Surveillance for hospitalized acute respiratory infection in Guatemala. *PloS one.* 2013;8(12):e83600.

93. McCracken JP, Prill MM, Arvelo W, et al. Respiratory syncytial virus infection in Guatemala, 2007-2012. *The Journal of infectious diseases.* 2013;208 Suppl 3:S197-206.

94. World Health O. *Global epidemiological surveillance standards for influenza.* Geneva: World Health Organization; 2013.

95. Vega T, Lozano JE, Meerhoff T, et al. Influenza surveillance in Europe: establishing epidemic thresholds by the moving epidemic method. *Influenza and other respiratory viruses.* 2013;7(4):546-558.

96. Vega T, Lozano JE, Meerhoff T, et al. Influenza surveillance in Europe: comparing intensity levels calculated using the moving epidemic method. *Influenza and other respiratory viruses.* 2015;9(5):234-246.

97. Vos LM, Teirlinck AC, Lozano JE, et al. Use of the moving epidemic method (MEM) to assess national surveillance data for respiratory syncytial virus (RSV) in the Netherlands, 2005 to 2017. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin.* 2019;24(20).

98. Grilc E, Prosenc Trilar K, Lajovic J, Sočan M. Determining the seasonality of respiratory syncytial virus in Slovenia. *Influenza and other respiratory viruses.* 2020.

99. Staadegaard L, Caini S, Wangchuk S, et al. Defining the seasonality of respiratory syncytial virus around the world: National and subnational surveillance data from 12 countries. *Influenza and other respiratory viruses.* 2021.

100. World Health Organization. Household air pollution and health. 2018; https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health. Accessed 29 September, 2020.

101. Smith KR, Mehta S. The burden of disease from indoor air pollution in developing countries: comparison of estimates. *International journal of hygiene and environmental health.* 2003;206(4-5):279-289.

102. ESMAP. *The State of Access to Modern Energy Cooking Services.* Washington, D.C.2020.

103. IEA. *SDG7: Data and Projections. Access to Affordable, Reliable, Sustainable and Modern Energy for All. Flagship Report — November 2019.* Paris (2019): International Energy Agency;2019.

104. World Health Organization. *WHO indoor air quality guidelines: household fuel combustion.* Geneva, Switzerland2014.

105. Rylance J, Chimpini C, Semple S, et al. Chronic Household Air Pollution Exposure Is Associated with Impaired Alveolar Macrophage Function in Malawian Non-Smokers. *PloS one.* 2015;10(9):e0138762.

106. Adaji EE, Ekezie W, Clifford M, Phalkey R. Understanding the effect of indoor air pollution on pneumonia in children under 5 in low- and middle-income countries: a systematic review of evidence. *Environmental science and pollution research international.* 2019;26(4):3208-3225.

107. Bruce NG, Dherani MK, Das JK, et al. Control of household air pollution for child survival: estimates for intervention impacts. *BMC public health.* 2013;13 Suppl 3:S8.

108. Jackson S, Mathews KH, Pulanic D, et al. Risk factors for severe acute lower respiratory infections in children: a systematic review and meta-analysis. *Croatian medical journal.* 2013;54(2):110-121.

109. Rudan I, Boschi-Pinto C, Biloglav Z, Mulholland K, Campbell H. Epidemiology and etiology of childhood pneumonia. *Bull World Health Organ.* 2008;86(5):408-416.

110. Sonego M, Pellegrin MC, Becker G, Lazzerini M. Risk factors for mortality from acute lower respiratory infections (ALRI) in children under five years of age in low and middle-income countries: a systematic review and meta-analysis of observational studies. *PloS one.* 2015;10(1):e0116380.

111. Smith KR, Samet JM, Romieu I, Bruce N. Indoor air pollution in developing countries and acute lower respiratory infections in children. *Thorax.* 2000;55(6):518-532.

112. Zar HJ, Ferkol TW. The global burden of respiratory disease-impact on child health. *Pediatric pulmonology.* 2014;49(5):430-434.

113. Buchner H, Rehfuess EA. Cooking and season as risk factors for acute lower respiratory infections in African children: a cross-sectional multi-country analysis. *PloS one.* 2015;10(6):e0128933.

114. Smith KR, McCracken JP, Weber MW, et al. Effect of reduction in household air pollution on childhood pneumonia in Guatemala (RESPIRE): a randomised controlled trial. *Lancet.* 2011;378(9804):1717-1726.

115. Mortimer K, Ndamala CB, Naunje AW, et al. A cleaner burning biomass-fuelled cookstove intervention to prevent pneumonia in children under 5 years old in rural Malawi (the Cooking and Pneumonia Study): a cluster randomised controlled trial. *Lancet.* 2017;389(10065):167-175.

116. Mortimer K, Lesosky M, Semple S, et al. Pneumonia and Exposure to Household Air Pollution in Children Under the Age of 5 Years in Rural Malawi: Findings From the Cooking and Pneumonia Study. *Chest.* 2020.

117. Jack DW, Ae-Ngibise KA, Gould CF, et al. A cluster randomised trial of cookstove interventions to improve infant health in Ghana. *BMJ Global Health.* 2021;6.

118. Chillrud SN, Ae-Ngibise KA, Gould CF, et al. The effect of clean cooking interventions on mother and child personal exposure to air pollution: results from the Ghana Randomized Air Pollution and Health Study (GRAPHS). *Journal of exposure science & environmental epidemiology.* 2021;31(4):683-698.

119. Selevan SG, Kimmel CA, Mendola P. Identifying critical windows of exposure for children's health. *Environmental health perspectives.* 2000;108 Suppl 3(Suppl 3):451-455.

120. Jedrychowski W, Galas A, Pac A, et al. Prenatal ambient air exposure to polycyclic aromatic hydrocarbons and the occurrence of respiratory symptoms over the first year of life. *Eur J Epidemiol.* 2005;20(9):775-782.

121. Kajekar R. Environmental factors and developmental outcomes in the lung. *Pharmacol Ther.* 2007;114(2):129-145.

122. Goshen S, Novack L, Erez O, et al. The effect of exposure to particulate matter during pregnancy on lower respiratory tract infection hospitalizations during first year of life. *Environ Health.* 2020;19(1):90.

123. Jedrychowski WA, Perera FP, Maugeri U, et al. Effect of prenatal exposure to fine particulate matter on ventilatory lung function of preschool children of non-smoking mothers. *Paediatric and perinatal epidemiology.* 2010;24(5):492-501.

124. Jedrychowski WA, Perera FP, Spengler JD, et al. Intrauterine exposure to fine particulate matter as a risk factor for increased susceptibility to acute broncho-pulmonary infections in early childhood. *International journal of hygiene and environmental health.* 2013;216(4):395-401.

125. Kinney PL, Asante K-P, Lee AG, et al. Prenatal and Postnatal Household Air Pollution Exposures and Pneumonia Risk: Evidence From the Ghana Randomized Air Pollution and Health Study. *Chest.* 2021;160(5):1634-1644.

126. McEvoy CT, Schilling D, Clay N, et al. Vitamin C supplementation for pregnant smoking women and pulmonary function in their newborn infants: a randomized clinical trial. *Jama.* 2014;311(20):2074-2082.

127. Korten I, Ramsey K, Latzin P. Air pollution during pregnancy and lung development in the child. *Paediatric respiratory reviews.* 2017;21:38-46.

128. McEvoy CT, Spindel ER. Pulmonary Effects of Maternal Smoking on the Fetus and Child: Effects on Lung Development, Respiratory Morbidities, and Life Long Lung Health. *Paediatric respiratory reviews.* 2017;21:27-33.

129. Lee KK, Bing R, Kiang J, et al. Adverse health effects associated with household air pollution: a systematic review, meta-analysis, and burden estimation study. *The Lancet Global Health.* 2020;8(11):e1427-e1434.

130. Collaborators; GL. 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;17(11):1133-1161.

131. Thongpan I, Vongpunsawad S, Poovorawan Y. Respiratory syncytial virus infection trend is associated with meteorological factors. *Scientific Reports.* 2020;10(1):10931.

132. Bloom-Feshbach K, Alonso WJ, Charu V, et al. Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. *PloS one.* 2013;8(2):e54445.

133. Colosia AD, Masaquel A, Hall CB, Barrett AM, Mahadevia PJ, Yogev R. Residential crowding and severe respiratory syncytial virus disease among infants and young children: a systematic literature review. *BMC infectious diseases.* 2012;12:95.

134. Weatherspark. Quetzaltenango Climate, Weather By Month, Average Temperature (Guatemala) - Weather Spark. https://weatherspark.com/y/11183/Average-Weather-in-Quetzaltenango-Guatemala-Year-Round. Accessed 2/12/2022, 2022.

135. El Instituto Nacional de Estadística. Portal de Resultados del Censo 2018. 2018; https://www.censopoblacion.gt/explorador. Accessed 10/10/2020.

136. Lindblade KA, Johnson AJ, Arvelo W, et al. Low usage of government healthcare facilities for acute respiratory infections in guatemala: implications for influenza surveillance. *BMC public health.* 2011;11:885.

137. JE. L. lozalojo/memapp: Second release of the MEM Shiny Web Application R package. 2018; https://zenodo.org/record/1173518#.YRvMEogzbgg. Accessed 2021, 15 May 2021.

138. Mullins JA, Lamonte AC, Bresee JS, Anderson LJ. Substantial variability in community respiratory syncytial virus season timing. *The Pediatric infectious disease journal.* 2003;22(10):857-862.

139. Foley DA, Yeoh DK, Minney-Smith CA, et al. The Interseasonal Resurgence of Respiratory Syncytial Virus in Australian Children Following the Reduction of Coronavirus Disease 2019-Related Public Health Measures. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2021.

140. Broberg EK, Waris M, Johansen K, Snacken R, Penttinen P. Seasonality and geographical spread of respiratory syncytial virus epidemics in 15 European countries, 2010 to 2016. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin.* 2018;23(5).

141. Tang JW, Loh TP. Correlations between climate factors and incidence--a contributor to RSV seasonality. *Reviews in medical virology.* 2014;24(1):15-34.

142. Dushoff J, Plotkin JB, Levin SA, Earn DJD. Dynamical resonance can account for seasonality of influenza epidemics. *Proceedings of the National Academy of Sciences of the United States of America.* 2004;101(48):16915.

143. Waris M. Pattern of respiratory syncytial virus epidemics in Finland: two-year cycles with alternating prevalence of groups A and B. *The Journal of infectious diseases.* 1991;163(3):464-469.

144. Choi EH, Lee HJ. Genetic diversity and molecular epidemiology of the G protein of subgroups A and B of respiratory syncytial viruses isolated over 9 consecutive epidemics in Korea. *The Journal of infectious diseases.* 2000;181(5):1547-1556.

145. Yu J, Liu C, Xiao Y, et al. Respiratory Syncytial Virus Seasonality, Beijing, China, 2007-2015. *Emerg Infect Dis.* 2019;25(6):1127-1135.

146. Madhi SA, Polack FP, Piedra PA, et al. Respiratory Syncytial Virus Vaccination during Pregnancy and Effects in Infants. *New England Journal of Medicine.* 2020;383(5):426-439.

147. Institute for Health Metrics and Evaluation (IHME). GBD 2019 Cause and Risk Summary: Upper Respiratory infections. 2019; https://www.healthdata.org/results/gbd_summaries/2019, 2022.

148. Collaborators GLRI. Quantifying risks and interventions that have affected the burden of lower respiratory infections among children younger than 5 years: an analysis for the Global Burden of Disease Study 2017. 2019.

149. IEA I. *Tracking SDG7: The Energy Progress Report.* UNSD, World Bank, WHO;2020.

150. Murray CJL, Aravkin AY, Zheng P, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet.* 2020;396(10258):1223-1249.

151. Lee AG, Le Grand B, Hsu HL, et al. Prenatal fine particulate exposure associated with reduced childhood lung function and nasal epithelia GSTP1 hypermethylation: Sex-specific effects. *Respir Res.* 2018;19(1):76.

152. Rosa MJ, Just AC, Kloog I, et al. Prenatal particulate matter exposure and wheeze in Mexican children: Effect modification by prenatal psychosocial stress. *Ann Allergy Asthma Immunol.* 2017;119(3):232-237.e231.

153. Lee AG, Kaali S, Quinn A, et al. Prenatal Household Air Pollution Is Associated with Impaired Infant Lung Function with Sex-Specific Effects. Evidence from GRAPHS, a Cluster Randomized Cookstove Intervention Trial. *American journal of respiratory and critical care medicine.* 2019;199(6):738-746.

154. Vanker A, Barnett W, Workman L, et al. Early-life exposure to indoor air pollution or tobacco smoke and lower respiratory tract illness and wheezing in African infants: a longitudinal birth cohort study. *The Lancet Planetary Health.* 2017;1(8):e328-e336.

155. Quinn AK, Adjei IA, Ae-Ngibise KA, et al. Prenatal household air pollutant exposure is associated with reduced size and gestational age at birth among a cohort of Ghanaian infants. *Environment international.* 2021;155:106659.

156. Greenough A. Does low birth weight confer a lifelong respiratory disadvantage? *American journal of respiratory and critical care medicine.* 2009;180(2):107-108.

157. Alexander DA, Northcross A, Karrison T, et al. Pregnancy outcomes and ethanol cook stove intervention: A randomized-controlled trial in Ibadan, Nigeria. *Environment international.* 2018;111:152-163.

158. Tielsch JM, Katz J, Khatry SK, et al. Effect of an improved biomass stove on acute lower respiratory infections in young children in rural Nepal: a cluster-randomised, step-wedge trial. *The Lancet Global Health.* 2016;4:S19.

159. Checkley W, Williams KN, Kephart JL, et al. Effects of a Household Air Pollution Intervention with Liquefied Petroleum Gas on Cardiopulmonary Outcomes in Peru. A Randomized Controlled Trial. *American journal of respiratory and critical care medicine.* 2021;203(11):1386-1397.

160. World Health Organization. *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide: executive summary.* Geneva: World Health Organization; 2021.

161. Katz J, Tielsch JM, Khatry SK, et al. Impact of Improved Biomass and Liquid Petroleum Gas Stoves on Birth Outcomes in Rural Nepal: Results of 2 Randomized Trials. *Glob Health Sci Pract.* 2020;8(3):372-382.

162. Johnson M, Pillarisetti A, Piedrahita R, et al. Exposure contrasts of pregnant women during the Household Air Pollution Intervention Network randomized controlled trial. *medRxiv.* 2021:2021.2011.2004.21265938.

163. Beltramo T, Levine DI. The effect of solar ovens on fuel use, emissions and health: results from a randomised controlled trial. *Journal of Development Effectiveness.* 2013;5(2):178-207.

164. Hartinger S, Lanata C, Hattendorf J, et al. Improving household air, drinking water and hygiene in rural Peru: a community-randomized–controlled trial of an integrated environmental home-based intervention package to improve child health. *International journal of epidemiology.* 2016;45(6):2089-2099.

165. National Institute of Statistics of Rwanda, Finance Mo, Economic Planning/Rwanda, Ministry of Health/Rwanda, ICF International. *Rwanda Demographic and Health Survey 2014-15.* Kigali, Rwanda: National Institute of Statistics of Rwanda, Ministry of Finance and Economic Planning/Rwanda, Ministry of Health/Rwanda, and ICF International;2016.

166. International Institute for Population Sciences - IIPS/India, ICF. *India National Family Health Survey NFHS-4 2015-16.* Mumbai, India: IIPS and ICF;2017.

167. Ministerio de Salud Pública y Asistencia Social - MSPAS/Guatemala, Instituto Nacional de Estadística - INE/Guatemala, Secretaría de Planificación y Programación del la Presidencia - Segeplán/Guatemala, ICF International. *Encuesta nacional de salud materno infantil 2014-2015: informe final.* Rockville, Maryland, USA: MSPAS, INE, Segeplán and ICF International;2017.

168. Johnson MA, Steenland K, Piedrahita R, et al. Air Pollutant Exposure and Stove Use Assessment Methods for the Household Air Pollution Intervention Network (HAPIN) Trial. *Environmental health perspectives.* 2020;128(4):47009.

169. Ballard TJ, Kepple AW, Cafiero C. *The food insecurity experience scale: developing a global standard for monitoring hunger worldwide.* Rome2013.

170. Kirby MA, Nagel CL, Rosa G, et al. Effects of a large-scale distribution of water filters and natural draft rocket-style cookstoves on diarrhea and acute respiratory infection: A cluster-randomized controlled trial in Western Province, Rwanda. *PLoS medicine.* 2019;16(6):e1002812.

171. Smith KR, Pillarisetti A. Household Air Pollution from Solid Cookfuels and Its Effects on Health. In: Mock CN, Nugent R, Kobusingye O, Smith KR, eds. *Injury Prevention and Environmental Health.* Washington (DC): The International Bank for Reconstruction and Development / The World Bank.; 2017.

172. Misra P, Srivastava R, Krishnan A, Sreenivaas V, Pandav CS. Indoor air pollution-related acute lower respiratory infections and low birthweight: a systematic review. *J Trop Pediatr.* 2012;58(6):457-466.

173. Agusti A, Faner R. Lung function trajectories in health and disease. *The Lancet Respiratory Medicine.* 2019;7(4):358-364.

174. Agustí A, Hogg JC. Update on the Pathogenesis of Chronic Obstructive Pulmonary Disease. *New England Journal of Medicine.* 2019;381(13):1248-1256.

175. Stanaway JD, Afshin A, Gakidou E, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet.* 2018;392(10159):1923-1994.

176. Northridge ME. Public health methods--attributable risk as a link between causality and public health action. *American journal of public health.* 1995;85(9):1202-1204.

177. Musher DM, Thorner AR. Community-acquired pneumonia. *N Engl J Med.* 2014;371(17):1619-1628.

178. Lu Y, Joseph L, Bélisle P, et al. Pneumococcal pneumonia prevalence among adults with severe acute respiratory illness in Thailand - comparison of Bayesian latent class modeling and conventional analysis. *BMC infectious diseases.* 2019;19(1):423.

179. Kodani M, Yang G, Conklin LM, et al. Application of TaqMan low-density arrays for simultaneous detection of multiple respiratory pathogens. *Journal of clinical microbiology.* 2011;49(6):2175-2182.

180. Pearce N. Analysis of matched case-control studies. *BMJ (Clinical research ed).* 2016;352:i969.

181. Falagas ME, Mourtzoukou EG, Vardakas KZ. Sex differences in the incidence and severity of respiratory tract infections. *Respiratory Medicine.* 2007;101(9):1845-1863.

182. Almirall J, Bolíbar I, Balanzó X, González CA. Risk factors for community-acquired pneumonia in adults: a population-based case-control study. *The European respiratory journal.* 1999;13(2):349-355.

183. Driscoll AJ, Deloria Knoll M, Hammitt LL, et al. The Effect of Antibiotic Exposure and Specimen Volume on the Detection of Bacterial Pathogens in Children With Pneumonia. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;64(suppl_3):S368-s377.

184. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *American journal of epidemiology.* 1985;122(5):904-914.

185. Pericone CD, Overweg K, Hermans PW, Weiser JN. Inhibitory and bactericidal effects of hydrogen peroxide production by Streptococcus pneumoniae on other inhabitants of the upper respiratory tract. *Infect Immun.* 2000;68(7):3990-3997.

186. Rowe AK, Powell KE, Flanders WD. Why population attributable fractions can sum to more than one. *American journal of preventive medicine.* 2004;26(3):243-249.

187. Poole C. A history of the population attributable fraction and related measures. *Ann Epidemiol.* 2015;25(3):147-154.

188. Wishaupt JO, Ploeg Tvd, Smeets LC, Groot Rd, Versteegh FGA, Hartwig NG. Pitfalls in interpretation of CT-values of RT-PCR in children with acute respiratory tract infections. *J Clin Virol.* 2017;90:1-6.

189. O'Brien KL, Walters MI, Sellman J, et al. Severe pneumococcal pneumonia in previously healthy children: the role of preceding influenza infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2000;30(5):784-789.

190. Morens DM, Taubenberger JK, Fauci AS. Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. *The Journal of infectious diseases.* 2008;198(7):962-970.

191. International Vaccine Access Center (IVAC) JHBSoPH. VIEW-hub. www.view-hub.org.

192. Imai C, Armstrong B, Chalabi Z, Mangtani P, Hashizume M. Time series regression model for infectious disease and weather. *Environmental research.* 2015;142:319-327.