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:

_____ _____

Catherine Bozio Eldridge Date

Imperfect Detections of Influenza and Non-Influenza Respiratory Viruses among Children and Adults Hospitalized with Pneumonia

By

Catherine Bozio Eldridge
Doctor of Philosophy

Epidemiology

_____
W. Dana Flanders, M.D., Sc.D.
Advisor

_____
Seema Jain, M.D.
Advisor

_____
Lyn Finelli, Dr.P.H., M.S.
Committee Member

_____
Neel Gandhi, M.D.
Committee Member

_____
Jorge Vidal, Ph.D., M.Sc.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Imperfect Detections of Influenza and Non-Influenza Respiratory Viruses among Children and Adults Hospitalized with Pneumonia

By

Catherine Bozio Eldridge
B.S., University of Kentucky, 2009
M.P.H., Emory University, 2011

Advisor: W. Dana Flanders, M.D., Sc.D.
Advisor: Seema Jain, M.D.

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
2016

Abstract

Imperfect Detections of Influenza and Non-Influenza Respiratory Viruses among Children and Adults Hospitalized with Pneumonia
By Catherine Bozio Eldridge

Community-acquired pneumonia (CAP) is a common cause of hospitalization among all ages in the United States. Respiratory viruses can be detected using real-time polymerase chain reaction (PCR) and serology; limitations of these tests and available results can lead to misclassification of pneumonia attributable to respiratory viruses. The Centers for Disease Control and Prevention Etiology of Pneumonia in the Community (EPIC) study was a prospective, multi-center, population-based, active surveillance study. Using these data, this dissertation addressed potential misclassification in estimating the proportions of CAP due to respiratory viruses and predicting influenza-associated pneumonia.

Missing serology results are a potentially important reason for misclassification of pneumonia etiology in the EPIC study. Among children and adults, 98.8-99.5% had naso-/oropharyngeal specimens and 37.5-48.6% had paired serum specimens available. We accounted for missing values using multiple imputation and estimated revised proportions of CAP with adenovirus, human metapneumovirus, parainfluenza virus, and respiratory syncytial virus detected (Aim 1), which were 14.4%, 15.0%, 8.9%, and 29.9%, respectively, among children. Among adults, the respective revised proportions were 4.2%, 4.7%, 4.0%, and 4.0%. These revised proportions were 0.8-3.2% higher than observed EPIC study estimates. In Aim 2, revised proportions of influenza virus detections were estimated to be 11.1% among children and 7.9% among adults, which were 2.1-4.4% higher than observed estimates.

For Aim 3, we developed and evaluated prediction models and scores for influenza-associated pneumonia, using factors readily available at clinical presentation. Two definitions of influenza-associated pneumonia were used: one based on imputed results from PCR and serology, and the other was based on observed EPIC study results. Significant predictors were age and influenza season among children and leukocytosis, underlying medical conditions, cough, abdominal pain, and influenza season among adults. Prediction scores among children and adults had consistently high negative predictive values and, overall, low positive predictive values.

By accounting for missing serology results, the proportions of CAP with specific respiratory viruses detected were higher than the observed results; thus, observed data may have underestimated the virus-specific burdens. Additionally, our prediction scores may reflect difficulty in predicting influenza-associated pneumonia and differentiating it from other causes of CAP, especially in children.

Imperfect Detections of Influenza and Non-Influenza Respiratory Viruses among Children and Adults Hospitalized with Pneumonia

By

Catherine Bozio Eldridge
B.S., University of Kentucky, 2009
M.P.H., Emory University, 2011

Advisor: W. Dana Flanders, M.D., Sc.D.
Advisor: Seema Jain, M.D.

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
2016

Acknowledgements

There are many people I would like to thank for their help and support on this research.

I feel incredibly fortunate for the committee I had. First, I would like to thank my committee co-chairs, Drs. Dana Flanders and Seema Jain, for their ideas, feedback, patience, and support throughout this process. I am also very thankful for the support and feedback from my committee members, Drs. Lyn Finelli, Neel Gandhi, and Jorge Vidal.

I also want to thank Anna Bramley and Carrie Reed as well as the other EPIC study investigators. Collaborating with CDC on my dissertation presented me with a unique and wonderful opportunity to present to and receive feedback from these investigators who are experts in this field. I have gained a lot throughout this collaboration, and I am incredibly grateful for the feedback that I have received from all of them.

I am also thankful for the love and support from my family and friends.

Finally, I thank Ron for his strength, love, and encouragement.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. Background and research plan

Section 1. Background

1.1 Background of Community-Acquired Pneumonia (CAP)

1.1.1 Description of CAP

Community-acquired pneumonia (CAP) is a common cause of hospitalization among all ages. Pneumonia and influenza combined is the ninth leading cause of death in the United States and the leading infectious cause of death [1]. In 2010, there were 1.1 million discharges of inpatients with pneumonia, which corresponds to a discharge rate of 366 per 100,000 people [2]. Incidence rates of CAP hospitalizations are the highest in young children (<2 years of age) and older adults (≥ 65 years of age) [3].

Pneumonia is an acute lower respiratory tract infection (LRTI) defined as inflammation of the lungs that causes mild to severe illness. Infectious pneumonia results from the invasion and overgrowth of microorganisms in lung parenchyma [4], mainly the alveoli. Pneumonia is defined by clinical signs (i.e., abnormal lung exam completed by the clinician) and symptoms (i.e., cough as reported by the patient) and relies on radiologic studies for more definitive diagnosis. Pneumonia can be caused by non-infectious and infectious agents, including bacteria, viruses, and fungi. The clinical presentation of pneumonia due to either a specific bacterium or virus overlaps and is nonspecific; thus presentation alone cannot be used to determine etiology.

Symptoms of pneumonia can vary among children, adults, and the elderly. The clinical presentation can also vary based on the causative agent, individual host factors, and severity of illness. However, many of the symptoms of pneumonia in all age groups (including fever, chills, and cough) are nonspecific and can overlap with other conditions.

The most common symptoms of pneumonia are cough (which may be productive and accompanied by yellow or greenish sputum), fever, shaking chills, and shortness of breath. Additional symptoms can include chest pain that gets worse upon deep breaths or cough, headache, temperature instability with chills and sweating, loss of appetite, fatigue, and confusion (especially in older people) [5]. In children, common physical findings include fever, rapid breathing, increasingly labored breathing, rhonchi[1], crackles, and wheezing [6, 7]. In some cases, patients with pneumonia associated with *Legionella* can present with gastrointestinal symptoms [8]. Reports of myalgia have been associated with influenza virus infection [9]. Despite these reports, few are based on prospectively collected epidemiological and laboratory data. Since it is challenging to discern specific community-acquired pneumonia pathogens with symptoms alone, age and epidemiologic clues might be helpful in narrowing possible detected agents.

Age is a well-known risk factor for pneumonia and children <5 years of age and adults ≥65 years of age are at an increased risk. The increased risk in children is partly due to their naïve immune system that results in infection [7]. Even among children, different pathogens are more commonly detected in younger children versus older children. For example, respiratory syncytial virus (RSV) is more common among children <5 years of age but *Mycoplasma pneumoniae* is more commonly detected among children >5 years of age [7]. For adults ≥ 65 years, the increased risk of community-acquired pneumonia is partially due to the aging process and decline in the immune system as well as the presence of underlying conditions that increase the risk for acquisition of respiratory pathogens and complications from pneumonia [9].

[1] A continuous snorelike sound in the throat or bronchial tubes.

Epidemiologic clues might also help in narrowing possible pathogens, particularly in adults. Most notably, staying in a hotel or cruise ship in the previous two weeks might be suggestive of *Legionella* species, which are transmitted from the environment to people through aerosolized water. Other important risk factors, including alcoholism, smoke exposure, underlying conditions such as chronic obstructive pulmonary disease (COPD), and the circulation of influenza in the community could also inform possible causative pathogens [10].

1.1.2 Case Definitions for Epidemiological Studies

Pneumonia is difficult to define and case definitions used in epidemiological studies can range in sensitivity and specificity. For example, the World Health Organization's (WHO) clinical definition is syndromic and thus very sensitive as it is defined by cough or difficulty breathing with rapid breathing relative to age [11]. Alternatively, a more specific definition is radiographically-confirmed pneumonia [12-14]. Even with more specific definitions that include radiographic-confirmation, there are challenges to final interpretation of films due to the subjective nature of the reading of a chest radiograph. However, a range of sensitive and specific case definitions has been used in different studies [7, 12] and there are no consistent definitions between studies.

Radiography is a diagnostic tool for pneumonia but there are few validated definitions for chest radiograph interpretation and it is vulnerable to inter- and intra-rater variability [13]. The radiographic definitions applied in epidemiological studies tend to be more specific than in clinical practice [14]. However, there are many indicators in a chest radiograph that are consistent with the diagnosis of pneumonia, which include lobar

consolidation, alveolar or interstitial infiltrates, and pleural effusions [14]. While lobar consolidation is the radiological finding most often associated with pneumococcal infection [14], it is difficult to distinguish between viral and bacterial CAP solely on radiographic evidence, especially in children. Given the nuances in radiograph interpretation, there is subjectivity to the interpretation of radiographic findings. The variability of radiographic interpretation can be within and between observers and differs for specific findings. For example, while a high degree of agreement can be attained about the presence or absence of consolidation, there is more disagreement about the description of infiltrates [14].

In 2005, as part of a WHO effort to standardize the use of radiographs for epidemiological pneumonia studies, Cherian et al. reported that even after training of clinicians and radiologists on radiographic terms, there was still varied agreement among different radiographic interpretations but certain terms allowed for more consistent interpretation [12, 13]. These terms became the standardized WHO radiology working group's definition of radiographic pneumonia and included the presence of consolidation (a dense or fluffy opacity with or without air bronchograms), other infiltrate (linear and patchy alveolar or interstitial densities), or pleural effusion (in the lateral space). Using alveolar consolidation or pleural effusions as the defined criteria for end-point pneumonia, there was 82.5% agreement in categorizing end-point pneumonia among 20 clinicians and radiologists [13]. Cherian et al. suggested that this reasonably high agreement may be achieved by employing a consistent definition of radiographic pneumonia and training based on WHO guidelines. In assessing intra-rater agreement for end-point pneumonia, there was 88.5% agreement of repeatability. The WHO guidelines

have been used to define pneumonia for many contemporary pediatric CAP studies, especially studies that focus on end-points for bacterial vaccine trials [13]. However, there are no such global guidelines for epidemiological studies in adults.

1.2 Etiology of CAP in the United States Based on Diagnostic Tests

Many pathogens can cause CAP, but, in the U.S.[15], the predominant causes are bacterial and viral pathogens [10], and in certain geographic areas, endemic fungi commonly cause pneumonia. The distribution of pathogens varies based on demographic and epidemiological risk factors for pneumonia including age, geographic region, exposures (i.e., smoking), underlying medical conditions, and access to preventative and treatment measures (i.e., vaccines and antimicrobials).

The most common bacterial causes of pneumonia are *Chlamydophila pneumoniae, Haemophilus influenzae, Legionella* species, *M. pneumoniae*, and *Streptococcus pneumoniae*. The most common viral causes of pneumonia are adenovirus (AdV), influenza A and B, parainfluenza (PIV) 1, 2, 3, and RSV [10, 15-18]. Prospective pneumonia etiology studies conducted in children and adults have shown that these are the most commonly detected pathogens, though the distribution varies between children and adults [16, 17]. However, these studies, which will be reviewed here, utilized different diagnostic methods, some of which are not commonly used currently. In addition, there have been molecular and antigenic advancements in diagnostic tests for different pathogens, which have informed contemporary pneumonia etiology studies.

1.2.1 Available Specimens for Detection of Pathogens Associated with Pneumonia

Specimens for pneumonia pathogen detection can be collected either directly from the lower respiratory tract/lung tissue or indirectly via upper respiratory tract specimens, whole blood, serum, and urine (in adults). Specimens from the lower respiratory tract include pleural fluid, bronchoalveolar samples, and endotracheal aspirates while upper respiratory tract samples include sputum and naso/oropharyngeal (NP/OP) swabs. Routine laboratory evaluation of specimens includes microscopy and culture of respiratory tract specimens and blood, antigen detection in urine, and detection of antibodies in serum. Currently, molecular detection using real-time polymerase chain reaction (PCR) is increasingly available for both clinical and research purposes and can be applied to most respiratory samples and in some cases blood for select pathogens [19] (Table 1.1).

Despite the available options for specimen collection and detection methods, the majority of specimens that can be practically obtained for both clinical and research purposes are not directly from the lung because this requires an invasive procedure that is not done except for rare clinical indications. This presents several challenges to the interpretation of pathogens detected among patients enrolled in pneumonia etiology studies. First, tests not directly from the lung may not reflect what is happening in the lung. For example, the presence of a virus in the upper respiratory tract, as detected by a NP/OP swab, may not be reflective of what is happening in the lower tract or may represent resolving infection rather than acute infection [20]. Second, the detection of bacteria from non-sterile sputum may represent colonization rather than true infection [21], as the oropharynx is colonized with normal bacterial flora that can also cause

pneumonia. High-quality sputum samples, with many white cells and few epithelial cells, may represent samples with reduced contamination from the oropharynx but obtaining such samples is not always possible [22-25].

1.2.2 Diagnostic Tests for Bacterial Pathogens

For detection of bacterial pathogens, the most commonly utilized methods include blood cultures, sputum Gram stain and culture, urine antigen tests, and whole blood PCR for select pathogens including pneumococcus (Table 1.1). Pneumococcus is considered the most common cause of CAP and thus this section is focused on pneumococcal methods of detection. Other bacteria, including *S. aureus* and Gram-negatives, also contribute to pneumonia, but specific detection methods for these pathogens outside of culture have not been developed. Blood culture has been traditionally used to identify bacteria because the presence of a bacterial species in a sterile site (i.e., blood) has been considered to be more indicative of the cause of pneumonia [21]. However, only a minority of patients with bacterial pneumonia has documented bacteremia [19]; additionally, blood cultures, while specific, lack sensitivity [26, 27], particularly following antibiotic treatment [28].

Urinary antigen tests for *S. pneumoniae* detection are widely available, but despite being rapid and available at the point of care [10], their clinical use is variable. Sensitivity estimates for pneumococcal urine antigen tests range from 64.3% to 88%, whereas the specificity estimates have been consistently high (89.7%-100%) [29-34] (Table 1.2). Despite the improvement of diagnostic yield in adults, evidence suggests that pneumococcal antigens can be detected in urine for several weeks following

pneumococcal pneumonia among patients with COPD [35]. In children, positive urinary antigen test results may reflect pneumococcal colonization [36] and thus such testing is not recommended in children <18 years old [15, 19, 21, 36].

Microscopy and culture of respiratory tract specimens, particularly sputum, have been used for pathogen detection among pneumonia patients both clinically and in previous pneumonia etiology studies in adults [10, 16]. However, these tests are limited in their ability to distinguish between infection and colonization with normal bacterial flora (i.e., *S. pneumoniae, H. influenzae*, and *S. aureus*) in the oropharynx [19, 22]. In addition, the process of collecting high-quality sputum samples, which would be more reliable samples, is difficult because sputum is most often collected by expectoration, thus allowing for potential contamination as it is coughed up through the oropharynx [22-25]. Collecting induced sputum with use of respiratory bronchodilators to induce a deep cough, and thus an improved expectoration sample, has been shown to be useful [37]. However, induced sputum is not practical due to additional cost and labor; thus proper sputum samples are infrequently collected [15]. [19]. If sputum samples are collected properly and are of high quality, pathogens detected by culture may be more indicative of the lower respiratory tract. However, recent antibiotic use before sputum collection may lower diagnostic yield [37, 38].

Due to its proximity to the lung parenchyma, pleural fluid samples can be useful for determining pneumonia etiology. Pleural fluid is collected invasively through thoracentesis when there is ample fluid available for collection; thoracentesis must be done by trained personnel (usually physicians) and may have adverse effects [39]. Thus, samples of pleural fluid are collected infrequently and often only in the most severely ill

patients. In addition, pleural fluid culture has low sensitivity ranging from 20%-30% [39-41]. These estimates could be underestimated by false-negative results due to prior antibiotic exposure [39], small sample volume, and/or improper collection, transport, and storage of sample [42]. Due to these limitations, pleural fluid PCR for detecting a pneumococcal gene target (*lytA*) has been developed and shows higher sensitivity (75%-88%) than culture with good specificity (71%-93%) [40-42]. Pleural fluid PCR for other bacteria have also been shown to be promising [43].

Whole blood PCR (using the autolysin (*lytA*) gene) has been reported to have improved sensitivity for pneumococcal detection relative to blood culture, especially among children and those with culture-negative invasive pneumococcal disease [44]. Studies have reported variable sensitivities (47%-100%) and consistently high specificities (88%-100%) when using whole blood PCR for *lytA* [27, 44-46]. Sensitivity and specificity depends whether the samples for blood culture and whole blood PCR were collected at the same time [47, 48]. Whole blood PCR does not require viable bacteria in the sample [44], can produce results within a clinically useful period [27, 47], and has a lower bacterial load required for detection [19, 49]. Additionally, some studies have reported that quantifying pneumococcal load using whole blood PCR correlates with severity of disease [28, 50, 51]. One challenge in analyses of whole blood PCR for pneumococcal detection is the lack of a good gold standard, as whole blood PCR may identify cases not captured by blood culture [27, 44, 52]. In addition, whole blood PCR is not optimized when there is low density of pathogen or small volume available for analysis [47-49].

Serology can also be used for bacterial detections, though it has not routinely been done in clinical and research settings for pneumococcus. Both acute and convalescent serum samples are necessary to compare antibody titer levels and determine whether there is evidence of infection. These samples need to be collected at least 3 weeks apart and thus serologic results have minimal impact on patient management. Pneumococcal surface adhesion A (PsaA) is a surface protein that is common to all pneumococcal serotypes and is highly immunogenic [53, 54], particularly among infants [55]. However, pneumococcal serology has limited sensitivity (42%-85%) [54, 56, 57] using a ≥2-fold rise in antibody concentration, which is more conservative than the standard 4-fold rise. Though it may be helpful in detecting bacteremic pneumonia in adults, the clinical utility is unclear for non-bacteremic cases and may be complicated for children [56, 58] due to higher rates of asymptomatic pneumococcal acquisition and colonization, which can elicit an increase in PsaA antibody.

Recent antibiotic use prior to specimen collection can also reduce diagnostic yield of bacterial detections [59, 60], as it can inhibit bacterial growth on blood, sputum, and pleural fluid cultures [26, 38, 48]. This has been particularly shown to affect the yield of pneumococcal detections [38]. This reduction in yield is dependent on initiation of antibiotic therapy relative to specimen collection and duration of antibiotic therapy and is difficult to estimate. A meta-analysis of 35 pneumococcal pneumonia studies estimated that this reduction of diagnostic yield was 67% (95% Confidence Interval (CI): 53%, 77%) for blood cultures and 26% (95% CI: 0%, 44%) for the urinary antigen test [61]. However, using PCR may help mitigate the reduction of diagnostic yield, as it does not

require viable bacteria in sample, even if it is collected after administration of antibiotics [44].

Though not a considered a diagnostic test, procalcitonin has been studied as a serum biomarker for potential differentiation between viral and bacterial infections. Procalcitonin is a precursor of the hormone calcitonin and is produced by parenchymal cells (found in the lung, liver and kidney) in response to microbial toxins and to proinflammatory mediators [62]. As it is down regulated in viral infection and increases in response to a bacterial infection [62], procalcitonin might be used to distinguish between viral and bacterial infections and, ultimately, to encourage or discourage antibiotic use for an individual patient. A meta-analysis of 21 studies with prospective procalcitonin measurements reported a pooled sensitivity of 92% and a pooled specificity of 73%, though these characteristics are dependent on the timing of serum collection, as it peaks after 6 hours [63]. A limitation is that the calcitonin gene that produces procalcitonin is ubiquitously expressed in parenchymal cells, which are not limited to the lung [64].

1.2.3 Diagnostic Tests for Viral Pathogens

Methods for detecting viral pathogens include culture of respiratory tract specimens, detection of antigens in respiratory specimen, detection of antibodies in serum, and PCR (Table 1.1). Appropriate specimens to collect for viral detections are nasal swabs or aspirates, throat swabs, NP/OP swabs, sputum, and sera [65]; however, the ideal specimen depends on the type of assay being performed [21]. Of the available detection methods, culture and PCR are the most reliable.

Viral culture was historically considered the gold standard for viral detections [66]. Viral isolates are useful for surveillance of circulating viruses and for susceptibility testing, though it can take 2-10 days to yield results and thus its clinical value is limited [66, 67]. Viral cultures can detect influenza A/B viruses, PIV, and rhinoviruses in addition to human metapneumovirus (HMPV) with special growth requirements [9]. In particular, detecting RSV by culture in adults can be difficult due to low viral titers in nasal secretions [68]. A challenge to the detection of viruses in NP/OP samples, even by culture methods, is that detection is not necessarily indicative of the cause of the pneumonia [18, 21] and may represent resolving infection rather than acute infection [20].

Antigen detection tests on respiratory specimens while rapid are less reliable viral detection methods because of potential false negative results. Both direct fluorescent antibody tests (DFA) and rapid diagnostic tests produce results in a more clinically relevant time frame, though their test performances are dependent on the type and quality of collected specimen [10, 69]. Rapid RSV diagnostic tests have sensitivities of 71%-95% and specificities of 80%-100% [9]. However, both DFA and rapid influenza diagnostic tests have lower sensitivity (27%-82%) and consistently high specificities (88%-99%) compared with viral culture [70-73]. Factors that can reduce the sensitivity of rapid influenza diagnostic tests are prolonged timing between illness onset and specimen collection [71, 73], viral RNA load at presentation, virus type (lower sensitivity has been observed for influenza B compared to that of influenza A), and test characteristics of the rapid tests as the validity of rapid influenza tests varies by test [71].

More sensitive PCR methods have increased not only the ability to detect and characterize respiratory pathogens but have also improved our understanding of the role of viruses in pneumonia [74]. While PCR may not capture every case, several studies have demonstrated that PCR had a higher proportion of pathogens detected than culture [66, 67, 75-77] and can also provide results in a more clinically relevant time frame (Table 1.2). PCR has increased sensitivities (86%-94%) and specificities (82%-99%) of influenza A/B viruses, PIV, and RSV detections relative to culture [77]. PCR test characteristics do not appear to depend on age but factors that can decrease the sensitivity of PCR are inadequate sampling, initiation of antiviral treatment, and timing between sample collection and illness onset [78]. Specifically, a longer duration between illness onset and the time to PCR specimen collection can lead to a lower likelihood of a positive result [75]. Studies have demonstrated that viral PCR detection and culture was higher within the 7 days after illness onset, as compared with 8-14 days after illness onset [77].

Studies have not adequately described the sensitivity of PCR relative to the timing between illness onset and PCR specimen collection. Weinberg et al. reported the proportion of positive PCR viral detections (RSV, influenza A/B viruses, and PIV combined) relative to timing: 8% on day 1 between illness onset and PCR specimen collection, 13%-42% from days 2-7, and <5% from days 8-14 [77]. As these do not reflect sensitivity, it is difficult to assess whether these proportions are the result of low prevalence of specific viruses in this study population, when PCR specimens are collected relative to illness onset, and/or the dynamics of a specific virus. Alternatively, Suess et al. reported the minimum duration of influenza viral shedding as 6.3 days in children and 6.7 days in adults [79], which is based on serial measurements and is

estimated from the time between illness onset to the last day a positive PCR specimen was taken. As with Weinberg et al., these results have limited ability in describing the sensitivity of PCR relative to timing because it does not estimate what proportion of viral detections could be missed due to prolonged timing. Additionally, due to the sample size and population (83 household contacts that were index cases, secondary cases, or non-cases), the results may have limited generalizability to hospitalized children and adults. To my knowledge, similar studies have not been performed for AdV, HMPV, PIV, and RSV.

Viral serology is not routinely performed in clinical settings because of the need for both acute and convalescent samples and the delay in getting results that can aid in clinical management. However, viral serology is commonly used in pneumonia etiology studies, and can increase yield of pathogens detected by PCR. The addition of serology to PCR can increase diagnostic yield by 0.4%-62% for non-influenza respiratory viruses [80, 81] and by 35%-130% for influenza viruses [78, 81]. However, there remain caveats to its use. For detection of antibodies to non-influenza respiratory viruses, enzyme immunoassays or complement fixation are used. Hemagglutination inhibition (HI) and/or microneutralization (MN) assays have been typically used for detecting antibody to influenza viruses [65, 69]. Sensitivities of HI and MN assays range from 60%-94% due to varying definitions of seroconversion, though specificities have a more narrow range of 83%-94% [82, 83]. In countries where influenza vaccination is used, a patient's vaccination status needs to be taken into consideration for determination and interpretation of influenza serology results.

1.2.4 Diagnostic Tests for Atypical Bacterial Pathogens

For detecting atypical bacterial pathogens, appropriate methods include sputum culture, detection of antigens in urine for select pathogens, detection of antibodies in serum, and PCR of respiratory sample (Table 1.1). Increasing attention has been placed on improving the diagnostic methods for these pathogens as there are no widely available and timely tests with the desired sensitivity and specificity for *C. pneumoniae* and *M. pneumoniae*.

Culturing atypical bacteria is more difficult than for other bacteria. Gram-negative *C. pneumoniae* and *Legionella* species require cell cultures or special media to grow, respectively [19]. Because *M. pneumoniae* is neither Gram-positive nor Gram-negative, it requires complex nutrients, grows slowly and thus cannot be readily cultured in a diagnostic laboratory [19]. Additionally, the sensitivity of detecting *M. pneumoniae* in culture may be no more than 60% using PCR as the gold standard [84]. For these reasons, culture methods have a low diagnostic yield [85] and thus results tend to not be available in a clinically relevant time frame. While sputum samples for *Legionella* detection may be difficult to obtain due to issues around ability of patients to expectorate, *Legionella* is not part of the normal naso-oropharyngeal flora and thus its detection in sputum culture is diagnostic [21].

In addition to sputum culture for *Legionella*, urine antigen assays are also available for both clinical and research purposes. All of the available urine antigen assays only detect *L. pneumophila* serogroup 1, as it accounts for 80%-95% of community-acquired cases of Legionnaire's disease in the United States [10, 19]. The

urinary antigen assay has sensitivity ranging between 70-90% and specificity of 99% relative to culture [86]. Additionally, in accordance with a positive urinary antigen test result, the diagnostic yield of sputum culture is increased [86]. Urinary antigen testing can show positive detection on the first day of illness onset and may remain positive for days after initiation of antibiotic therapy [35]. However, results may continue to be positive for at least 6 weeks after illness onset [10, 35]. The development of PCR assays has improved detection of *Legionella* as well as *M. pneumoniae* and *C. pneumoniae*. One study reported the diagnostic yield of *C. pneumoniae*, *M. pneumoniae*, and *Legionella* species combined as 7% by conventional methods (serology and culture of sputum, blood, or throat swabs), though PCR increased this yield to 17% [76]. Additionally, PCR has been shown to have a higher proportion of *M. pneumoniae* detections (11.3%-29%) relative to serology (5.6%-21%) [87, 88]. However, the sensitivity of PCR decreases as the timing between illness onset and PCR specimen collection increases. For example, the sensitivity of PCR for *M. pneumoniae* is 48% when timing is between days 0-21, though the sensitivity decreases to 29% for days 22-59 and to 12% after 60 days [89]. Thus, PCR is more ideal during the early phases of *M. pneumoniae* infection. Despite its usefulness in the detection of atypical pathogens, only a few commercially available PCR reagents have been sufficiently validated [10, 22]. However, PCR assays are not readily available or used for clinical settings.

Serology is a less reliable detection method for atypical bacteria but has been historically used in previous pneumonia etiology studies and may still be used in certain clinical settings. The antibody response to *C. pneumoniae* can complicate serologic testing as the IgG antibody may reach a diagnostically high titer 6-8 weeks after onset of

illness for the primary infection, but if reinfection occurs, the levels of IgG antibody titer increases quickly within 1-2 weeks [90]. Additionally, evidence suggests serologic cross-reactivity between *M. pneumoniae* and *C. pneumoniae* [91] and between *M. pneumoniae* and other *Mycoplasma* species or Gram-negative bacteria [92]. For detecting *M. pneumoniae*, the sensitivity of IgM assays ranged from 32% to 84% and the specificity was >95% [92]. As age increases, sensitivity may decrease and, in contrast, the specificity may increase [89]. Because IgM antibody levels for *M. pneumoniae* may persist for years [84, 89] and *C. pneumoniae* testing is complicated by persistent or chronic infection [90], the most reliable serologic evidence may be a ≥4-fold rise in IgG from paired serum samples. Sensitivities of IgG assays for *M. pneumoniae* ranged between 52%-78% in children and between 89%-92% in adults [84]. In addition, commercially available serology assays for both *C. pneumoniae* and *M. pneumoniae* lack validation [84]. Ultimately, serology is not ideal for patient management due to the need for both acute and convalescent sera samples timed 3-10 weeks apart, though can be useful in epidemiological studies.

1.2.5 Etiology of CAP in Children and in Adults

Few population-based studies with a comprehensive approach to detecting pneumonia etiology have been conducted in the United States, the most recent one of which was in the 1990's. The two main studies were conducted by Michelow et al. in children and Marston et al. in adults. Both these studies were conducted before widespread use of pneumococcal vaccination among children in the United States [3].

Michelow et al. prospectively collected data on 154 children who were hospitalized with radiographically-confirmed LRTI between the ages of >6 weeks and ≤18 years from January 1999 to March 2000 in one hospital in Dallas, TX. Diagnostic methods included blood cultures; pleural fluid culture (when available); viral DFA and culture for AdV, influenza A/B, PIV, and RSV; viral culture for rhinovirus and enterovirus; pneumolysin-based whole blood PCR; and paired serology for AdV, *C. pneumoniae*, influenza A/B, *M. pneumoniae*, PIV, and RSV. A pathogen was identified in 79% of children (n=122) and bacteria and viruses accounted for 60% and 45% of detections, respectively. As both bacteria and viruses can be detected within an individual, the sum of the proportions of viral and bacterial detections can exceed 100%. The most common pathogens detected were *S. pneumoniae* (44% of all cases), influenza A (17%), *M. pneumoniae* (14%), RSV (13%), PIV 1-3 (13%) and *C. pneumoniae* (9%) [17]. Mixed bacterial/viral detections were present in 23% of children (Table 1.3). No microbial agent was detected in 21% of children, which was thought to be due to antibiotic use prior to specimen collection, the absence of tests for other specific pathogens, and missing convalescent viral and atypical serology in 21% of patients. Since the study was conducted, two events have occurred: 1) more sensitive molecular diagnostic tests have been developed that are used in children [93, 94], and 2) there is widespread use in the United States *H. influenzae* serotype b (Hib) conjugate vaccine, heptavalent pneumococcal conjugate vaccine (PCV-7), and later 13-valent PCV. PCV-7 has been reported to contribute to reduced all-cause pneumonia hospitalizations in all age groups, especially in children <5 years of age [3, 95].

Some microbial causes of pneumonia are more common among specific age groups [7]. For example, *M. pneumoniae* is more common in school-aged children. Viruses cause a significant proportion of CAP, especially in children <2 years of age [6, 7, 17]. However, the prevalence of viral pneumonia decreases, though not dramatically, with age [96-98]. Michelow et al. observed that viral pathogens were detected in approximately 42% of those <6 months of age, though decreased to approximately 12% among those ≥5 years of age [17].

Marston et al. collected data on 2776 adults ≥18 years hospitalized with CAP in two counties of Ohio from January 1 to December 31, 1991. Diagnostic tests included blood cultures; sputum cultures; pleural fluid cultures (when available); *L. pneumophila* urinary antigen testing, and serology for *C. pneumoniae*, influenza, *L. pneumophila*, *M. pneumoniae*, and RSV. Marston had different definitions of pneumonia based on laboratory methods which included definite, probable, and possible cases. Definite cases included those that had selected bacteria isolated from blood or pleural fluid cultures; had 4-fold rise in antibody titer for *L. pneumophila, M. pneumoniae, C. pneumoniae*, RSV, or influenza virus; or had *Legionella* species or influenza virus isolated from respiratory secretions. Possible cases could have been based on a single acute serology sample. Among all cases, including the possible category, the most common pathogens detected were *M. pneumoniae* (32.5%), *S. pneumoniae* (12.6%), *C. pneumoniae* (8.9%), and seasonal influenza A (7.4%) [16] (Table 1.3). Among the definite cases, the most common pathogens detected were seasonal influenza A (7.4%), *S. pneumoniae* (5.5%), and *M. pneumoniae* (5.4%). No microbial agents were detected in 55.7% of adults when including all cases, but it was 83.5% among the definite cases. The lack of detections

was thought to be due to the insensitivity of available diagnostic tests and partially due to antibiotic use before specimen collection, particularly for pneumococcal detections. Detection of ≥2 pathogens was observed in 50 (1.8%) adults. Events that have affected the epidemiology of pneumonia among adults have also occurred including widespread use of the Hib vaccine and PCV-7 in children. The indirect herd effects of pediatric immunization with PCV has reduced all-cause pneumonia hospitalizations [3, 95]; invasive Hib disease [99] has also decreased in older age groups.

Co-detection is another important aspect of CAP etiology. Comparing the proportions of co-infections from the studies of Michelow et al. and Marston et al., infections of at least two detected pathogens appear to be higher in children (29%) [17] than in adults (1.8%) [16]. However, more recent estimates of co-detections are needed, which requires systematic diagnostic testing in research studies. When multiple pathogens are detected, the significance of a single pathogen is made more difficult [100], especially in distinguishing whether one or both agents are contributing to infection. Bacterial and viral co-detections are also of interest, as primary viral infection may make an individual more susceptible to a secondary bacterial infection (as in the case of influenza and pneumococcal pneumonia) [101, 102]. More so than single pathogen detections, identifying co-detections are dependent on the selection of diagnostic tests, the quality and type of specimens analyzed, and systematically testing patients for all pathogens of interest [15].

Despite the availability of diagnostic tests, the etiology of CAP remains undetermined for a proportion of hospitalized patients. Michelow et al. reported unknown etiology in 21% of children, whereas Marston et al. reported unknown etiology

in 56% of adults. In a study of 109 adults with CAP presenting to an emergency department that used both conventional methods (blood and sputum cultures, and serology) and lung aspirates to detect bacterial or atypical bacterial pathogens, there were 55 (50%) patients with no pathogens detected. Among 13 identified pathogens, 36 (65%) were identified from lung aspirates, 18 of which were *S. pneumoniae* [103]. Possible explanations for the unknown etiology could be false-negative results due to antibiotic use prior to specimen collection, low sensitivity from available detection methods, variable specimen quality, and/or emergence of pathogens not being captured by used detection methods [104].

1.2.6 Influenza-Associated Pneumonia

Pneumonia is a known complication of seasonal influenza [65, 105-108] and pandemic influenza [109-113]. Each year in the United States, influenza virus infection causes an estimated 54,000-430,000 hospitalizations [114]. Data from the CDC Etiology of Pneumonia in the Community (EPIC) study report overall influenza-associated pneumonia hospitalization incidence of 1.1/10,000 children and 1.5/10,000 adults, with rates being highest in young children (<5 years of age) and older adults (≥ 65 years of age), through prospective, population-based, active surveillance [115, 116].

Several studies based on laboratory-confirmed influenza hospitalization surveillance data have indicated that pneumonia is the most common complication of influenza hospitalization affecting between 15-43% patients depending on age and season surveyed. Among children hospitalized with laboratory-confirmed influenza infection, the proportion of pneumonia associated with seasonal influenza (from 2003-2009

seasons) and influenza A(H1N1)pdm09 (from 2009-2010 season) were 23% and 36%, respectively [117]. Among adults hospitalized with laboratory-confirmed influenza infection, the proportion of pneumonia associated with seasonal influenza (from 2005-2009 seasons) and influenza A(H1N1)pdm09 (from 2009-2010) were 35.4% and 42.8%, respectively [113]. More recent post-pandemic data from October 2010 through April 2011 reported the proportion of pneumonia discharges to be 15.5%-28.8% among children and 32.1%-38.6% among adults hospitalized with influenza A(H1N1)pdm09, H3N2, and B viruses combined [118]. These estimates are dependent on the proportion of patients with available chest radiographs and could be different if all hospitalized patients with suspected pneumonia had radiographs performed.

While pneumonia is the most common complication of influenza, clinical and radiographic findings of influenza-associated pneumonia have only been documented in a few studies before and during the 2009 H1N1 pandemic. In a case series of 451 hospitalized adults and children with laboratory-confirmed influenza A(H1N1)pdm09 infection, patients with and without influenza-associated pneumonia had similar clinical presentation, but patients with pneumonia were more likely to have shortness of breath (73% vs 52%) and diarrhea (27% vs 18%) than those without pneumonia [107]. Among studies of patients hospitalized with influenza A(H1N1)pdm09, radiographic findings of influenza A(H1N1)pdm09 pneumonia included bilateral infiltrates, infiltrate limited to one lobe, and bilateral lobar/multilobar infiltrates [107, 109, 119, 120].

Underlying medical conditions have been previously described among patients hospitalized with influenza with and without pneumonia. Among adults and children hospitalized with influenza A(H1N1)pdm09 virus infection, patients with pneumonia

were less likely to have underlying medical condition (61% vs 71%), including asthma or chronic obstructive pulmonary disease (COPD) (28% vs 38%), but were more likely to have neurological disease (15% vs 7%) than those without pneumonia [107]. Using laboratory-confirmed influenza hospitalization surveillance data from 2003 to 2008, pediatric patients with pneumonia were more likely to have asthma (24% vs 19%), but were less likely to have hemoglobinopathy (2% vs 4%) than pediatric patients without pneumonia [121].

Influenza antiviral agents are recommended for patients with suspected influenza who are hospitalized, including for those with pneumonia [122]. Timely initiation of antiviral therapy is crucial, as influenza-associated pneumonia is associated with an increased risk of severe outcomes, including intensive care unit (ICU) admission, acute respiratory distress syndrome, sepsis and death [107, 113, 119]. Among influenza A(H1N1)pdm09 virus hospitalizations, those with pneumonia had longer median time from illness onset to antiviral initiation and were less likely to receive influenza antiviral agents within two days of illness onset compared to those without pneumonia [107]. Because delayed initiation of antiviral therapy may be correlated with severity of illness, it is important to quickly identify those with suspected influenza-associated pneumonia to guide diagnostic testing and, if appropriate, to initiate antiviral therapy [65, 123, 124].

Clinical features of influenza-associated pneumonia are difficult to distinguish from other causes of CAP. In addition, not all patients with suspected influenza-associated pneumonia get tested for influenza. However, most patients who present with CAP, regardless of etiology, receive antibiotics [116]. Understanding which patients have influenza, another respiratory pathogen, or co-infections, could help guide treatment

with antivirals and antibiotics, especially in situations when influenza testing is not available, not accurate (settings which only have rapid tests), and not timely.

Clinical prediction models can be developed to help predict those with influenza-associated pneumonia using symptoms and signs, rather than solely relying on clinical judgment. However, no such formal model for influenza-associated pneumonia has been developed. In addition, prediction models for influenza in general are also limited in discerning influenza from other etiologies and most have been limited to outpatient settings. Of the few predictions of influenza infection that focus on hospitalized populations, one reported balanced sensitivity and specificity, but weak positive predictive value and high negative predictive value for influenza-like illness and FARI (fever and either cough, sore throat, runny nose, difficulty breathing, or earache) [125]. Bjarnason et al. used bivariate analyses to compare symptoms and radiographic findings between 22 patients with influenza A(H1N1)pdm09 pneumonia and 291 patients with CAP due to other etiology, which included bacteria and atypical bacteria but not viruses. Hospitalized adults with influenza A(H1N1)pdm09 pneumonia were more likely to present with expectoration of blood (27% vs 10%, $p=0.02$), dyspnea (95% vs 68%, $p=0.01$), headache (55% vs 33%, $p=0.06$), and diarrhea (32% vs 15%, $p=0.06$) than those with CAP due to other etiology. In this same study, hospitalized adults with influenza A(H1N1)pdm09 pneumonia were also less likely to have lobar infiltrates (32% vs 69%, $p<0.001$) but more likely to have bilateral interstitial infiltrates (50% vs 9%, $p<0.001$) than those with CAP due to other etiology [126]. Understanding the clinical and radiographic differences between influenza A(H1N1)pdm09 pneumonia and CAP due to other etiology is important, but these bivariate analyses have limited ability to provide

meaningful combinations of symptoms and/or signs needed to quickly identify patients with influenza-associated pneumonia. Ultimately, a clinical prediction model of influenza-associated pneumonia incorporating symptoms and signs as well as other meaningful variables may be helpful.

1.3 Latent Class Analysis

Rationale for Needing a Latent Variable and for Selecting Latent Class Analysis

Imperfect viral detection methods and prolonged timing between illness onset and specimen collection for PCR can misclassify pneumonia cases associated with influenza and other respiratory viruses. In addition to being imperfect measurements, these variables also describe an underlying latent construct of viral pneumonia. A latent construct is a variable that is not directly observed or measured and therefore must be constructed through the observation of related variables. For example, while the presence of infection cannot be directly observed (and therefore is considered to be latent), it can be inferred through the observed results of positive diagnostic tests and other laboratory measurements. In this dissertation, the latent variable of interest is pneumonia etiology, which requires a latent variable model for analysis. A latent variable model measures one or more latent variables from a set of related observed variables using a statistical model. Latent class analysis (LCA) is a type of latent variable model that is based on the assumption that observed categorical variables are imperfect measurements of an underlying categorical latent construct. LCA utilizes these multiple imperfect measurements to construct a latent variable in order to address misclassification. LCA has been applied to other infectious diseases, such as tuberculosis [127], Human

Immunodeficiency Virus infection [128], brucellosis [129], pneumococcal pneumonia [59, 130], and canine Echinococcus [131], to better estimate of disease prevalence and/or test characteristics. For this dissertation, the application of LCA can be used to address the problem of misclassification and may improve estimates of the proportion of CAP cases associated with a specific viral pathogen among children and adults enrolled in the EPIC study.

Based on the defined aims, traditional regression analysis, factor analysis, and cluster analysis are not suitable. Traditional regression analysis can only utilize data from observed variables, whereas LCA can create a categorical latent variable based on observed categorical indicators. Further, LCA also does not require the traditional modeling assumptions, such as multivariate normality and linear relationships [132, 133]. LCA is a categorical analogue to factor analysis (which usually uses continuous latent variables and observed variables), which is also a type of a latent variable model. Factor analysis is more focused on the structure of latent variables and their covariances, whereas LCA is more focused on the structures of cases, which is more of interest in this dissertation. Cluster analysis is closely related to LCA as it identified groups of cases with similar characteristics and each cluster is represented in a variable [132]. However, these clusters are not identified in utilizing statistical theory. Instead, LCA identifies unobserved heterogeneity from observed variables that are only explained by a latent variable through maximum likelihood estimation [132]. Finally, cases are deterministically assigned to classes in cluster analysis, but LCA estimates a probability of membership for each class. Therefore, LCA is more appropriate in identifying classes

of pneumonia etiology and in being able to apply maximum-likelihood estimation in order to better guide the structure of this latent variable.

Fundamentals of Latent Class Analysis

For LCA, two assumptions need to be made. First, distributional assumptions need to be made for the observed variables and the latent variable [134]. As the observed variables are categorical, the multinomial distribution is assumed. Depending on how many latent classes are created, either the binomial or multinomial distribution is assumed. Second, conditional independence is assumed, which means that the observed variables are independent of each other conditional on the latent variable [133, 135].

In LCA, the two fundamental parameters of interest are the latent class probability and the item-response probability. The latent class probability describes the probability of being in a latent class. For example, the probability of being in class t of a latent variable X is expressed as $P(X=t)$. The sum of latent class probabilities over all classes should equal one. Two important aspects of these latent classes are the number and the relative size of them. The number of classes can be determined by an a priori hypothesis; alternatively, using statistical tests of fit can inform an appropriate number of classes. The relative size gives an idea of how much of the study population is categorized in a latent class.

The other fundamental parameter of LCA is the item-response probability, or the probability of observing a response pattern given a latent class. An example of the item-response probability of observing c response in a dichotomous variable C given latent class t is expressed as $P(C=c|X=t)$. An alternative form is expressing this probability

logistically: $P(C = 1|X = 1) = \frac{e^{\alpha+\beta_1}}{1+e^{\alpha+\beta_1}}$, in which α is the log-odds of C=1 for those who have X=0 and β1 is the log of the odds ratio of C=1 comparing X=1 to X=0. This conditional probability describes the relationship between the latent variable and observed variables and indicates how likely or unlikely an observation is to be in a latent class [133]. The number of item-response probabilities equals the number of unique combinations of the observed variables [133]. For example, if there are 4 observed variables (with A, B, C, and D number of levels for each variable), there will be A+B+C+D item-response probabilities for each class of the latent variable. Within a latent class, the item-response probabilities for a given observed variable should also sum to one. These item-response probabilities are then used to ascribe labels to the latent classes – in this case, whether the latent classes can be labeled as pneumonia due to a specific virus, pneumonia due to other viruses, or pneumonia not due to a virus.

If these parameters of latent class probability and item-response probabilities are multiplied together, P(C)*P(C|X), the result is the joint probability of P(X and C). When applied to all observed variables and the latent variable, equation 1 can be expressed as (where A, B, …, E are observed variables and i, j, …, m are levels within each observed variable, respectively and there are t classes in the latent variable X):

$$P\left(A_i B_j \dots E_m X_t\right) = P(A_i|X_t) * P\left(B_j\middle|X_t\right) * \dots * P(E_m|X_t) * P(X_t) \qquad \text{(Eqn. 1)}$$

This equation supports the conditional independence assumption that the observed measurements are conditionally independent given the latent variable [133].

The standard approach to estimating these parameters is the maximum-likelihood estimation via the Expectation-Maximization (EM) algorithm. The EM algorithm calculates the likelihood function's expected value and then finds the parameter values that maximize this function [136]; this is the likelihood function, $L = \sum_{h=1}^{n} ln\left\{\sum_{j=0}^{K-1} \eta_j \prod_{i=1}^{p} \pi_{ij}^{X_{ih}} (1-\pi_{ij})^{1-X_{ih}}\right\}$ [137]. To maximize this function, a logit equation, written in a log-linear formulation, for each observed variable is solved:

$$\ln P(X=x, Y=y) = b + b_x^X + \sum_{\ell=1}^{L} b_{y_\ell}^{Y_\ell} + \sum_{\ell=1}^{L} b_{x,y_\ell}^{X,Y_\ell}$$ [138] (Eqn. 2),

in which X represents the latent variable with x levels and Y_l is one of the L observed variables with l levels. This iterative process continues until halted by either the completion of a specified number of iterations or, preferably, convergence of a maximum.

The potential problems in conducting LCA are model identifiability and presence of local maxima. First, model parameters may not be identifiable in the model estimation process, meaning that there is not a unique set (i.e., more than two sets) of parameter estimates that yields the maximum likelihood. The model identifiability can be investigated to see if the information matrix (based on the second derivatives of the log-likelihood function) are positive definite and/or to estimate the model parameters with different starting values [139]. Another possible problem associated with the estimation in LCA is the presence of local maxima. In traditional regression analysis when maximum likelihood estimation is employed, it is assumed that the global maximum of the likelihood function is estimated, mainly because the data is directly observed. However, because the latent variable is not directly observed, it is uncertain whether the

local or global maximum of the likelihood function has been maximized. However, the best way to proceed is to use different random starting values; if several starting values converge to the same highest likelihood value, one can be confident in having the global maximum solution.

Exploratory LCA and Confirmatory LCA

Exploratory LCA is primarily concerned with the structure of the latent variable, especially the number of latent classes. For exploratory LCA, it is not necessary to have explicit hypotheses to test, but it is helpful to have a general idea of possible labels of the latent classes and to think about the relative size of each group (as reflected in the latent class probability). Both the statistical fit and an appropriate number of classes for the hypothesis should be considered in deciding the final number of latent classes that will be modeled in confirmatory LCA. As the latent variable of interest is pneumonia etiology, it was ultimately decided to have three mutually exclusive latent classes with presumable labels of pneumonia due to a specific virus, pneumonia due to other viruses, and pneumonia not due to a virus.

Unlike exploratory LCA in which no restrictions have been imposed on parameter estimates, confirmatory LCA is performed to test specific hypotheses on item-response probabilities and latent class probabilities. The two types of restrictions that can be placed on parameters are equality constraints and specific value constraints across latent classes [133]. If an equality constraint is imposed, the estimates of two or more parameters take on the same unspecified value, whereas the specific value constraint

required that one or more parameters equal a value of prior specification. The number and the type of such hypotheses are ultimately at the investigator's discretion.

Simulations were conducted to ensure model identifiability based on realistic observed variables that are intended to be in the latent class model. For the creation of three latent classes, at least three tests tend to be sufficient in order for the model to be identifiable. In this process, the intention was for the simulation to be similar to what would be realistic, which included plausible sensitivities and specificities for PCR and serology. Dr. Flanders and I both did simulations independently with 3 markers and the latent variable was well identified.

Section 2. Dissertation Research Plan

2.1 Objective and List of Specific Aims

The overall objective of this dissertation was to account for misclassification of pneumonia etiology from imperfect detections of influenza and non-influenza respiratory viruses, which was translated into two goals. The first goal was to provide a revised proportion of CAP due to AdV, HMPV, PIV, RSV, and influenza A/B viruses. The second goal was to develop and evaluate prediction models and scores for influenza-associated pneumonia. For both goals, we compared the findings using the observed data from the EPIC study to the findings after accounting for potential misclassification. To achieve these objectives and goals, the following three specific aims were addressed using data from the CDC Etiology of Pneumonia in the Community (EPIC) study:

1) Estimate the revised proportions of CAP due to AdV, HMPV, PIV, and RSV among children and adults hospitalized with CAP, accounting for missing

serology results using multiple imputation,

2) Estimate the revised proportion of CAP due to influenza A/B viruses among children and adults hospitalized with CAP, accounting for missing serology results using multiple imputation, and
3) Develop and evaluate prediction models and scores for influenza-associated pneumonia using two definitions (one that was based on multiply imputed results for PCR and serology, and the other that used only observed results from PCR and serology). Predictors of influenza-associated pneumonia using both outcome definitions were compared to determine any differences.

The rationale for separating the first and second aims for influenza and other respiratory viruses was the difference in serologic assays performed. Indirect enzyme immunoassays were used for AdV, HMPV, PIV, and RSV detections, whereas HI and MN assays were used for detection of influenza A/B viruses. Each aim was performed separately in children and adults primarily because the distribution of pneumonia etiologies differs between these two age groups.

Finally, secondary analyses were performed with the goal of performing an additional analytic technique to address the first goal and of providing an alternative definition for a key variable in the second goal.

1) Estimate the revised proportions of CAP due to AdV, HMPV, PIV, and RSV through the application of latent class analysis (LCA),
2) Estimate the revised proportion of CAP due to influenza A/B viruses through the application of LCA, and

3) Develop and evaluate prediction models and scores for influenza-associated pneumonia using an alternative definition for influenza season using CDC surveillance data.

2.2 Data Source: CDC Etiology of Pneumonia in the Community (EPIC) Study

The CDC EPIC study was a prospective, multicenter, population-based, active surveillance study that systematically enrolled and tested patients of all ages using comprehensive diagnostic methods to determine etiology and estimate population-based incidence of CAP requiring hospitalization among adults and children in the United States. From January 1, 2010 to June 30, 2012, children <18 years of age were enrolled from Le Bonheur Hospital (Memphis, TN), Vanderbilt Hospital (Nashville, TN), and Primary Children's Hospital (Salt Lake City, UT). Adults ≥ 18 years of age were enrolled from Northwestern Memorial Hospital, Rush University Medical Center, and John H. Stroger Jr. Memorial Hospital of Cook County (Chicago, IL) as well as Vanderbilt Hospital and Baptist Hospital (Nashville, TN).

Patients were prospectively identified and enrolled for those who met the case definition of CAP and were admitted for inpatient care, including those who were admitted for observation. Case ascertainment occurred in study network sites through emergency rooms, outpatient clinics, and hospital admitting departments. Informed consent was obtained prior to enrollment.

People admitted to a study hospital were eligible if they 1) resided within the hospital catchment area; 2) had evidence of acute infection on presentation defined as documented or reported fever, hypothermia, or abnormal white blood count (leukocytosis

or leukopenia); 3) had evidence of an acute respiratory illness defined as new cough, sputum production, chest pain, shortness of breath, tachypnea, abnormal finding(s) from chest examination, or acute respiratory failure requiring mechanical ventilation; and 4) had a chest radiograph of evidence consistent with pneumonia within 72 hours of admission. Final inclusion in the study included blinded independent confirmation from a study radiologist and met the criteria of consolidation, infiltrate, or effusion.

Patients were excluded from the study if they were recently hospitalized (<7 days for immunocompetent children, <28 days for immunocompetent adults, or <90 days for immunosuppressed children or adults), enrolled in this study <28 days earlier, resided in an extended care facility, newborns who never left the hospital, or had a clear alternative non-pneumonia diagnosis. Patients with the following conditions were also excluded: human immunodeficiency virus infection with CD4 cell count <200 cells/mm^3, solid organ or hematopoietic stem cell transplant ≤90 days earlier, active graft versus host disease or bronchiolitis obliterans, tracheostomy or percutaneous endoscopic gastrostomy tube, cancer with neutropenia, or cystic fibrosis.

At the time of enrollment, clinical specimens were obtained as soon as possible and ideally before antimicrobial administration and were then processed appropriately to allow for multiple-pathogen testing. Whole blood, acute sera, and NP/OP swabs were collected in all enrolled patients and urine was only collected in adults ≥18 years old (Figure 1.1). The collection of convalescent sera required a follow-up visit in the hospital 3-10 weeks after enrollment. Clinical, demographic and epidemiologic information were collected through interviews and medical chart review.

For all enrolled patients, laboratory testing was completed for the detection of viruses (AdV, coronavirus, HMPV, influenza A/B viruses, PIV, rhinovirus, and RSV) and of bacteria (*C. pneumoniae*, *Legionella* species, *M. pneumoniae*, *S. aureus*, and *S. pneumoniae*). Urine samples from adults were tested for the detection of *Legionella* and *S. pneumoniae*. Sites had the option of collecting induced sputum from subjects, though only high-quality samples were processed for routine Gram stain and culture and included in the final analysis. For any patients that required thoracentesis or collection of lower respiratory tract samples (i.e., bronchoalveolar lavage, bronchial washing or endotracheal aspiration), samples were also tested for the aforementioned viral and bacterial pathogens. For any PCR testing, all sites were required to use CDC primers and probes in order to standardize results across sites.

A positive PCR test for AdV, coronavirus, HMPV, influenza A/B viruses, PIV, RSV, or rhinovirus was defined as a cycle threshold value of <40 from a NP/OP PCR assay. A positive serologic test for adenovirus, HMPV, influenza A/B viruses, parainfluenza virus, or RSV was defined as a ≥4-fold rise in agent-specific IgG antibody titer between paired acute and convalescent sera.

A bacterial pathogen was defined as being present if *Chlamydophila pneumoniae* or *Mycoplasma pneumoniae* were detected from a NP/OP PCR assay, or if bacteria (e.g., *Staphylococcus aureus*, *Haemophilus influenzae*, *S. pneumoniae*, and *S. pyogenes*) were detected in blood, endotracheal aspirate, or bronchoalveolar-lavage specimen by culture or, in pleural fluid, by culture or a PCR assay. For children, a bacterial pathogen was also defined as being present if bacteria were detected in whole blood by PCR for *S. pneumoniae* or *S. pyogenes*. For adults, a bacterial pathogen was also defined as being

present if *S. pneumoniae* or *L. pneumophila* was detected by the urine antigen assay. Procalcitonin was categorized as <0.1 μg/L, 0.1-<0.25 μg/L, 0.25-<0.5 μg/L, ≥0.5 μg/L [140], in which higher levels may correspond to bacterial infections and lower levels may correspond to viral infections [141, 142].

Serology for non-influenza respiratory viruses was performed using indirect enzyme immunoassays. PIV types 1-3 were combined for analysis due to antigenic cross-reactivity. For both influenza A and B viruses, HI assay was performed but because the HI assay is overly sensitive for influenza B viruses [143], serology samples that were positive for influenza B virus by the HI assay were further tested using the MN assay to improve specificity. If the influenza serology results indicated seroconversion when the vaccine was administered (based on self/caregiver report or vaccine verification) within 2 weeks before acute-phase serum collection, or between acute-phase and convalescent-phase serum collection, results were deemed inconclusive and were considered as missing serology results for these analyses.

2.3 Analytic Plan for Aims

Rationale for Aims 1 and 2

The most easily available pneumonia diagnostic tests are not from the lung, which makes it challenging to interpret results from these tests. An additional challenge in diagnosis of viral pneumonia has been the imperfect sensitivity and specificity of diagnostic tests and the lack of a true gold standard. Several studies have demonstrated that PCR has increased sensitivity in viral detections relative to culture [66, 67, 75-77] and can also provide results in a more clinically relevant time frame. In some cases, the

addition of serology to PCR can increase diagnostic yield for influenza [78, 83, 144] and other respiratory viral detections [80, 81]. Because some proportion of viral cases are undetected by PCR but captured by serology, PCR may not have perfect sensitivity. In addition, a proportion of viral cases may be undetected by serology but captured by PCR. Moreover, with the EPIC study, most patients enrolled had an available PCR results, but serology results were available for less than half of the patients. Ultimately, there is a risk of misclassification of pneumonia etiology attributable to influenza and other respiratory viruses due to missing diagnostic test results and imperfect test characteristics.

Timely and careful specimen collection is important to test performance. For example, a longer duration between onset of symptoms and the time to PCR specimen collection can lead to a lower likelihood of a positive result [75]. Studies have demonstrated that influenza detection by PCR and culture was higher within the 7 days after illness onset, as compared to 8-14 days after illness onset [77]. Careful specimen collection with adequate sampling is also vital. Furthermore, the additional diagnostic yield from serology is minimized if either the acute or convalescent serum samples are not collected (thus not paired specimens and cannot interpret results) or if both are collected but they are outside the appropriate time frame that is consistent with timing from acute infection.

Analytic Plan for Aim 1

The objective of the first aim was to estimate the revised proportions of CAP associated with AdV, HMPV, PIV, and RSV among children and adults hospitalized with CAP, accounting for missing test results using multiple imputation. For this analysis,

2222 children and 2259 adults were included, as they were enrolled, met final radiographic criteria for CAP based on a study radiologist review, and had samples available for both viral and bacterial testing. This analysis focused on AdV, HMPV, PIV, and RSV individually as both PCR and serology were done for these viruses. While not every patient had a NP/OP swab collected, 99% of children and 99.5% of adults had a NP/OP sample available; 49% of children and 40% of adults were missing serology results, largely due to lack of an available convalescent sample. Due to the missing serology data, it is possible that positive serologic results were undetected and thus the proportion of pneumonia cases associated with respiratory viruses might be underestimated.

Imputing missing data can minimize bias due to missingness, attempts to predict reasonable estimates, and maximizes the number of observations used for analysis. Methods of handling missing data vary based on the type of missing data. The three types of missing data are missing completely at random (MCAR), missing not at random (MNAR), and data that are neither MCAR nor MNAR (though this class is sometimes referred to as missing at random (MAR)) and each have different assumptions. When data are MCAR, the reason for missingness is completely random and the probability that an observation is missing is unrelated to any other measurable characteristic. This class has the strongest assumption and it is rare that data are MCAR. When data are MNAR, the probability that an observation is missing depends on information that is not observed. MNAR data indicates that they are missing in a systematic way and there is no universal method of handling the missing data properly. Finally, when data are MAR, the probability that an observation is missing is random within each stratum of observed

variables. It is easier for this assumption to be satisfied than it is for MCAR and there are options for handling these missing data. Related to these assumptions of missing data types is the idea of replacement. Depending on the type of missing data, subjects with missing data based on known variables can be randomly replaced by another subject based on the same combination of observed variables [145].

The main proposed imputation method was multiple imputation through regression. Performing multiple imputation can be considered an improvement over simple imputation. With multiple imputation, data are imputed multiple times to create multiple datasets; each dataset is analyzed and then results are combined into one analytic dataset. While simple imputation provides natural variability in the missing data, the added advantage of multiple imputation is the incorporation of uncertainty from estimating the missing data and thus also produces unbiased parameter estimates and approximately unbiased standard error [146]. The regression method requires a model with independent variables that are correlated with the missing variable and/or are explain some of the missingness to be fit for all observations with complete data. For observations missing data on a given variable, the predicted value of the missing variable is then estimated by plugging in the values of the observed variables and the estimated regression coefficients from this model – this represents the creation of one imputed dataset. This process is repeated to create multiple imputed datasets, as other regression models with reasonable regression coefficients can be fit. Once this is accomplished, these multiple imputed datasets are summarized into one overall dataset to produce estimates for inference.

For this analysis, we identified a priori variables to impute missing serology results, including age, sex, any bacterial detections, and the corresponding virus-specific NP/OP PCR result. We also assessed whether additional variables (including symptoms, underlying medical conditions, radiographic characteristics, and pneumonia severity) could be used to impute missing serology results. We then used observed and imputed PCR and serology results to create a binary outcome of any positive detection; a person had a positive detection if either PCR or serology results were positive. We then estimated a revised proportion of any positive detection for each virus.

Secondary Analysis of LCA:

LCA may estimate the proportion of CAP cases associated with AdV, HMPV, PIV, and RSV. The pneumonia-etiology latent variable may have three mutually exclusive classes: 1) pneumonia due to a specific virus, 2) pneumonia due to another virus, and 3) pneumonia not due to a virus. LCA models were constructed separately for children and adults. This latent class model were run a total of eight times, as there were four viruses of interest (AdV, HMPV, PIV, and RSV) and two age groups (children and adults). Due to the cross-reactivity of PIV 1-3, the strains for PIV were combined.

The latent variable of pneumonia etiology was constructed using a statistical model that included observed categorical variables. The model used to create the latent variable of pneumonia due to AdV, HMPV, PIV, and RSV contained test results from serology and from PCR (Figure 1.2). We categorized PCR and serology test results into 3 categories: negative for all viruses, positive for any virus except for the one of interest, and positive for the virus of interest regardless of whether another virus was detected. In

addition to using the diagnostic test results on a specific respiratory virus, we examined other potentially useful variables, including any bacterial detections, white blood cell count, and procalcitonin (adults only). Data were available on 60% of adults, and preliminary analyses indicated no meaningful differences in demographic and clinical characteristics between those with and without procalcitonin data available. A dichotomous variable of the timing between illness onset and sample collection for PCR was created using a cut-point of ≤7 days and >7 days; in the EPIC study, 79.4% of children and 70.8% of adults had NP/OP specimens within 7 days of illness onset. The LCA models for children included PCR result, serology result, and any bacterial detections. For adults, the LCA models included PCR result, serology result, and procalcitonin.

The interpretation of the latent class probabilities and the item-response probabilities is key in answering the questions of interest. As a latent class was labeled as pneumonia due to a specific virus, this latent class probability could be interpreted as the revised probability of having pneumonia due to a specific virus and could reflect the revised proportion of CAP with a given virus detected. Because LCA was performed for adenovirus, HMPV, parainfluenza, and RSV separately, these labels reflected the respiratory virus of interest. The item-response probabilities reflected the sensitivity of diagnostic test, including serology and PCR.

One strength of this analysis is the population-based nature of the surveillance study. The impact of selection bias is likely minimal for two reasons: 1) the EPIC study included all eligible pneumonia hospitalizations within each hospital's geographic catchment area, and 2) patients who were eligible were similar to those who were

enrolled. Another strength is the nearly systematic specimen collection and multiple pathogen testing. Because the EPIC study prospectively tested all patients for viruses and bacteria that can cause pneumonia, testing was not clinically driven. Another strength is the novel investigation of potential misclassification of pneumonia etiology due to respiratory viruses by accounting for missing test results via multiple imputation and when diagnostic tests are not perfectly specific. Another strength is the novel application of LCA to estimate a revised proportion of CAP associated with specific respiratory viruses. LCA accounts for the non-independence between the virus-specific NP/OP PCR and serologic test results and can reduce measurement error, especially when using both detection methods increases the diagnostic yield over using either method alone. Another strength of this analysis is the estimation of sensitivity and specificity for both detection methods, which can then be compared to those found in the literature.

There are limitations to this analysis. First, LCA relies on assumptions, including the relationship between and among observed and latent variables. Second, despite the robust dataset from the EPIC study, the only diagnostic tests available for each specific respiratory virus were NP/OP PCR and serology. Third, as we had only these two strong discriminators between latent classes, we included weaker discriminators that might distinguish between bacteria and viruses. Additionally, as detecting one pathogen does not mean that another pathogen cannot be detected, the proportion of co-detections in children (25%) and adults (5%), may have resulted in difficulty in identifying a class labeled as a specific respiratory virus. Fourth, the summary variable of any bacterial detections combined positive detections from multiple diagnostic tests, each of which has its own imperfect test characteristics, possibly introducing potential misclassification

which could not be accounted. Fifth, the detection of viruses in NP/OP swabs is not necessarily indicative of the cause of pneumonia [18, 21] and may represent resolving infection rather than acute infection or an infection limited to the upper tract and not lower tract [20]. Finally, illness onset was self or caregiver-reported, which introduces potential recall bias.

Analytic Plan for Aim 2

The objective of the second aim was to estimate the revised proportion of CAP due to influenza A/B viruses among children and adults hospitalized with CAP, accounting for missing test results using multiple. As the analytical approach was the same as that of the first aim, only the differences will be highlighted here. Among 2222 children, 99% had NP/OP specimens and 43% had paired serum specimens available. Among 2259 adults, 99% had NP/OP specimens and 38% had paired serum specimens available. Due to the missing serology data, it is possible that positive serologic results were undetected and thus the proportion of CAP cases associated with influenza A/B viruses might be underestimated.

For this analysis, we identified a priori variables to impute missing serology results, including age, sex, influenza PCR result, and any bacterial detections. We also assessed whether additional variables could be used to impute missing serology results. We then used observed and imputed PCR and serology results to create a binary outcome of any positive detection; a person had a positive detection if either PCR or serology results were positive. We then estimated a revised proportion of any positive detection for influenza viruses.

Secondary Analysis of LCA:

As with the first aim, the latent variable of pneumonia etiology could have three mutually exclusive classes: CAP due to influenza A/B viruses, CAP due to other respiratory viruses for which we tested, and CAP not due to a respiratory virus for which we tested. Models will be constructed separately for children and adults. Due to the low number of detections for each influenza type, influenza types A and B will be combined. For these models, the variables were categorized in the same way as in Aim 1, and the same variables were included in the final LCA models (Figure 1.3).

This aim had the same strengths and limitations as Aim 1.

Rationale for Aim 3

Pneumonia is a known complication of seasonal influenza [65, 105-108] and pandemic influenza [109-113]. Among hospitalized adults and children with influenza A(H1N1)pdm09 infection, patients with and without influenza-associated pneumonia had similar clinical presentation, but patients with pneumonia were more likely to have shortness of breath and diarrhea than those without pneumonia [107]. Additionally, radiographic findings of influenza A(H1N1)pmd09 pneumonia included bilaterial infilrates, infiltrate limited to one lobe, and multilobar infilrates [107, 109, 119, 120]. Early diagnosis and identification of influenza-associated pneumonia is crucial because of increased risk of severe outcomes [107, 113, 119] and need for timely initiation of antiviral therapy [107, 109, 120]. Because testing is clinically driven and clinical behaviors may shift, it is possible that cases of influenza-associated pneumonia might be

missed if radiographs and/or appropriate diagnostic testing are not performed. Thus, the use of signs, symptoms, and commonly used laboratory measurements could be helpful in predicting influenza-associated pneumonia if a clinical prediction model is developed. However, no formal clinical prediction model has been developed for pneumonia due to influenza specifically, including in a hospitalized population.

Analytic Plan for Aim 3

The objective of the third aim was to develop and evaluate prediction models and scores for influenza-associated pneumonia (using the outcome based on multiple imputation from Aim 2) among children and adults hospitalized with CAP. Within this aim, there were three specific sub-objectives: 1) to identify predictors of influenza-associated pneumonia (using the outcome based on multiple imputation), including symptoms and signs, using logistic regression to construct a prediction model and from this, 2) to develop prediction scores and 3) to evaluate the performance of the prediction models and scores and to perform an internal validation of the prediction models using bootstrapping. For this analysis, we used the same analytic population as that in Aims 1 and 2. Among these 2222 children and 2259 adults, 149 children (6.7%) and 132 adults (5.8%) had observed detections of influenza virus by NP/OP PCR or serology.

Potential Predictors

The potential predictors evaluated were reflective of those that are quickly available, which included symptoms and clinical signs, and initially excluded radiographic characteristics. Over 20 self-reported or caregiver-reported symptoms during the patient interview were evaluated as potential predictors of influenza-associated

pneumonia. In addition, clinical signs that were recorded from a hospital physical exam were also evaluated; such findings included altered mental status, cyanosis, chest indrawing (for children only), rales/crackles, rhonchi/coarse breath sounds, wheezing, dullness to percussion, egophony/increased resonance of voice sounds, and decreased breathing sounds. White blood cell count, a commonly available laboratory finding available at presentation, was also assessed. The presence of any underlying medical conditions as well as specific underlying medical conditions were evaluated. Demographic characteristics of race/ethnicity, sex, and age groups within children and within adults were also evaluated as potential predictors. Influenza season was also assessed as a potential predictor. Finally, radiographic characteristics of consolidation, other infiltrates, or pleural effusion and the verified receipt of the influenza vaccine were evaluated only after the multivariable models were constructed.

Outcome Definition of Influenza-associated Pneumonia based on Multiple Imputation

Within Aim 2, we performed multiple imputation for missing serology results for influenza virus in children and adults separately. Our final multiple imputation model for children included age, sex, any bacterial detection, influenza PCR result, and influenza season. For adults, our final model included age, sex, any bacterial detections, influenza PCR result, and self-reported abdominal symptoms. Using these final variables, we applied multiple imputation to impute missing serology results. We combined observed and imputed results to create a binary outcome of influenza-associated pneumonia, which was defined as either positive or negative for influenza virus. A person was classified as having influenza-associated pneumonia if results from PCR or serology were positive; a person was classified as negative otherwise.

Modeling Strategy

Bivariate analyses were conducted for each potential predictor with the dichotomous outcome of influenza-associated pneumonia using logistic regression. Variables with a $p<0.20$ were eligible for inclusion in the multivariable prediction model. The multivariable prediction model were constructed from a forward-building approach using logistic regression until all retained variables had $p<0.05$. Biological plausibility was also considered in selecting the variables retained in the final model. Interaction was assessed for a limited number of variables based on a priori knowledge and the literature. To ensure consistent variables were retained in the final model, a backwards elimination approach was also used to eliminate one of the eligible variables at a time that was the least non-statistically significant until all retained variables had $p<0.05$, which also utilized logistic regression. Prediction models were constructed for children and adults separately. After the models were constructed, we also examined whether the inclusions of radiographic characteristics and verified receipt of the influenza vaccine improved the predictive performance of the model. To develop an influenza-associated pneumonia score, a point value was assigned to each predictor in the multivariable model by rounding each beta coefficient to the nearest integer.

Following the construction of the prediction models and scores, their performances were evaluated using the measures of discrimination and calibration. Discrimination measures how well the model distinguishes between patients with and without the outcome of interest and was evaluated using the c-statistic, the area under the receiver operating characteristic (ROC) curve. Values range between 0.5 (no discrimination) and 1.0 (perfect discrimination). Calibration measures how well the

model fits the data and was evaluated using the Hosmer-Lemeshow goodness of fit test. For each numerical score, sensitivity, specificity, positive predictive value, and negative predictive value were calculated.

Validation of a prediction regression model is important, as a prediction model can perform well in one population though can fail to predict well when validated in another; this phenomenon describes optimism. Rather than using an external dataset, an internal validation could be performed using cross-validation or bootstrapping; the application of the latter may be more appropriate in this study, as optimism can be assessed using bootstrapping. Bootstrapping draws a random sample with replacement within the original sample. Typically, bootstrapping is typically replicated at least 300 times in order to provide stable estimates. The average difference in the c-statistic between the bootstrap sample and the original sample indicates the optimism [147, 148].

Sensitivity Analysis using outcome definition using observed PCR and serology results:

We defined the outcome of influenza-associated pneumonia based only on observed PCR and serology results. We classified individuals with and without the outcome the same way as we did in the main analysis. We applied the same methodology for developing and evaluating the prediction models.

Once the primary and sensitivity analyses were completed, the predictors of influenza-associated pneumonia from both analyses were compared to determine any differences using the outcome definitions of influenza-associated pneumonia from the primary and sensitivity analyses.

In addition to the other strengths, another strength is that the EPIC study prospectively tested all enrolled patients for influenza, and thus testing was not clinically driven, which may allow for a less biased assessment of the clinical and radiographic presentation of influenza-associated pneumonia. Another strength is the breadth of symptoms and radiographic evidence available in the EPIC dataset, which will allow for a more complete assessment of which clinical characteristics predict the outcome.

A limitation is that 54% of children and 61% of adults were missing influenza serology, so it is possible that some proportion of influenza A/B viruses were undetected, which could result in misclassification. A limitation of the analysis is the self-reported or caregiver-reported symptoms. In particular, young children or older adults may not have the ability to verbally express their symptoms and the caregivers are relied upon for the report of the symptoms, which may not be accurately capture what symptoms were experienced. Another limitation is the potential for inter- and intra-observer variability in the final determination of radiographically-confirmed pneumonia.

Software Program Used for Analysis:

SAS was used for data management, cleaning, and analysis. Mplus was used for LCA.

Section 3. Novelty, Significance, and Impact of the Dissertation

CAP is a common cause of hospitalization among all ages. Multiple bacterial and viral respiratory pathogens contribute to pneumonia etiology and can be detected through a variety of specimens and methods. In the EPIC study, systematic collections of

multiple specimens and use of the most current bacterial and viral diagnostic tests available were attempted, but there were limitations, including inability to obtain lower respiratory tract samples and inability to obtain every sample in each individual. The estimated proportion of pneumonia cases associated with specific pathogens is dependent on the detection methods used and the type and quality of collected specimens. Even with systematic testing using multiple detection methods, there is the potential for biased estimates of the proportion of pneumonia cases associated with specific pathogens due to imperfect characteristics of each detection method – an issue commonly discussed but its implications have not been thoroughly investigated.

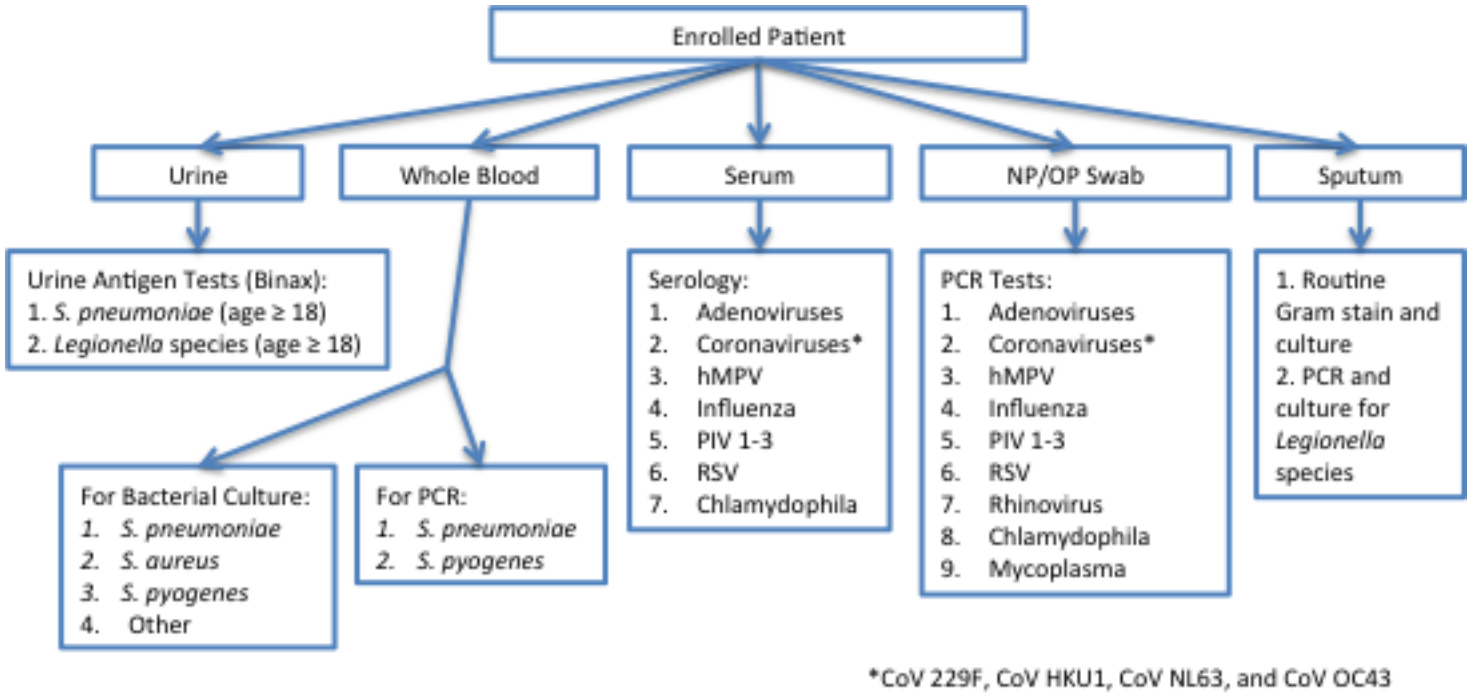
This was the first investigation into the potential misclassification of pneumonia etiology attributable to respiratory viruses from missing diagnostic test results and from imperfect test characteristics, especially in a hospitalized population. Specifically, to our knowledge, this was the first time that revised viral-specific pneumonia burdens have been estimated by applying multiple imputation to account for missing diagnostic test results.

This dissertation also included a novel application of LCA in estimating revised proportions of CAP with specific respiratory viruses detected; LCA has been previously applied to other infectious diseases, including pneumococcal pneumonia [59, 130], but not to viral-specific pneumonia. Though the use of LCA addresses misclassification of diagnostic tests, this particular application aimed to provide a revised estimate of the proportion of CAP associated with specific pathogens in addition to accounting for the timing between illness onset and PCR specimen collection, a factor not previously considered in estimating disease prevalence. Accounting for such misclassification could

have elucidated improved estimates of the viral-specific pneumonia burdens among hospitalized children and adults, which could help public health policy and clinical guidance.

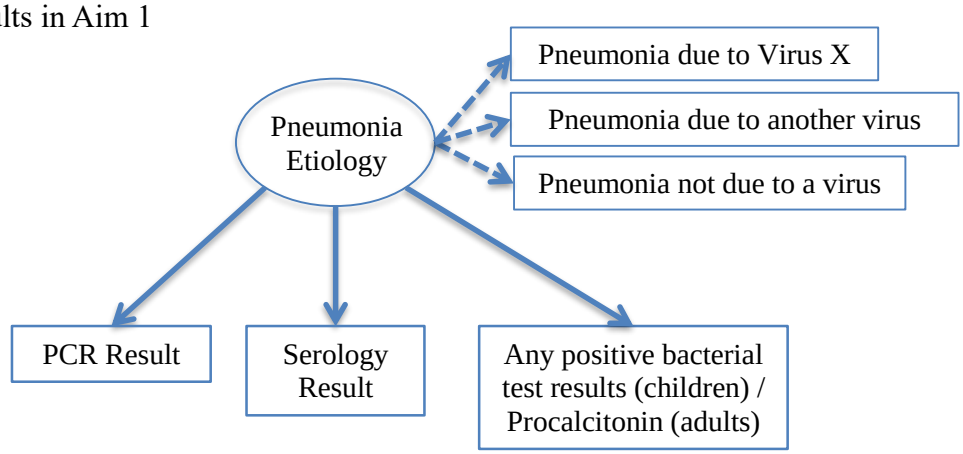
Early diagnosis and identification of influenza-associated pneumonia is crucial because of increased risk of severe outcomes [107, 113, 119] and need for timely initiation of antiviral therapy [107, 109, 120]. Because testing is clinically driven and clinical behaviors may shift, it is possible that cases of influenza-associated pneumonia might be missed if radiographs and/or appropriate diagnostic testing are not performed. Thus, the use of symptoms and radiographic characteristics could be maximized if a clinical prediction model is developed to predict influenza-associated pneumonia. However, no formal clinical prediction model has been developed for influenza-associated pneumonia, particularly in a hospitalized population. This dissertation also focused on the development and evaluation of prediction models and scores using a breadth of symptoms and radiographic evidence. Thus, the identification of meaningful combination(s) of symptoms and radiographic evidence could be more easily adopted in a clinical setting to inform diagnostic testing and appropriate antimicrobial therapy.

Figure 1.1. Modified Specimen Collection and Laboratory Testing Flow Chart



Modified from the Incidence and Etiology of Influenza-Associated Community-Acquired Pneumonia in Hospitalized Persons Study Protocol Version 032910.

Figure 1.2. Latent class model for non-influenza respiratory viruses among children and adults in Aim 1



Virus X is a placeholder for AdV, HMPV, PIV, and RSV, as LCA will be performed separately for each of these four viruses. The latent variable of pneumonia etiology is indicated with circle and the observed variables are indicated with squares. The dotted lines emanating from the latent variable are the proposed labels for the latent classes.

Figure 1.3. Latent class model for influenza A/B viruses among children and adults in Aim 2

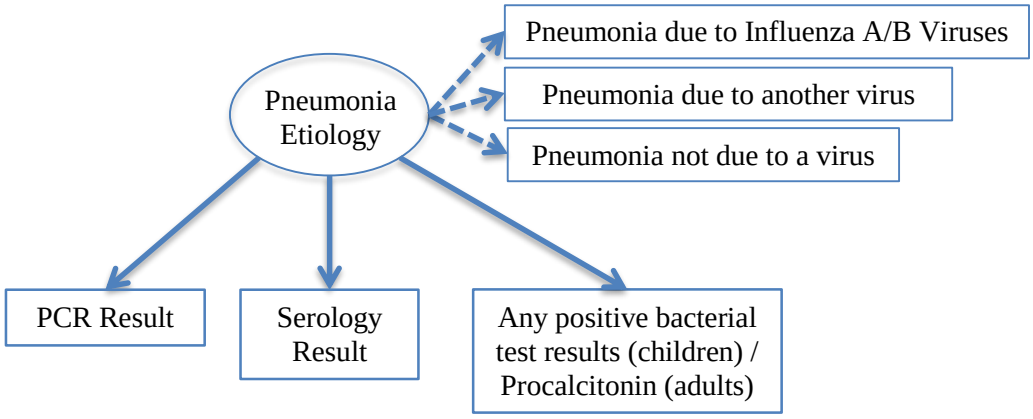


Table 1.1. Possible Detection Methods of Pathogens Included in the EPIC Study

Pathogen	Classification	Microscopy and Culture	Whole Blood Culture	Detection of Antigen in Urine and Respiratory Specimen	Detection of Antibodies in Serum	Nucleic Acid Tests
S. pneumoniae	Gram-positive	X	X	X		X
S. aureus	Gram-positive	X	X			
S. pyogenes	Gram-positive	X	X			X
H. influenzae	Gram-negative	X	X			
Legionella species	Gram-negative	X		X	X	X
M. pneumoniae	Neither Gram-positive or Gram-negative	X			X	X
C. pneumoniae	Gram-negative				X	X
Adenovirus	ds DNA	X	X	X	X	X
HMPV	ss RNA	X		X	X	X
Influenza A, B	ss RNA	X	X	X	X	X
PIV 1, 2, 3	ss RNA	X	X	X	X	X
RSV	ss RNA	X	X	X	X	X
Rhinovirus	ss RNA	X	X	X		X

NR=not reported
ss = single-stranded, ds = double-stranded

Table 1.2. Sensitivity and Specificity of Each Detection Method as well as its Advantages and Disadvantages			
Method Type	Sensitivity; Specificity of a pathogen in a diagnostic test	Advantages of Detection Method	Disadvantages of Detection Method
Microscopy and Culture	*S. pneumoniae* Gram stain: 82%; 93% [149]; 57%; not reported [38] *Staphylococci* Gram stain: 76%; 96% [149] Gram-negative bacteria Gram stain: 79%; 96% [149]	Gram Stain: rapid; helps to characterize detected pathogen; very specific with high quality of specimen [150]; sputum Gram stain broadens initial empiric coverage for less common etiologies; validate sputum culture results [10]	Does not reveal organisms that do not stain; sputum specimen frequently unreliable due to difficulty in high-quality specimen collection, especially in children [21, 151]; low sensitivity in patients who have received previous antibiotic treatment; can't perform antibiotic susceptibility testing [10, 149]
	S. pneumoniae Sputum Culture: 86%; not reported [10]; 79%; not reported [38] *Legionella* species Culture: <10-80%; 100% [152] RSV Culture: 36%; not reported [77] Influenza Culture: 59%; not reported [77] PIV Culture: 58.5%; not reported [77]	Culture: low cost; high specificity; allows for susceptibility testing and serotyping [151]; high sensitivity and specific for pneumococcal pneumonia in adults not exposed to antibiotics before sputum collection [19]	Wait 24-48h for results; can be contaminated by oropharyngeal flora; low sensitivity in patients who have received previous antibiotic treatment; tendency for *S. pneumoniae* to autolyse once reaching the stationary growth phase ; diagnostic yield of sputum culture is variable and is affected by the quality of specimen collected, transport, rapid processing, absence of prior antibiotic therapy, and skill in interpretation [10]
Whole Blood Culture	*S. pneumoniae:* <30%; not reported [60, 151]	Identified organisms are considered to be cause of pneumonia [21]; recommended for patients with severe CAP [10]	Low sensitivity and diagnostic yield, especially in children and in patients who have received previous antibiotic treatment; wait 24 hours for results [15, 150, 151]; possible limited utility due to low prevalence of bloodstream infections [19]

Detection of Antigen in Urine and Respiratory Specimen	RSV Rapid: 71-95%; 80-100% [19] *L. pneumophila* DFA: 25-70%; >95% [152] *L. pneumophila* Urinary Antigen Test: 70-90%; >99% [152] *S. pneumoniae* Urinary Antigen Test (adults only): 88%; 96% [33]; 65.9%; 100% [32]; 70.4%; 89.7% [30]; 64.3%; 98.8% [29]; 80.4%; 97.2% [34] Rapid Influenza Diagnostic Test: 44%; 100% [70]; 70%-82%; 100% [71]; 88%; 93% [72]; 27%; 97% [73]	Rapid Tests: High specificity; able to distinguish between influenza A and B; rapid results; possibly reduced use of antibacterial agents [10] UAT: ease of getting specimen; improvement of diagnostic yield; able to detect pathogen after antibiotic treatment **;** detects all pneumococcal strains (as C polysaccharide antigen is in all pneumococcal serotypes [31]; detection of *Legionella* improves likelihood of identifying environmental source; rapidity; simplicity; able to detect pneumococcal pneumonia after administration of antibiotic therapy [10]	Rapid Tests: Cost; high rates of false-negative test results; false-positive assays with adenovirus infections [10] UAT: *(S. pneumoniae)*: proportion of patients will have positive blood or sputum cultures and negative antigen tests; antigen test may cross-react with other streptococci; UAT can be positive for weeks after disease onset; UAT is unreliable in children due to false-positive in detecting pneumococcal carriage ; sensitivity and specificity are less in adults with non-bacteremic pneumonia [21, 22]; (*Legionella)*: *Legionella* urinary antigen only identifies *L. pneumophila* serogroup 1 (which causes 80% of sporadic cases of Legionnaire's disease) [152]
Detection of Antibodies in Serum	Influenza MN assay: 94%; 83% [82]; 83%; 86% [83] Influenza HI assay: 75%; 97% [82]; 60%; 94% [83]	Useful in detecting fastidious organisms and for pathogens that have prolonged shedding in the nasopharynx or are highly prevalent in a control population [21]; helpful in some series for *Legionella*, *Mycoplasma*, *Chlamydophila*; widely available and relatively simple [15]	Results not available for weeks due to requiring convalescent serum; major limitation is single measurements of acute-phase serum specimens lack sensitivity [19, 93]; potentially useful in diagnosing secondary bacterial infections from initial viral infection [19]

Nucleic Acid Tests	*S. aureus* (PCR): 97%; 92% [149]; 96.2%; 93.2% [153] *H. influenzae* (PCR): 95.8%; 95.4% [153] *S. pyogenes* (PCR): 100%; 100% [153] *M. pneumoniae* (PCR): 100%; 95.4% [153] RSV PCR: 93.6%; not reported [77] Influenza PCR: 92.9%; not reported [77] PIV PCR: 85.7%; not reported [77]	Rapid availability; high throughput, less labor intensive, more cost-effective [19, 153]; improved sensitivity in patients who are taking antimicrobial drugs for diagnosis of pneumococcal pneumonia [59]; possibly use multiplex platform to detect multiple pathogens in one assay; apply molecular serotyping; detect pathogen and/or determine antibiotic susceptibility ; detect low levels of nucleic acid from respiratory pathogens; don't depend on viability of target microbe; results available in clinically relevant time frame; less affected by prior antibiotic administration [19]	Not readily available; culture is still required to obtain an isolate for antimicrobial-susceptibility testing [153]; lack of suitable comparator gold standard; very high analytic sensitivity may not guarantee high clinical sensitivity [19]; cost; lack of adequate specimen in respiratory tract; limited data on quantitation thresholds to define significance; interpretation if multiple pathogens are detected [22]

Table 1.3. Review of Michelow and Marston Study Results

Pathogen	Number and Proportion of Pneumonia Cases Attributable to Detected Pathogen in Children from Michelow Study N (%) (N=154)				Proportion of Pneumonia Cases Attributable to Detected Pathogen in Adults from Marston Study N (%) (N=2776)
	No Co-infection	Co-infection with Bacteria*	Co-infection with Viruses*	Total Number of Episodes	
S. pneumoniae	35 (22.7%)	12 (7.8%)	21 (13.6%)	68 (44.2%)	351 (12.6%)
S. aureus	0 (0%)	2 (1.3%)	0 (0%)	2 (1.3%)	94 (3.4%)
S. pyogenes	0 (0%)	2 (1.3%)	2 (1.3%)	2 (1.3%)	17 (0.5%)
H. influenzae	Not reported	Not reported	Not reported	Not reported	184 (6.6%)
Legionella species	Not reported	Not reported	Not reported	Not reported	63 (3%)
M. pneumoniae	11 (7.1%)	6 (3.9%)	8 (5.2%)	21 (13.6%)	404 (32.5%)
C. pneumoniae	6 (3.9%)	7 (4.5%)	7 (4.5%)	14 (9.1%)	172 (8.9%)
Adenovirus	2 (1.3%)	9 (5.8%)	5 (3.2%)	11 (7.1%)	Not reported
HMPV	Not reported	Not reported	Not reported	Not reported	Not reported
Influenza A, B	9 (5.8%) (A),	16 (10.4%),	10 (6.5%),	26 (16.9%),	58 (7.4%) (season for A),
	1 (0.6%) (B)	6 (3.9%)	6 (3.9%)	7 (4.5%)	17 (2.2%) (season for B)
PIV 1, 2, 3	6 (3.9%)	12 (7.8%)	10 (6.5%)	20 (13%)	Not reported
RSV	6 (3.9%)	11 (7.1%)	8 (5.2%)	20 (13%)	14 (3.1%) (season),
					3 (1.4%) (off-season)
Rhinovirus	1 (0.6%)	2 (1.3%)	2 (1.3%)	5 (3%)	Not reported

* The categories of co-infection with bacteria and with viruses are not mutually exclusive

From: Michelow IC, Olsen K, Lozano J, Rollin NK, Duffy LB, et al. (2004) Epidemiology and clinical characteristics of community-acquired pneumonia in hospitalized children. Pediatrics 113: 701-707.

Marston BJ, Plouffe JF, File TM, Jr., et al. (1997) Incidence of community-acquired pneumonia requiring hospitalization: Results of a population-based active surveillance study in Ohio. Arch Intern Med 157: 1709-1718.

CHAPTER 2. Use of multiple imputation and other methods to calculate revised proportions of non-influenza respiratory viral detections among children and adults hospitalized with community-acquired pneumonia

Abstract

Background: Real-time polymerase chain reaction (PCR) and serology are detection methods for respiratory viruses, though limitations of these tests can lead to misclassification of pneumonia etiology attributable to respiratory viruses. Specifically, due to missing test results and imperfect test characteristics, revised estimates of viral-specific community-acquired pneumonia burdens are needed.

Methods: We analyzed data from an active population-based surveillance study for community-acquired pneumonia requiring hospitalizations among children (<18 years old) and adults. Adenovirus, human metapneumovirus (HMPV), parainfluenza virus types 1-3, and respiratory syncytial virus (RSV) were considered detected if PCR or serology results were positive, assuming no false-positive results. Multiple imputation was applied to impute missing serology results for each of these viruses individually. The revised proportions of pneumonia due to a given virus were estimated based on observed and imputed results from PCR and serology.

Results: Among 2222 children and 2259 adults with radiographically-confirmed pneumonia, 98.8-99.5% had NP/OP specimens and 39.7-48.6% had paired serum specimens available. By accounting for missing serology results, the revised proportions for adenovirus, HMPV, parainfluenza virus, and RSV detections ranged from 8.9-29.9%

among children and from 4.0-4.7% among adults, which were 0.8-3.2% higher than the observed estimates which had missing serology results.

Conclusions: The proportion of community-acquired pneumonia with these respiratory viruses detected using observed results may have underestimated the virus-specific burdens.

Introduction

Community-acquired pneumonia (CAP) is a common cause of hospitalization among all ages in the United States [154, 155]. Both bacterial and viral respiratory pathogens can cause pneumonia and can be detected using a broad range of diagnostic tests. Timing of collection and quality of specimens affects diagnostic test performance [75]. Further, the specimens used for diagnostic testing are usually not from the site of infection (e.g., lung tissue), making the interpretation of results more challenging. An additional challenge is the imperfect sensitivity and specificity of currently available diagnostic tests for respiratory pathogens and lack of a true gold standard [78, 83, 156]. PCR and serology are detection methods for respiratory viruses, though serology requires collection of paired serum specimens and thus has limited impact on patient management. These limitations of currently available diagnostic tools can lead to misclassification of pneumonia etiology attributable to respiratory viruses in pneumonia etiology burden studies.

Real-time polymerase chain reaction (PCR) has improved sensitivity for viral detections relative to culture [66, 67, 75-77], and the addition of serology to PCR has also increased diagnostic yield for respiratory viral detections in research studies [80, 81]. The Centers for Disease Control and Prevention (CDC) Etiology of Pneumonia in the Community (EPIC) study used PCR and serology for detection of respiratory viruses in a multi-center active surveillance study of the incidence and etiology of CAP among hospitalized U.S. adults and children [157, 158]. Adenovirus, human metapneumovirus (HMPV), parainfluenza virus types 1-3, and respiratory syncytial virus (RSV) were detected using both PCR and serology, which was our motivation to study them. While

most patients enrolled in the EPIC study had an available PCR result, serology results were available for less than half of the patients. We used multiple imputation to assign values for missing serology test results based on data from patients with available test results. The goal of the study was to estimate the revised proportions of CAP due to adenovirus, HMPV, parainfluenza virus types 1-3, and RSV among children and adults hospitalized with CAP, accounting for missing serology test results using multiple imputation.

Materials and Methods

Enrollment, Specimen Collection, and Laboratory Methods

The CDC EPIC study was a prospective, multi-center, population-based, active surveillance study, as previously described in detail elsewhere [157, 158]. Briefly, from January 1, 2010 to June 30, 2012, children <18 years of age were enrolled in three pediatric hospitals in Memphis, TN, Nashville, TN, and Salt Lake City, UT, and adults were enrolled in three hospitals in Chicago, IL, and two hospitals in Nashville, TN. Patients admitted to a study hospital were eligible if they resided within the hospital catchment area and had evidence of acute respiratory infection and radiographically-confirmed pneumonia within 72 hours of admission. Patients were included in the etiologic analysis if they met final radiographic criteria for CAP based on a study radiologist review and had samples available for both bacterial and viral testing. Clinical, demographic, and epidemiologic information were collected through interviews and medical chart review. Informed consent was obtained and the study protocol was approved by the institutional review boards at each institution and the CDC.

Different specimen types were collected for comprehensive diagnostic testing of respiratory pathogens from enrolled patients [157, 158]. Whole blood, acute serum, and naso-oropharyngeal (NP/OP) swabs were collected as soon as possible after enrollment; urine was only collected in adults. For adults with a productive cough, sputum was also collected; results from high-quality sputum samples were included. Pleural fluid, endotracheal aspirates, and bronchoalveolar-lavage specimens when collected in select cases for clinical care were included in the analysis. Convalescent serum was collected 3-10 weeks after enrollment.

Real-time PCR was performed on NP/OP swabs for detection of multiple viruses and atypical bacteria (e.g., *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae*) [157, 158]. Serologic testing for viruses was performed on available paired acute and convalescent serum specimens. Available adult serum specimens were analyzed to quantify the biomarker procalcitonin [62, 63]. Bacterial culture was performed on specimens from blood, sputum, pleural fluid, endotracheal aspirates, and bronchoalveolar-lavage. To test for specific bacteria, PCR assays were performed on pleural fluid and, in children only, on whole blood. Urine antigen testing was performed for *Legionella pneumophila* and *Streptococcus pneumoniae*.

Categorization of Variables Based on Diagnostic Test Results

A positive PCR test for adenovirus, coronavirus, HMPV, influenza A/B viruses, parainfluenza virus types 1-3, RSV, or rhinovirus was defined as a cycle threshold value of <40 from a NP/OP PCR assay. A positive serologic test for adenovirus, HMPV, influenza A/B viruses, parainfluenza virus, or RSV was defined as a ≥4-fold rise in agent-specific IgG antibody titer between paired acute and convalescent sera. We categorized

viral tests dichotomously as either positive or negative using definitions described above. In the EPIC study and for these analyses, a virus was considered to be detected if there was a positive result from either PCR or serology, except for rhinovirus and coronaviruses which only depended on PCR results. This analysis focuses on adenovirus, HMPV, parainfluenza, and RSV, which were tested by both PCR and serology.

A bacterial pathogen was defined as being present if *Chlamydophila pneumoniae* or *Mycoplasma pneumoniae* was detected in the NP/OP swab by a PCR assay, or if bacteria (e.g., *Staphylococcus aureus, Haemophilus influenzae*, *S. pneumoniae,* and *S. pyogenes*) were detected from blood, endotracheal aspirate, or bronchoalveolar-lavage specimen by culture or, in pleural fluid, by culture or a PCR. For children, a bacterial pathogen was also defined as being present if bacteria were detected in whole blood by PCR for *S. pneumoniae* or *S. pyogenes*. For adults, a bacterial pathogen was also defined as being present if *S. pneumoniae* or *L. pneumophila* was detected by the urinary antigen test. A summary variable of any bacterial detection was created; no bacterial detections were observed if none of these criteria were met.

Procalcitonin was categorized as <0.1 μg/L, 0.1-<0.25 μg/L, 0.25-<0.5 μg/L, ≥0.5 μg/L [140], in which higher levels may correspond to bacterial infections and lower levels may correspond to viral infections [141, 142]; this variable was only used in a secondary analysis presented in the Supplementary Appendix.

Statistical Analysis

Multiple Imputation

We performed multiple imputation for missing serology results for adenovirus, HMPV, parainfluenza, and RSV in children and adults separately. Multiple imputation

uses variables within non-missing data, in this case data from patients who had serology results available, to impute missing data, in this case, data from patients who did not have serology results available, via a multivariable regression model [145, 146, 160]. To assess whether these data were missing at random [145, 146], we compared demographic and clinical characteristics among patients with and without available serology results.

Using both bivariate and multivariable analyses, we identified independent variables that could be used to impute whether missing serology results for each respiratory virus as either positive or negative. First, we constructed "*a priori*" models using age, sex, any bacterial detection, and the corresponding viral PCR results chosen *a priori* based on biological plausibility. We hypothesized that a given virus's PCR results would be positively correlated with its corresponding serology results, given the concordant results from both methods. Additionally, we hypothesized that any bacterial detection would be negatively correlated with serology results for a given virus, as patients with a positive serology result for a given virus were less likely to have any bacteria detections compared to those with negative serology results for a given virus. Second, we assessed whether additional variables, including cough, fever, diarrhea, sore throat, abdominal pain, myalgia, chills, headache, chest indrawing/retraction, underlying medical conditions, radiographic findings (e.g., consolidation, pleural effusion, and other infiltrates), ICU admission, and pneumonia severity (defined as having experienced at least one of the following: ICU admission mechanical ventilation, acute respiratory distress syndrome, shock, or death), were significant in bivariate analyses. For the variables that were significant in the bivariate analyses, each variable was added one at a time to the *a priori* model to assess whether it was significant in the multivariable model.

Model fit was evaluated using the Hosmer-Lemeshow goodness of fit test; a p-value <0.05 was used to indicate that model did not fit data well. Additionally, we compared the area under the curve between the *a priori* model with and without the potential variable. A non *a priori* variable was used to impute serology data if it was significant in both the bivariate and multivariable analyses and the area under the curve was higher for the *a priori* model with this variable relative to the model without it. Potential variables were assessed for each virus separately and for children and adults separately, thus different variables could have been used for multiple imputation for the separate models.

For children, our final multiple imputation model for adenovirus, parainfluenza, and RSV included age, sex, any bacterial detections, and the corresponding viral PCR result; the model for HMPV included these variables and the presence of any underlying medical condition. For adults, our final multiple imputation models for adenovirus, HMPV, and parainfluenza included age, sex, any bacterial detection, and the corresponding viral PCR result; the model for RSV included these variables and wheezing. Using these final variables, we applied multiple imputation to impute missing serology results. Twenty imputed datasets were created for improved efficiency [161]. We combined observed and imputed results from PCR and serology to create a binary outcome, which was defined as either positive or negative for a given virus. We then estimated a revised proportion of any positive result for each virus and its corresponding 95% confidence interval (CI), referred to as the proportion of CAP with a given virus detected. We also performed sensitivity analysis in which the multiple imputation model excluded patients for whom NP/OP specimens were collected after 28 days from illness

onset, as the likelihood of a positive PCR results can decrease with a longer duration between illness onset and NP/OP specimen collection [75, 77].

We compared the revised proportions to the observed EPIC study results in terns of absolute difference by subtracting the proportion observed in the EPIC study from the proportion estimated from multiple imputation. Relative difference was calculated by dividing the proportion estimated from multiple imputation by the proportion observed within the EPIC study. Finally, among patients who had both PCR and serology results available, diagnostic yield of serology was calculated. We applied the formula: [(serology positive/PCR negative specimens) ÷ (specimens positive by both methods + serology negative/PCR positive specimens)]; diagnostic yield was calculated separately for children and adults.

Secondary Analysis: Latent Class Analysis

Latent class analysis (LCA) was also considered for estimating revised proportions of CAP due to influenza. However, LCA proved to be unstable and potentially misleading. Specifically, 25% of children and 5% of adults likely had more than one pathogen detected, modeled latent classes were difficult to interpret, and the required assumptions for LCA did not appear to be supported including absence of three separate moderately predictive diagnostic tests. Furthermore, results were sensitive to model selection strategy. LCA methods, including the assumptions, details of the model results, and a discussion of results are described in the supplementary appendix.

Bivariate and multivariable regression analyses and multiple imputation were performed using SAS 9.3 (Cary, NC). LCA was performed in Mplus 7.

Results

Children

Of 2638 enrolled children, 2358 (89.4%) met the final radiographic criteria for CAP, among whom 2222 (94.2%) had samples available for both bacterial and viral diagnostic tests. Median age was 2 years (interquartile range [IQR] 1-6); 50.9% of children had an underlying medical condition, including 33.4% that reported asthma (Table 2.1). Among these 2222 children, 2196 (98.8%) had NP/OP specimens and 1081 (48.6%) had paired serum specimens available. We found no meaningful differences in the demographic and clinical characteristics among children with and without available serology data (Table S2.1 in the Supplementary Appendix). For 2190 (98.6%) children who had an available NP/OP swab and illness onset data, the median time between illness onset and specimen collection was 4.6 days (IQR 2.8-7.4).

For all 2222 children, the revised proportions of CAP with adenovirus, HMPV, parainfluenza, and RSV detected from multiple imputation were 14.4%, 15.0%, 8.9%, and 29.9%, respectively; these were 1.6-3.2% higher in absolute differences and were 1.1-1.3 times higher in relative terms for each virus than the observed EPIC study results (Table 2.2). When limiting multiple imputation to the 2151 (96.8%) children who had NP/OP specimens collected within 28 days of illness onset, the revised proportions for adenovirus, HMPV, parainfluenza, and RSV were within 0.1-1.3% of those among all children. The contribution of serology-positive and PCR-negative detections above any PCR-positive detections results in a diagnostic yield of serology for adenovirus, HMPV, parainfluenza virus, and RSV of 29.9%, 25.2%, 61.1%, and 10.3%, respectively, among

955 (43.0%) children who had both PCR and serology results available. Results from LCA are described in the supplement (Table S2.2 in the Supplementary Appendix).

Adults

Of 2488 enrolled adults, 2320 (93.2%) met the final radiographic criteria for CAP, among whom 2259 (97.3%) had specimens available for bacterial and viral diagnostic testing. Median age was 57 years (IQR 46-71); 78.2% of adults had an underlying medical condition (Table 2.1). Among these 2259 adults, 2248 (99.5%) had NP/OP specimens and 897 (39.7%) had paired serum specimens available. Data on procalcitonin were available for 1339 (59.3%) adults. There were no meaningful differences in demographic and clinical characteristics between adults with and without available serology data (Table S2.3 in the Supplementary Appendix) and patients with and without procalcitonin data available (Table S2.4 in the Supplementary Appendix). For 2246 (99.4%) adults that had a NP/OP swab and illness onset data available, the median time between illness onset and specimen collection was 4.8 days (IQR 2.8-8.8).

For all 2259 adults, the revised proportions of CAP who had adenovirus, HMPV, parainfluenza, and RSV detected from multiple imputation were 4.2%, 4.7%, 4.0%, and 4.0%, respectively; these were 0.8-3.2% higher for each virus in absolute differences and 1.2-3 times higher in relative terms than the observed results (Table 2.3). When limiting the multiple imputation model to the 2142 (94.8%) adults who had NP/OP specimens collected within 28 days of illness onset, the revised proportions for adenovirus, HMPV, parainfluenza, and RSV were within 0.1-1.9% of those among all adults. The contribution of serology-positive and PCR-negative detections above any PCR-positive detections results in a diagnostic yield of serology for adenovirus, HMPV, parainfluenza

virus, and RSV of 133.3%, 24.3%, 35.5%, and 50.0%, respectively, among 833 (36.9%) adults who had both PCR and serology results available. Results from LCA are described in the supplement (Table S2.5 in the Supplementary Appendix).

Discussion

We estimated revised proportions of CAP with adenovirus, HMPV, parainfluenza, and RSV detected by applying multiple imputation to account for missing serology results, using prospectively collected clinical and microbiological data from patients hospitalized with CAP enrolled in the EPIC study. To our knowledge, this is the first time that multiple imputation has been applied to estimate revised viral-specific burdens in hospitalized CAP. Multiple imputation estimates of the revised proportions for adenovirus, HMPV, parainfluenza, and RSV detections were 0.8-3.2% higher in absolute differences and 1.1-3.0 times higher in relative terms than those estimated from the EPIC study among children and adults. Our results illustrate the potential underestimation of virus-specific pneumonia burden when serology results are either not included or missing.

More than half of enrolled children and adults were missing serology results, largely due to the lack of a convalescent sample. While obtaining convalescent serum samples is important for etiologic studies of CAP, there are challenges to this exercise. Patients are lost to follow-up, and the potential methods required to improve follow-up and thus increase specimen collection require intense resources that are often not available. Thus, using modeling methods, such as multiple imputation, could be informative in estimating virus-specific burdens. Multiple imputation requires an

assumption that the data are missing at random. As we found no meaningful differences between characteristics between patients with and without available serology results, we believed that this assumption was valid. We found minimal differences in revised estimates when multiple imputation was restricted to include NP/OP specimens collected within 28 days of illness onset; these estimates are likely similar because 3.2% of children and 5.2% of adults were excluded in this sensitivity analysis.

The revised estimates obtained using multiple imputation were consistently higher than the observed estimates from the EPIC study, as missing results can misclassify pneumonia due to respiratory viruses. Among 955 (43.0%) children and 833 (36.9%) adults who had both PCR and serology results available, 0.2-7.4% were PCR-positive and serology-negative. PCR may detect respiratory viruses not captured by serology because PCR detects the presence of viruses in the naso-/oropharynx whereas serology measures the virus-specific antibody response. Another explanation for the higher revised estimates, even after accounting for missing serology results, is the likelihood that serology detects viruses that are not captured by PCR, as 1.0-3.5% of these results were PCR-negative and serology-positive. One possible reason for this is because the convalescent serum sample is obtained 3-7 weeks after the acute sample (thus the timing is not precisely the same as when samples for PCR were collected), and the positive result could represent a subsequent respiratory viral infection after the initial hospitalization. Adding serology to PCR may increase the diagnostic yield by 10.3-61.1% in children and 24.3-133.3% in adults, depending on the virus of interest. Other studies have also demonstrated that serology can increase diagnostic yield for these viruses by 3.0-86.7% when added to PCR [80, 81]. A few factors could explain this wide

range among studies. First, most studies do not provide details regarding the timing between NP/OP specimen collection relative to illness onset. If this is duration is longer, there may be false-negative PCR results, which could lead to higher estimates of diagnostic yield of serology. Second, diagnostic yield may differ by age; both studies combined both children of varying ages and adults for their analyses.

There are limitations to these analyses. First, more than 50% of enrolled children and adults were missing serology results. Imputing these data assumed they were missing at random; based on our analyses, we believed this assumption was reasonable. Second, estimates from multiple imputation could have changed if other independent variables were used for imputation. This concern may be minimal because results from another diagnostic test (PCR) were available and were a strong predictor of serology results. Third, the detection of viruses in NP/OP swabs is not necessarily indicative of the cause of CAP [18, 21] and may represent resolving infection rather than acute infection or an infection limited to the upper tract and not lower tract [20]. Fourth, multiple imputation may help to estimate burden at the population level from studies that collect convalescent samples, but development of more sensitive PCR methods for clinical purposes and to inform model-based estimates are still needed. Fifth, in the sensitivity analysis of estimating proportions of respiratory viral detections assuming imperfect specificities, we assumed conditional test independence for PCR and serology, though it is difficult to assess whether this is a reasonable assumption. Finally, illness onset was self or caregiver-reported, which introduces potential recall bias.

In conclusion, the proportion of CAP with adenovirus, HMPV, parainfluenza, and RSV detected using observed results may have underestimated the virus-specific burdens

because of missing specimens. Multiple imputation estimates may help inform virus-specific burden estimates in pneumonia etiology studies as well as other studies in which convalescent serum samples are available for a proportion, but not necessarily all, patients. If the proportion of available serology results had varied, it would be difficult to assess the impact of multiple imputation on burden estimates; the utility of multiple imputation is driven by the strength of the association between independent variables used for imputation and serology among patients with available data. Our sensitivity analyses indicate that there were few differences in virus-specific burden estimates when accounting for timing of NP/OP specimen. Our reported higher estimates of viral-specific pneumonia burdens based on multiple imputation underscore the need for enhancement of respiratory viral diagnostic tests at the patient level which would also inform population-based estimates and better inform clinical guidance and public health policy.

Supplementary Appendix

Methods for Sensitivity Analysis: LCA

LCA is a statistical technique that constructs a latent variable as inferred from multiple observed diagnostic tests using a model [133]. Two key model assumptions were made. First, we assumed the study population consisted of mutually exclusive latent classes, in which subjects in the same latent class were assumed to be homogeneous with respect to the likelihood of disease [134, 162]. A priori, we thought the latent classes should represent pneumonia due to a specific respiratory virus, pneumonia due to another respiratory virus for which we tested, or pneumonia not due to

a respiratory virus for which we tested; these labels directly correspond to the categorization of PCR and serology results. Because LCA was performed for adenovirus, HMPV, parainfluenza, and RSV separately, the labels reflected the respiratory virus of interest. Second, we assumed the observed diagnostic tests were independent of each other conditional on the latent variable [133, 135]. When these assumptions are met, LCA yields estimates of the probability of a diagnostic test result within a latent class and the probability of being in a latent class. These probabilities reflect the sensitivities of PCR and serology and the revised proportion of CAP associated with a specific respiratory virus, respectively.

For the LCA models, we selected diagnostic tests that would best characterize the latent variable of pneumonia etiology, which included PCR result, serology result, any positive bacterial test results, white blood cell count, and procalcitonin (adults only). We considered these variables for all viruses simultaneously. We categorized PCR and serology test results into 3 categories: negative for all viruses, positive for any virus except for the one of interest, and positive for the virus of interest regardless of whether another virus was detected.

After running models with combinations of these variables, we selected the final LCA model based on either the lowest Bayesian Information Criterion or the highest entropy (an indicator of latent class separation) and on model interpretation. LCA was performed separately for children and adults. The final LCA models for each virus included PCR result, serology result, and any positive bacterial results for children and PCR result, serology result, and procalcitonin for adults. LCA was also performed excluding outliers (defined as NP/OP specimen collected after 28 days from illness

onset). Further, we estimated PCR sensitivity by restricting to subjects that had NP/OP specimen collected within 7 days of illness onset. Observations that were missing data on any observed variables were retained in the models. To calculate 95% CI, bootstrapping methods of 4000 resampling draws were used. LCA was performed using Mplus 7.11.

Results for Sensitivity Analysis: LCA

Children

From LCA, the revised proportions of CAP with adenovirus, HMPV, parainfluenza, and RSV detected were 3.1%, 16.4%, 6.4%, and 26.1%, respectively; for HMPV and parainfluenza, these results were 2.5-3.5% higher but adenovirus was 8.1% lower and RSV was 2.2% lower than the observed EPIC study results. For adenovirus, HMPV, parainfluenza, and RSV, the sensitivities of PCR were 85.6%, 70.7%, 65.7%, and 90.9% and the sensitivities of serology were 60.9%, 66.2%, 71.5%, and 82.3%, respectively (Table S2.2). We also performed LCA among 2151 (96.8%) children who had NP/OP specimen collected within 28 days of illness onset. Compared with LCA estimates with all children, the sensitivities of PCR and serology were within 0.4-3.8% for adenovirus, parainfluenza and RSV and increased by 28.4% and by 13.2% for HMPV, respectively, though the revised proportion decreased to 9.4%.

To better refine the dataset for timing of specimen collections, we performed LCA among 1765 (79.4%) children who had NP/OP specimens collected within 7 days after illness onset; median time between illness onset and specimen collection was 3.8 days (IQR 2.6-5.5). The sensitivity of PCR decreased for adenovirus (73.4%) and

parainfluenza (50.2%), remained the same for HMPV (70.4%), and slightly increased for RSV (92.8%) when timing was restricted to within 7 days compared to estimates with no timing restrictions (Table S2.2). The observed proportions for each virus from these timing restrictions were within 0.2-0.8% of those with no restrictions (data not shown).

Adults

From LCA, the revised proportions of CAP with adenovirus, HMPV, parainfluenza, and RSV detected were 1.1%, 4.6%, 6.2%, and 2.1%, respectively; this was 0.3-0.9% higher for adenovirus and RSV respectively but 0.7-3.2% lower for HMPV and parainfluenza, respectively. For adenovirus, HMPV, parainfluenza, and RSV, the sensitivities of PCR were 100%, 76.3%, 39.8%, and 82.8% and the sensitivities of serology were 66.2%, 43%, 35.1%, and 100%, respectively (Table S2.5). We also performed LCA among 2142 (94.8%) adults who had NP/OP specimen collected within 28 days of illness onset. For adenovirus, HMPV, parainfluenza, and RSV, the sensitivities of PCR and serology were within 0-1.5% compared to estimates with all adults.

We also performed LCA among 1600 (71%) adults who had NP/OP specimens collected within 7 days after illness onset; median time between illness onset and specimen collection was 3.7 days (IQR 2.5-5.5). The sensitivity of PCR remained similar for adenovirus (100%) and parainfluenza (38.9%), increased for HMPV (100%), and decreased for RSV (82.8%) when timing was restricted to within 7 days compared to estimates with no timing restrictions (Table S2.5). From these timing restrictions, the observed proportions for each virus were within 0-0.5% of those with no restrictions (data not shown).

Discussion of Sensitivity Analysis: LCA

Multiple imputation estimates of the revised proportions for each of these respiratory viruses were higher than those estimated from LCA, except for HMPV in children. LCA estimates were more variably higher or lower compared with observed results and also varied from virus to virus.

As described in the main body of the manuscript, we had only two strong discriminators and a third, weak discriminator in our LCA models. One potential consequence was an underestimate of the revised proportion and possible over- or underestimates of PCR and/or serology sensitivity. For example, in children, not all of the adenovirus detections were included in the latent class of pneumonia due to adenovirus. Specifically, the probabilities of being PCR positive for adenovirus within the latent classes labeled as pneumonia due to another virus and pneumonia not due to a virus were 8.8% and 4.8%, respectively. Additionally, 79.8% of adenovirus detections also had another pathogen detected. These may explain why we estimated a lower revised proportion for adenovirus in children from LCA relative to the observed results (3.1% vs. 11.2%, respectively). These factors may also explain the lower revised proportions from LCA relative to the observed results for parainfluenza and RSV in children and adenovirus and RSV in adults. Similarly, when timing was restricted to within 7 days, if the revised proportion was underestimated, it may explain the decreased PCR sensitivity for adenovirus and parainfluenza in children and RSV in adults compared with estimates with no timing restrictions.

We also estimated PCR sensitivity for non-influenza respiratory viruses from LCA. Among hospitalized children, PCR sensitivity was 65.7% and 90.9% for parainfluenza and RSV, respectively. In comparison to studies of respiratory tract infections hospitalizations among children <5 years old, PCR sensitivity for parainfluenza was higher at 85.7-100% using culture [77] with immunofluorescence assay [168] as a referent standard and PCR sensitivity estimates for RSV were similar at 84.5-97.5% using culture [77, 169, 170] or an RSV antigenic test [169] as the referent standard. In a study of hospitalized adults >64 years old or who had underlying heart and lung disease, PCR sensitivity for RSV was 69% using culture and serology [68] as a composite reference standard. In comparison, our PCR sensitivity was higher at 82.8% for RSV among hospitalized adults; almost 70% had underlying heart or lung disease or were ≥65 years old. Differences in estimates could be due to differences in populations, especially with respect to age, and the choice of the referent standard.

We had concerns about the revised estimates from LCA. First, it is preferable to have ≥3 indicators of the latent variable [162], whereas we only had two strong discriminators (PCR and serology). We could not identify a strong, third indicator, so we included a discriminator that probably could only weakly distinguish between bacteria and viruses (e.g., any bacterial detection). Consequently, some with a positive test for a specific virus may not have been in the latent class labeled as pneumonia due to that virus. One potential consequence of this is an underestimate of the revised proportion and possible over- or underestimates of PCR and/or serology sensitivity. The proportion of children (25%) and adults (5%) who had positive tests for at least two pathogens, coupled with having only two strong discriminators, may have poorly identified the latent

classes and contributed to making interpretation unclear. Additionally, applying different model selection criteria led to different models; thus, results differed based on what variables were included in the latent class models and the criteria used for selecting them. However, to our knowledge, there is limited guidance in the LCA literature about variable selection and/or modeling strategy for such models [171].

Table 2.1. Demographic and clinical characteristics of children and adults hospitalized with community-acquired pneumonia.

	Children (n=2222) No. (%)	Adults (n=2259) No. (%)
Age groups		
<2 years	980 (44.0%)	---
2-4 years	559 (25.2%)	---
5-9 years	408 (18.4%)	---
10-17 years	275 (12.4%)	---
18-44 years	---	509 (22.5%)
45-64 years	---	945 (41.8%)
65+ years	---	805 (35.6%)
Age, median years (interquartile range, IQR)	2 (1-6)	57 (46-71)
Sex		
Male	1226 (55.2%)	1104 (48.9%)
Race/Ethnicity		
Non-Hispanic White	872 (39.2%)	1054 (46.7%)
Non-Hispanic Black	765 (34.4%)	874 (38.7%)
Hispanic	414 (18.6%)	238 (10.5%)
Other	171 (7.8%)	93 (4.1%)
Underlying Medical Conditions		
None reported	1469 (66.1%)	622 (27.5%)
Asthma	743 (33.4%)	584 (25.9%)
Chronic obstructive pulmonary disease	---	520 (23.0%)
Congenital heart disease	159 (7.2%)	0 (0%)
Coronary artery disease	0 (0%)	663 (29.4%)
Heart failure	---	430 (19.0%)
Diabetes mellitus	7 (0.3%)	584 (25.9%)
Chronic kidney disease	26 (1.2%)	356 (15.8%)
Chronic liver disease	6 (0.3%)	126 (5.6%)
Preterm birth in children under 2 years old	205 (9.2%)	0 (0%)
Immunosuppression	33 (1.5%)	368 (16.3%)
Cancer	9 (0.4%)	457 (20.2%)
Time from illness onset to PCR specimen collection, median days (IQR)	4.6 (2.8-7.4)	4.8 (2.8-8.8)
CDC Study City		
Chicago	---	1507 (66.7%)
Memphis	842 (37.9%)	---
Nashville	600 (27%)	752 (33.3%)
Salt Lake City	780 (35.1%)	---

*Underlying medical conditions included asthma, chronic obstructive pulmonary disease (adults only); congenital heart disease (children only), coronary artery disease; pre-term birth (defined as gestational age <37 weeks at birth in children under 2 years old); diabetes mellitus; chronic kidney disease; chronic liver disease; immunosuppression; any cancer (excluding skin cancers); neurological disorders (including seizure, cerebral palsy, scoliosis); and chromosomal disorders (including Down's syndrome).

Table 2.2. Multiple imputation estimations of a revised proportion of community-acquired pneumonia with detections of specific respiratory viruses among children hospitalized with community-acquired pneumonia (n=2222).

	Estimate (95% CI)			
	Adenovirus	HMPV	Parainfluenza	RSV
Observed proportions of any positive test result from the EPIC Study	11.2% (9.9%, 12.5%)	12.9% (11.5%, 14.3%)	6.8% (5.8%, 7.8%)	28.3% (26.4%, 30.2%)
Multiple imputation estimate Revised proportion of any positive test result	14.4% (10.7%, 18.0%)	15.0% (13.3%, 16.7%)	8.9% (7.5%, 10.4%)	29.9% (27.9%, 31.9%)
Restriction of timing for NP/OP specimen collection within 28 days of illness onset (n=2151)				
Observed proportions of any positive test result from the EPIC Study	11.3% (10.0%, 12.6%)	12.9% (11.5%, 14.3%)	6.9% (5.8%, 8.0%)	28.7% (26.8%, 30.6%)
Multiple imputation estimate Revised proportion of any positive test result	15.7% (9.1%, 22.2%)	14.9% (13.0%, 16.8%)	9.2% (7.6%, 10.9%)	30.3% (28.2%, 32.4%)

Table 2.3. Multiple imputation estimations of a revised proportion of community-acquired pneumonia with detections of specific respiratory viruses among adults hospitalized with community-acquired pneumonia (n=2259).

	Estimate (95% CI)			
	Adenovirus	HMPV	Parainfluenza	RSV
Observed proportions of any positive test result from the EPIC Study	1.4% (0.9%, 1.9%)	3.9% (3.1%, 4.7%)	3.0% (2.3%, 3.7%)	3.0% (2.3%, 3.7%)
Multiple imputation estimate Revised proportion of any positive test result	4.2% (-13.2%, 21.6%)	4.7% (3.7%, 5.7%)	4.0% (3.0%, 4.9%)	4.0% (2.9%, 5.0%)
Restriction of timing for NP/OP specimen collection within 28 days of illness onset (n=2142)				
Observed proportions of any positive test result from the EPIC Study	1.4% (0.9%, 1.9%)	4.0% (3.2%, 4.8%)	3.0% (2.3%, 3.7%)	3.1% (2.4%, 3.8%)
Multiple imputation estimate Revised proportion of any positive test result	2.3% (-1.1%, 5.7%)	4.8% (3.7%, 5.9%)	4.1% (2.9%, 5.4%)	4.1% (3.1%, 5.1%)

Supplementary Table S2.1: Characteristics of Hospitalized Children with Community-acquired Pneumonia: Comparison of Children with and without available serology data

	Children with available paired serology data (N=1023) No. (%)	Children without available paired serology data (N=1199) No. (%)
Age groups		
<2 years	396 (38.7%)	584 (48.7%)
2-4 years	254 (24.8%)	305 (25.4%)
5-9 years	224 (21.9%)	184 (15.4%)
10-17 years	149 (14.6%)	126 (10.5%)
Sex		
Male	580 (56.7%)	646 (53.9%)
Female	443 (43.3%)	553 (46.1%)
Race/Ethnicity		
Non-Hispanic White	454 (44.4%)	418 (34.9%)
Non-Hispanic Black	271 (26.5%)	494 (41.2%)
Hispanic	220 (21.5%)	194 (16.2%)
Other	78 (7.6%)	93 (7.8%)
Underlying Medical Conditions		
None reported	510 (49.9%)	580 (48.4%)
Asthma	336 (32.8%)	407 (33.9%)
Congenital heart disease	83 (8.1%)	76 (6.3%)
Preterm birth in children under 2 years old	78 (7.6%)	127 (10.6%)
Immunosuppression	18 (1.8%)	15 (1.3%)
Heart failure	11 (1.1%)	11 (0.9%)
Chronic kidney disease	8 (0.8%)	18 (1.5%)
Cancer	6 (0.6%)	3 (0.3%)
Diabetes mellitus	4 (0.4%)	3 (0.3%)
Chronic liver disease	1 (0.1%)	5 (0.4%)
Hospital Indicators		
Disease severity	228 (22.3%)	238 (19.9%)
Intensive care unit admission	226 (22.1%)	237 (19.8%)
Mechanical ventilation	70 (6.8%)	79 (6.6%)
Death	1 (0.1%)	2 (0.2%)

Supplementary Table S2.2: Latent class analysis estimations of a revised proportion of community-acquired pneumonia associated with specific respiratory viruses as well as sensitivities of PCR and serology among children hospitalized with community-acquired pneumonia (n=2222).

	Estimate (95% CI)			
	Adenovirus	HMPV	Parainfluenza	RSV
Observed proportions of any positive test result from the EPIC Study [158]	11.2% (9.9%, 12.5%)	12.9% (11.5%, 14.3%)	6.8% (5.8%, 7.8%)	28.3% (26.4%, 30.2%)
LCA estimate				
Revised proportion	3.1% (1.0%, 5.6%)	16.4% (14.1%, 19.8%)	6.4% (3.5%, 11.2%)	26.1% (22.8%, 30.2%)
Sensitivity of PCR	85.6% (47.2%, 100%)	70.7% (61.6%, 97.8%)	65.7% (39.4%, 100%)	90.9% (85.5%, 98.7%)
Sensitivity of Serology	60.9% (28.2%, 100%)	66.2% (53.3%, 74.6%)	71.5% (50.9%, 100%)	82.3% (72.9%, 91.2%)
LCA estimate when restricting timing of specimen collection within 28 days of illness onset (n=2151)				
Revised proportion	3.2% (1.3%, 5.8%)	9.4% (7.2%, 12.6%)	7.1% (3.9%, 12.4%)	26.3% (23.0%, 30.5%)
Sensitivity of PCR	84.7% (57.6%, 100%)	99.1% (75.2%, 100%)	61.9% (37.1%, 96.2%)	91.3% (85.3%, 99.2%)
Sensitivity of Serology	57.2% (27.0%, 100%)	79.4% (65.0%, 97.4%)	69.3% (49.2%, 100%)	82.9% (73.2%, 91.8%)
LCA estimate when restricting timing of specimen collection within 7 days of illness onset (n=1765)				
Revised proportion	3.2% (0.8%, 7.3%)	17.5% (14.5%, 21.5%)	6.6% (3.2%, 12.3%)	26.8% (22.8%, 31.4%)
Sensitivity of PCR	73.4% (34.6%, 100%)	70.4% (59.6%, 96.7%)	50.2% (30.4%, 86.7%)	92.8% (87.2%, 100%)

Supplementary Table S2.3: Characteristics of Hospitalized Adults with Community-acquired Pneumonia: Comparison of Adults with and without available serology data

	Adults with available paired serology data (N=854) No. (%)	Adults without available paired serology data (N=1405) No. (%)
Age groups		
18-44 years	187 (21.9%)	322 (22.9%)
45-64 years	384 (45.0%)	561 (39.9%)
65+ years	283 (33.1%)	522 (37.2%)
Sex		
Male	435 (50.9%)	669 (47.6%)
Female	419 (49.1%)	736 (52.4%)
Race/Ethnicity		
Non-Hispanic White	394 (46.1%)	660 (47.0%)
Non-Hispanic Black	330 (38.6%)	544 (38.7%)
Hispanic	97 (11.4%)	141 (10.0%)
Other	33 (3.9%)	60 (4.3%)
Underlying Medical Conditions		
None reported	171 (20.0%)	322 (22.9%)
Coronary artery disease	278 (32.6%)	385 (27.4%)
Asthma	230 (26.9%)	354 (25.2%)
Diabetes mellitus	225 (26.4%)	359 (25.6%)
Chronic obstructive pulmonary disease	189 (22.1%)	331 (23.6%)
Cancer	163 (19.1%)	242 (17.2%)
Heart failure	160 (18.7%)	270 (19.2%)
Immunosuppression	153 (17.9%)	215 (15.3%)
Chronic kidney disease	151 (17.7%)	205 (14.6%)
Chronic liver disease	53 (6.2%)	73 (5.2%)
Hospital Indicators		
Disease severity	183 (21.4%)	322 (22.9%)
Intensive care unit admission	175 (20.5%)	307 (21.9%)
Mechanical ventilation	33 (3.9%)	84 (6.0%)
Death	4 (0.5%)	45 (3.2%)

Supplementary Table S2.4: Characteristics of Hospitalized Adults with Community-acquired Pneumonia: Comparison of Adults with and without available procalcitonin data

	Adults with available procalcitonin data (N=1339) No. (%)	Adults without available procalcitonin data (N=920) No. (%)
Age groups		
18-44 years	305 (22.8%)	204 (22.2%)
45-64 years	579 (43.2%)	366 (39.8%)
65+ years	455 (34.0%)	350 (38.0%)
Sex		
Male	666 (49.7%)	438 (47.6%)
Female	673 (50.3%)	482 (52.4%)
Race/Ethnicity		
Non-Hispanic White	504 (37.6%)	550 (59.8%)
Non-Hispanic Black	568 (42.4%)	306 (33.3%)
Hispanic	199 (14.9%)	39 (4.2%)
Other	68 (5.1%)	25 (2.7%)
Underlying Medical Conditions		
None reported	322 (24.1%)	171 (18.6%)
Coronary artery disease	426 (31.8%)	237 (25.8%)
Asthma	336 (25.1%)	248 (27.0%)
Diabetes mellitus	325 (24.3%)	259 (28.2%)
Chronic obstructive pulmonary disease	243 (18.2%)	277 (30.1%)
Heart failure	241 (18.0%)	189 (20.5%)
Cancer	236 (17.6%)	169 (18.4%)
Chronic kidney disease	209 (15.6%)	147 (16.0%)
Immunosuppression	200 (14.9%)	168 (18.3%)
Chronic liver disease	58 (4.3%)	68 (7.4%)
Hospital Indicators		
Disease severity	291 (21.7%)	214 (23.3%)
Intensive care unit admission	288 (21.5%)	194 (21.1%)
Mechanical ventilation	65 (4.9%)	52 (5.7%)
Death	23 (1.7%)	26 (2.8%)

Supplementary Table S2.5: Latent class analysis estimations of a revised proportion of community-acquired pneumonia associated with specific respiratory viruses as well as sensitivities of PCR and serology among adults hospitalized with community-acquired pneumonia (n=2259).

	Estimate (95% CI)			
	Adenovirus	HMPV	Parainfluenza	RSV
Observed proportions of any positive test result from the EPIC Study [157]	1.4% (0.9%, 1.9%)	3.9% (3.1%, 4.7%)	3.0% (2.3%, 3.7%)	3.0% (2.3%, 3.7%)
LCA estimate				
Revised proportion	1.1% (0.6%, 1.6%)	4.6% (2.9%, 7.7%)	6.2% (3.6%, 16.6%)	2.1% (1.1%, 3.2%)
Sensitivity of PCR	100% (34.2%, 100%)	76.3% (46.5%, 100%)	39.8% (14.9%, 68.3%)	82.8% (56.3%, 100%)
Sensitivity of Serology	66.2% (6.1%, 100%)	43.0% (27.4%, 60.7%)	35.1% (16.6%, 59.8%)	100% (76.7%, 100%)
LCA estimate when restricting timing of specimen collection within 28 days of illness onset (n=2142)				
Revised proportion	1.1% (0.6%, 1.6%)	4.8% (2.9%, 7.9%)	6.6% (3.8%, 19.8%)	2.1% (1.1%, 3.4%)
Sensitivity of PCR	100% (50.7%, 100%)	75.6% (45.2%, 100%)	38.3% (12.4%, 64.1%)	83.0% (57.5%, 100%)
Sensitivity of Serology	66.7% (0%, 100%)	42.9% (25.0%, 60.8%)	34.3% (15.6%, 54.7%)	100% (59.0%, 100%)
LCA estimate when restricting timing of specimen collection within 7 days of illness onset (n=1600)				
Revised proportion	1.1% (0.5%, 2.0%)	3.8% (2.8%, 4.9%)	5.2% (2.1%, 15.5%)	3.1% (1.6%, 4.8%)
Sensitivity of PCR	100% (15.7%, 100%)	100% (100%, 100%)	38.9% (14.0%, 67.5%)	73.0% (45.7%, 92.9%)

CHAPTER 3. Use of multiple imputation and other methods to calculate a revised proportion of influenza virus detections among children and adults hospitalized with community-acquired pneumonia

Abstract

Background: Real-time polymerase chain reaction (PCR) and serology are detection methods for influenza virus infection; limitations of these tests can lead to misclassification of pneumonia attributable to influenza. Thus, revised estimates of influenza-associated community-acquired pneumonia (CAP) are needed to account for missing test results and imperfect test characteristics.

Methods: We analyzed data from an active population-based surveillance study for CAP requiring hospitalization among children (<18 years old) and adults. Influenza virus was considered detected if PCR or serology results were positive, assuming no false-positive results. Multiple imputation was applied to impute missing influenza serology results. The revised proportion of CAP due to influenza was estimated based on observed and imputed results from PCR and serology.

Results: Among 2222 children and 2259 adults with radiographically-confirmed CAP, 98.8% and 99.5% had NP/OP specimens and 43.2% and 37.5% had paired serum specimens available respectively. Using imputed serology results, the revised proportion of influenza-associated CAP increased in children from 6.7% to 11.1% and in adults from 5.8% to 7.9%.

Conclusions: The proportion of CAP with influenza virus detected using available test results may have underestimated the influenza burden. Imputing missing data could improve burden estimation.

Introduction

Community-acquired pneumonia (CAP) is a common cause of hospitalization among all ages in the United States [154, 155]. Both bacterial and viral respiratory pathogens, including influenza virus, can cause pneumonia and can be detected using a broad range of diagnostic tests. Timing of collection and quality of specimens affects diagnostic test performance [75]. Further, the specimens used for diagnostic testing are often not from the lower respiratory tract (e.g., lung tissue), making the interpretation of results more challenging. An additional challenge is the imperfect sensitivity and specificity of currently available diagnostic tests for respiratory pathogens and lack of a true gold standard [78, 83, 156]. PCR and serology are detection methods for respiratory viruses, though serology requires collection of paired serum specimens and thus has limited impact on patient management. These limitations of currently available diagnostic tools can lead to misclassification of pneumonia etiology attributable to influenza virus in pneumonia etiology studies.

Real-time polymerase chain reaction (PCR) has improved sensitivity for influenza virus detections relative to culture [66, 67, 75-77], and the addition of serology to PCR has also increased diagnostic yield for influenza virus detections in research studies [78, 80, 81, 83, 144]. The Centers for Disease Control and Prevention (CDC) Etiology of Pneumonia in the Community (EPIC) study used PCR and serology for detection of influenza virus in a multi-center active surveillance study of the incidence and etiology of CAP among hospitalized U.S. adults and children [157, 158]. Our focus is on influenza virus because it was detected by both PCR and serology. While most patients enrolled in the EPIC study had an available PCR result, serology results were available for less than

half of the patients. The goal of the study was to estimate the revised proportion of CAP due to influenza virus among children and adults hospitalized with CAP, accounting for missing serology test results using multiple imputation.

Materials and Methods

Enrollment, Specimen Collection, and Laboratory Methods

The CDC EPIC study was a prospective, multi-center, population-based, active surveillance study, as previously described in detail elsewhere [157, 158]. Briefly, from January 1, 2010 to June 30, 2012, children <18 years of age were enrolled in three pediatric hospitals in Memphis, TN, Nashville, TN, and Salt Lake City, UT, and adults were enrolled in three hospitals in Chicago, IL, and two hospitals in Nashville, TN. Patients admitted to a study hospital were eligible if they resided within the hospital catchment area and had evidence of acute respiratory infection and radiographically-confirmed pneumonia within 72 h of admission. Patients were included in the etiologic analysis if they met final radiographic criteria for CAP based on a study radiologist review and had samples available for both bacterial and viral detection. Clinical, demographic, and epidemiologic information were collected through interviews and medical chart review. Informed consent was obtained and the study protocol was approved by the institutional review boards at each institution and the CDC.

Different specimen types were collected for comprehensive diagnostic testing of respiratory pathogens from enrolled patients [157, 158]. Whole blood, acute serum, and naso-oropharyngeal (NP/OP) swabs were collected as soon as possible after enrollment; urine was only collected in adults. For adults with a productive cough, sputum was also

collected; only results from high-quality sputum samples were included. Pleural fluid, endotracheal aspirates, and bronchoalveolar-lavage specimens when collected in select cases for clinical care were included in the analysis. Convalescent serum was collected 3-10 weeks after enrollment.

Real-time PCR was performed on NP/OP swabs for detection of multiple viruses and atypical bacteria [157, 158]. Serologic testing for viruses was performed on available paired acute and convalescent serum specimens. Influenza serology utilized both hemagglutination inhibition (HI) and microneutralization (MN) assays to define positive results [157, 158]. Available adult serum specimens were analyzed to quantify the biomarker procalcitonin [62, 63]. Bacterial culture was performed on specimens from blood, sputum, pleural fluid, endotracheal aspirates, and bronchoalveolar-lavage. To test for specific bacteria, PCR assays were performed on pleural fluid and, in children only, on whole blood. Urine antigen testing was performed for *Legionella pneumophila* and *S. pneumoniae*.

Categorization of Variables Based on Diagnostic Test Results

A positive PCR test for adenovirus, coronavirus, human metapneumovirus, influenza A and B viruses, parainfluenza virus types 1-3, respiratory syncytial virus, or rhinovirus virus was defined as a cycle threshold value of <40 from a NP/OP PCR assay. A positive serologic test for adenovirus, human metapneumovirus, influenza A/B viruses, parainfluenza virus, or respiratory syncytial virus was defined as a ≥4-fold rise in agent-specific IgG antibody titer between paired acute and convalescent sera.

For both influenza A and B viruses, HI assay was performed but because the HI assay is not specific for influenza B viruses [143], serological samples that were positive

for influenza B virus infection by the HI assay were further tested using the MN assay to improve specificity. If the influenza serology results indicated seroconversion when vaccine was administered (based on self/caregiver report or vaccine verification) within 2 weeks before acute-phase serum collection, or between acute-phase and convalescent-phase serum collection, results were deemed inconclusive and were considered as missing serology results for this analysis.

In the EPIC study and for these analyses, influenza virus infection was considered present if there was a positive result from PCR or serology. A non-influenza virus was defined as being present if there was a positive result from PCR or serology, except for rhinovirus and coronaviruses which only depended on PCR results.

A bacterial pathogen was defined as being present if *Chlamydophila pneumoniae* or *Mycoplasma pneumoniae* was detected in the NP/OP swab by a PCR assay or if bacteria (e.g., *Staphylococcus aureus*, *Haemophilus influenzae*, *S. pneumoniae*, and *S. pyogenes*) were detected in blood, endotracheal aspirate, or bronchoalveolar-lavage specimen by culture or, in pleural fluid, by culture or a PCR. For children, a bacterial pathogen was also defined as being present if bacteria were detected in whole blood by PCR for *S. pneumoniae or S. pyogenes.* For adults, a bacterial pathogen was also defined as being present if *S. pneumoniae* or *L. pneumophila* was detected by the urine antigen test or if bacteria were detected in high-quality sputum by culture. A summary variable of any bacterial detections was created; no bacterial detections were observed if none of these criteria were met.

Procalcitonin was categorized as <0.1 μg/L, 0.1-<0.25 μg/L, 0.25-<0.5 μg/L, ≥0.5 μg/L [140], in which the two higher categories may correspond to bacterial infections and

the two lower categories may correspond to viral infections [141, 142]; this variable was only used in a secondary analysis presented in the Supplementary Appendix.

Influenza Vaccination Status and Other Variables for Analyses

For the EPIC study, influenza vaccination history was collected during the patient interview and medical charts were further reviewed to verify vaccination status. Vaccine receipt included the monovalent influenza A(H1N1)pdm09 vaccine (2009-2010 influenza season) or trivalent inactivated or live attenuated seasonal influenza vaccines (2010-2011 and 2011-2012 influenza seasons). Influenza season was defined as October 1 through April 30 for each study year.

Statistical Analysis

Multiple Imputation

We performed multiple imputation for missing serology results for influenza virus in children and adults separately. Multiple imputation uses variables from non-missing data, in this case the patient data from those who had serology results available, to impute data onto the missing data, in this case, the patient data who did not have serology results available, via a multivariable regression model [145, 146, 160]. To assess whether these data were missing at random [145, 146], we compared demographic and clinical characteristics among patients with and without available serology results.

Using both bivariate and multivariable analyses, we identified independent variables that could be used to impute missing serology results as either positive or negative. First, we constructed "*a priori*" models for children and adults using age, sex, any bacterial detection, and influenza PCR result chosen *a priori* based on biological plausibility. We hypothesized that influenza PCR results would be positively correlated

with influenza serology results, given the concordant results from both methods. Additionally, we hypothesized that any bacterial detection would be negatively correlated with influenza serology results, as patients with positive influenza serology results were less likely to have any bacteria detections compared to patients with negative influenza serology results.

Second, we assessed whether additional variables, including cough, fever, diarrhea, sore throat, abdominal pain, myalgia, chills, headache, influenza season, chest indrawing/retraction, underlying medical conditions, radiographic findings (e.g., consolidation, pleural effusion, and other infiltrates), ICU admission, and pneumonia severity (defined as having experienced at least one of the following: ICU admission, mechanical ventilation, acute respiratory distress syndrome, shock, or death), were significant in bivariate analyses. For the variables that were significant in the bivariate analyses, each variable was added one at a time to the *a priori* model to assess whether it was significant in the multivariable model. Model fit was evaluated using the Hosmer-Lemeshow goodness of fit test; a p-value <0.05 was used to indicate that the model did not fit the data well. Additionally, we compared the area under the curve between the *a priori* model with and without the potential variable. A non *a priori* variable was used to impute serology data if it was significant in both the bivariate and multivariable analyses and the area under the curve was higher for the *a priori* model with this variable relative to the model without it. Potential variables were assessed for children and adults separately.

Our final multiple imputation model for children included age, sex, any bacterial detection, influenza PCR result, and influenza season. For adults, our final model

included age, sex, any bacterial detection, influenza PCR result, and self-reported abdominal symptoms. Using these final variables, we applied multiple imputation to impute missing serology results. Twenty imputed datasets were created for improved efficiency [161]. We combined observed and imputed results from PCR and serology to create a binary outcome, which was defined as either positive or negative for influenza virus infection. We then estimated a revised proportion of any positive result for influenza virus infection and its corresponding 95% confidence interval (CI), referred to as the revised proportion of CAP with influenza virus detected. We also performed a sensitivity analysis in which the multiple imputation model excluded patients for whom NP/OP specimens were collected after 28 days from illness onset, as the likelihood of a positive PCR result can decrease with a longer duration between illness onset and NP/OP specimen collection [75, 77].

We compared the revised proportions to the observed EPIC study results in terms of absolute difference by subtracting the proportion observed in the EPIC study from the proportion estimated from multiple imputation. Relative difference was calculated by dividing the proportion estimated from multiple imputation by the proportion observed in the EPIC study. Finally, among patients who had both PCR and serology results available, diagnostic yield of serology was calculated. We applied the formula: [(serology positive/PCR negative specimens) ÷ (specimens positive by both methods + serology negative/PCR positive specimens)]; diagnostic yield of serology was calculated for children and adults separately.

Secondary Analysis: Latent Class Analysis

Latent class analysis (LCA) was also considered for estimating revised proportions of CAP due to influenza. However, LCA proved to be unstable and potentially misleading. Specifically, 25% of children and 5% of adults likely had more than one pathogen detected, modeled latent classes were difficult to interpret, and the required assumptions for LCA did not appear to be supported including absence of three separate moderately predictive diagnostic tests. Furthermore, results were sensitive to model selection strategy. LCA methods, including the assumptions, details of the model results, and a discussion of results are described in the supplementary appendix.

Bivariate and multivariable regression analyses and multiple imputation were performed using SAS 9.3. LCA was performed in Mplus 7.

Results

Children

Of 2638 enrolled children, 2358 (89.4%) met the final radiographic criteria for CAP, among whom 2222 (94.2%) had samples available for both bacterial and viral diagnostic tests. Median age was 2 years (interquartile range [IQR] 1-6); 50.9% of children had an underlying medical condition, including 33.4% that reported asthma (Table 3.1). Among the 2222 children, 2196 (98.8%) had NP/OP specimens and 961 (43.2%) had paired serum specimens available. We found no meaningful differences in the demographic and clinical characteristics among children with and without available serology data (Table S3.1 in the Supplementary Appendix). For 2190 (98.6%) children who had an available NP/OP swab and illness onset data, the median time between illness onset and specimen collection was 4.6 days (IQR 2.8-7.4).

For all 2222 children, the revised proportion of CAP with influenza virus detected from multiple imputation was 11.1%, which is 4.4% higher in an absolute difference and is 1.7 times higher in relative terms than the observed EPIC study results (Table 3.2). When limiting multiple imputation to the 2151 (96.8%) children who had NP/OP specimens collected within 28 days of illness onset, the revised proportion was 0.2% lower than the corresponding estimate among all children. The contribution of serology-positive and PCR-negative detections above any PCR-positive detections resulted in a diagnostic yield of serology of 184.4% among the 870 (39.2%) children who had both PCR and serology results available. Results from the LCA are described in the supplement (Table S3.2 in the Supplementary Appendix).

Adults

Of 2488 enrolled adults, 2320 (93.2%) met the final radiographic criteria for CAP, among whom 2259 (97.3%) had specimens available for bacterial and viral diagnostic testing. Median age was 57 years (IQR 46-71); 78.2% of adults had an underlying medical condition (Table 3.1). Among these 2259 adults, 2248 (99.5%) had NP/OP specimens and 846 (37.5%) had paired serum specimens available. Data on procalcitonin were available for 1339 (59.3%) adults. There were no meaningful differences in demographic and clinical characteristics between adults with and without available serology data (Table S3.3 in the Supplementary Appendix) and between patients with and without procalcitonin data available (Table S3.4 in the Supplementary Appendix). For 2246 (99.4%) adults that had a NP/OP swab and illness onset data available, the median time between illness onset and specimen collection was 4.8 days (IQR 2.8-8.8).

For all 2259 adults, the revised proportion of CAP with influenza detected was 7.9% from multiple imputation; this is 2.1% higher in an absolute difference and 1.4 times higher in relative terms than the observed results (Table 3.2). When limiting the multiple imputation model to the 2142 (94.8%) adults who had NP/OP specimens collected within 28 days of illness onset, the revised proportion was 0.2% higher than that among all adults. The contribution of serology-positive and PCR-negative detections above any PCR-positive detections resulted in a diagnostic yield of serology of 78.6%, among the 741 (32.8%) adults who had both PCR and serology results available. Results from the LCA are described in the supplement (Table S3.2 in the Supplementary Appendix).

Discussion

We estimated revised proportions of CAP with influenza virus detected by applying multiple imputation to account for missing serology results, using prospectively collected clinical and microbiological data from patients hospitalized with CAP enrolled in the EPIC study. Multiple imputation increased the estimates of influenza-associated CAP from 6.7% to 11.1% in children and from 5.8% to 7.9% in adults. Our results illustrate the potential underestimation of influenza-associated pneumonia burden when serology results are either not included or missing.

More than half of enrolled children and adults were missing serology results, largely due to the lack of a convalescent sample. Patients are lost to follow-up, and the potential methods required to improve follow-up and thus increase specimen collection require intense resources that are often not available for convalescent specimen

collection. Thus, using modeling methods, such as multiple imputation, could be informative in estimating influenza burden. Multiple imputation requires an assumption that the data are missing at random. As we found no meaningful differences between characteristics between patients with and without available serology results, we believed that this assumption was valid. We found minimal differences in revised estimates when multiple imputation was restricted to include NP/OP specimens collected within 28 days of illness onset; these estimates are likely similar because only 3.2% of children and 5.2% of adults were excluded in the sensitivity analysis.

The revised estimates obtained using multiple imputation were higher than the observed estimates from the EPIC study, as missing serology results can misclassify pneumonia due to influenza virus. Among 870 (39.2%) children and 741 (32.8%) adults who had both PCR and serology results available, 67.9-71.9% of PCR-positive results were also serology-positive. PCR may detect influenza virus not captured by serology because PCR detects the presence of influenza virus in the naso/oropharynx whereas serology measures the influenza-specific antibody response. Several studies have shown that influenza detection by PCR was higher when NP/OP specimen are collected within 7 days of illness onset, compared with 8-14 days after illness onset [77, 123, 172]; in our study, 1765 (79%) of children and 1600 (71%) of adults had NP/OP specimens collected within 7 days of illness onset. In contrast, 46.3% and 28.1% of serology-positive results were also PCR-positive in children and adults, respectively. Thus, it is likely that serology detects influenza virus that is not captured by PCR. One possible reason for this is because the convalescent serum sample is obtained 3-7 weeks after the acute sample (thus the timing is not precisely the same as when samples for PCR were collected),

which may capture detections missed by PCR due to late collection of NP/OP specimens relative to illness onset. Adding serology to PCR may increase the diagnostic yield by 78.6% in adults and 184.4% in children. Other studies have also demonstrated that serology can increase diagnostic yield by 15-130% when added to PCR [78, 80, 81, 83]. There may be several factors that could explain this wide range among studies. First, most studies do not provide details regarding the duration between NP/OP specimen collection relative to illness onset. If this duration is longer, there may be false-negative PCR results, which could lead to higher estimates of diagnostic yield of serology. Second, diagnostic yield may differ by age; most studies combined both children of varying ages and adults for diagnostic yield calculations, though one only included adults [78]. Finally, the diagnostic yield of 110-130% for influenza B viruses may be over-estimated because these studies utilized the HI assay [80, 81], which is not specific for influenza B viruses [143].

There are limitations to these analyses. First, more than 50% of enrolled children and adults were missing serology results. Imputing these data assumed they were missing at random. Based on our analyses, we believed this assumption was reasonable. Second, estimates from multiple imputation could have changed if other independent variables were used for imputation. This concern may be minimal because PCR results were available and were a strong predictor of serology results. Third, the detection of viruses in NP/OP swabs is not necessarily indicative of the cause of CAP [18, 21] and may represent resolving infection rather than acute infection or an infection limited to the upper tract and not lower tract [20]. Fourth, multiple imputation may help to estimate burden at the population level from studies that collect convalescent samples, but

development of more sensitive PCR methods for clinical purposes and to inform model-based estimates are still needed. Fifth, in the sensitivity analysis of estimating proportions of influenza virus detection assuming imperfect specificities, we assumed conditional test independence for PCR and serology, though it is difficult to assess whether this is a reasonable assumption.

In conclusion, the proportion of CAP with influenza virus detected using observed results may have underestimated the influenza burden because of missing specimens and imperfect test characteristics. Multiple imputation estimates may help inform influenza burden estimates in pneumonia etiology studies as well as other studies in which convalescent serum samples are available for a proportion but not necessarily all patients. If the proportion of available serology results had varied, it would be difficult to assess the impact of multiple imputation on burden estimates; the utility of multiple imputation is driven by the strength of the association between independent variables used for imputation and serology among patients with available data. Our sensitivity analyses indicated that there were few differences in influenza burden estimates when accounting for timing of NP/OP specimen collection. Our reported higher estimates of influenza-associated pneumonia burden based on multiple imputation underscore the need for enhancement of influenza diagnostic tests at the patient level which would also inform population-based estimates and better inform clinical guidance and public health policy.

Supplementary Appendix

Methods for Secondary Analysis: Latent Class Analysis

LCA is a statistical technique that constructs a latent variable as inferred from multiple observed diagnostic tests using a model [133]. Two key model assumptions were made. First, we assumed the study population consisted of mutually exclusive latent classes, in which subjects in the same latent class were assumed to be homogeneous with respect to the likelihood of disease [134, 162]. A priori, we determined that the latent classes should represent pneumonia due to influenza A/B viruses, pneumonia due to any other respiratory virus for which we tested, or pneumonia not due to a respiratory virus for which we tested; these labels directly correspond to the categorization of PCR and serology results. Second, we assumed the observed diagnostic tests were independent of each other conditional on the latent variable [133, 135]. When these assumptions are met, LCA yields estimates of the probability of a diagnostic test result within a latent class and the probability of being in a latent class. These probabilities reflect the sensitivities of PCR and serology and the revised proportion of CAP associated with a specific respiratory virus, respectively.

For the LCA models, we selected diagnostic tests that would best characterize the latent variable of pneumonia etiology, which included PCR result, serology result, any positive bacterial test results, white blood cell count, and procalcitonin (adults only). We considered the tests for all viruses simultaneously. We categorized PCR and serology test results into three categories: negative for all viruses, positive for any virus except for influenza A/B, and positive for influenza A/B regardless of whether another virus was detected.

After running models with combinations of these variables, we selected the final LCA model based on either the lowest Bayesian Information Criterion or the highest

entropy (an indicator of latent class separation) and on model interpretation. LCA was performed separately for children and adults. The final LCA models for each virus included PCR result, serology result, and any bacterial detections for children and PCR result, serology result, and procalcitonin for adults. LCA was also performed excluding subjects with NP/OP specimen collected after 28 days of illness onset. Further, we estimated PCR sensitivity by restricting to subjects that had NP/OP specimen collected within 7 days of illness onset. Observations that were missing data on any observed variables were retained in the models. To calculate 95% CI, bootstrapping methods of 4000 resampling draws were used. LCA was performed using Mplus 7.11.

In a sensitivity analysis, we also explored the impact of assuming fixed values of serology sensitivity and specificity on PCR sensitivity in children. We set serology sensitivity and specificity each to six values, ranging from 70%-100%. For each of these 36 combinations, we noted the corresponding PCR sensitivity, as estimated from the LCA model.

Results for Secondary Analysis: Latent Class Analysis

Children

From LCA, this revised proportion was 9.5%; this result was 2.8% higher than the observed result (Table S3.2). The sensitivities of PCR and serology were 34.3% and 87.9%, respectively. We also performed LCA among 2151 (96.8%) children who had NP/OP specimen collected within 28 days of illness onset. Compared with LCA estimates among all children, the sensitivities of PCR and serology increased by 52.5% and decreased by 7.2%, respectively, though the revised proportion decreased to 4.1%.

To better refine the dataset for timing of specimen collections, we performed LCA among 1765 (79.4%) children who had NP/OP specimens collected within 7 days after illness onset; median time between illness onset and specimen collection was 3.8 days (IQR 2.6-5.5). The sensitivity of PCR increased (88.2%) when timing was restricted to within 7 days compared to estimates with no timing restrictions (Table S3.2). The observed proportions of influenza A/B detections from each of these timing restrictions were within 0.4% of those with no restrictions (data not shown).

Additionally, from fixing serology test characteristics in 36 combinations, we observed a range of PCR sensitivity estimates from 30-38%.

Adults

Among 1339 (59.3%) adults who had procalcitonin data available, 593 (26.3%) had <0.1 μg/L, 216 (9.6%) had 0.1-<0.25 μg/L, 128 (5.7%) had 0.25-<0.5 μg/L, and 402 (17.8%) had ≥0.5 μg/L.

From LCA, the revised proportion was 5.9%; this was 0.1% higher (Table S3.2). The sensitivities of PCR and serology were 83.1% and 67.8%, respectively. We also performed LCA among 2142 (94.8%) adults who had NP/OP specimen collected within 28 days of illness onset. From LCA, the sensitivities of PCR and serology were within 2.4-2.8% of the estimates among all adults.

We also performed LCA among 1600 (71%) adults who had NP/OP specimens collected within 7 days after illness onset; median time between illness onset and specimen collection was 3.7 days (IQR 2.5-5.5). The sensitivity of PCR increased (100%) when timing was restricted to within 7 days compared to estimates with no timing restrictions (Table S3.2). The observed proportions of influenza A/B detections

from each of these timing restrictions were within 0.8% of those with no restrictions (data not shown).

Discussion of Secondary Analysis: LCA

Multiple imputation estimates of the revised proportions for influenza A/B viruses were higher than those estimated from LCA. Both multiple imputation and LCA estimates were 0.1-4.4% higher compared to the observed results.

For example, among all children, the probability of having a positive serology result for influenza within the latent class labeled as pneumonia not due to a virus was 4.8% and the probabilities of having a positive PCR result for influenza was 0.8-0.9% in the latent classes of pneumonia due to another virus and pneumonia not due to a virus. If some positive detections of influenza are not in the latent class of pneumonia due to influenza A/B viruses, one potential consequence is an underestimate of the revised proportion and possible over- or underestimates of PCR and/or serology sensitivity. For example, among children who had NP/OP specimen collected within 28 days of illness onset, the probability of having a positive serology results for influenza within the latent classes of pneumonia not due to influenza and of pneumonia due to other viruses were 5.0% and 7.1%, respectively. This may explain why we estimated a lower revised proportion for influenza in children from LCA relative to the observed results (4.1% vs. 6.6%, respectively), despite that the sensitivity estimates of PCR and serology from this model (86.8% and 80.7%, respectively) are more consistent with our expectations than those estimates with no timing restrictions (34.3% and 87.9%, respectively).

Timing between illness onset and NP/OP specimen collection is important to PCR test performance. Studies have described that prolonged timing can lead to a lower

likelihood of a positive result [75, 81] and can impact sensitivity [78]. This could explain why PCR sensitivity increased from 34.3% among all children to 86.8% among those who had NP/OP specimen collected within 28 days of illness onset as well as the slight increase from 83.1% to 85.5% in PCR sensitivity in adults, respectively. More specifically, studies have described that the time from illness onset to the last day of a positive PCR test was 5-7 days in a household study of naturally acquired influenza infection [79], challenge studies [172, 173], and studies of outpatients and patients presenting to emergency departments of all ages [174, 175]. Some studies of adults hospitalized with influenza have reported prolonged viral replication among those with severe disease, including those with comorbidities [123, 176]. Among children and adults who had NP/OP specimen collected within 7 days of illness onset, we estimated PCR sensitivity to be 88.2% and 100%, respectively. These estimates may indicate that collecting NP/OP specimen within 7 days of illness onset remains meaningful for those hospitalized with influenza-associated pneumonia.

We had concerns about the revised estimates from LCA. First, it is preferable to have ≥3 indicators of the latent variable [162]. We could only identify two strong discriminators (PCR and serology), so we included a discriminator that could only weakly distinguish between bacterial and viral detections. LCA estimates for adults and children were higher than those from the EPIC study, though not all positive test results for influenza A/B viruses were in the corresponding latent class. Consequently, the revised proportion could be underestimated, resulting in a possible over- or underestimates of PCR and/or serology sensitivity. The proportion of co-detections in children (25%) and adults (5%), coupled with having only two strong discriminators, may

have poorly identified the latent classes and obscured the interpretation of the results. Additionally, applying different model selection criteria led to different models; thus, results differed based on what variables were included in the latent class models. However, to our knowledge, there is limited guidance to the LCA literature about variable selection and/or modeling strategy [171].

Table 3.1. Demographic and clinical characteristics of children and adults hospitalized with community-acquired pneumonia

	Children (n=2222) No. (%)	Adults (n=2259) No. (%)
Age groups		
<2 years	980 (44.0%)	---
2-4 years	559 (25.2%)	---
5-9 years	408 (18.4%)	---
10-17 years	275 (12.4%)	---
18-44 years	---	509 (22.5%)
45-64 years	---	945 (41.8%)
65+ years	---	805 (35.6%)
Age, median years (interquartile range, IQR)	2 (1-6)	57 (46-71)
Sex		
Male	1226 (55.2%)	1104 (48.9%)
Race/Ethnicity		
Non-Hispanic White	872 (39.2%)	1054 (46.7%)
Non-Hispanic Black	765 (34.4%)	874 (38.7%)
Hispanic	414 (18.6%)	238 (10.5%)
Other	171 (7.8%)	93 (4.1%)
Underlying Medical Conditions*		
None reported	1090 (49.1%)	493 (21.8%)
Asthma	743 (33.4%)	584 (25.9%)
Preterm birth in children under 2 years old	205 (9.2%)	---
Congenital heart disease	159 (7.2%)	---
Immunosuppression	33 (1.5%)	368 (15.8%)
Chronic kidney disease	26 (1.2%)	356 (15.8%)
Heart failure	22 (1%)	430 (19.0%)
Cancer	9 (0.4%)	405 (17.9%)
Diabetes mellitus	7 (0.3%)	584 (25.9%)
Chronic liver disease	6 (0.3%)	126 (5.6%)
Coronary artery disease	---	663 (29.4%)
Chronic obstructive pulmonary disease	---	520 (23.0%)
Time from illness onset to NP/OP specimen collection, median days (IQR)	4.6 (2.8-7.4)	4.8 (2.8-8.8)
Study City		
Chicago	---	1507 (66.7%)
Memphis	842 (37.9%)	---
Nashville	600 (27%)	752 (33.3%)
Salt Lake City	780 (35.1%)	---

*Underlying medical conditions included asthma, chronic obstructive pulmonary disease (adults only); congenital heart disease (children only), coronary artery disease; pre-term birth (defined as gestational age <37 weeks at birth in children under 2 years old); diabetes mellitus; chronic kidney disease; chronic liver disease; immunosuppression; any cancer (excluding skin cancers); neurological disorders (including seizure, cerebral palsy, scoliosis); and chromosomal disorders (including Down's syndrome).

Table 3.2. Multiple imputation estimations of a revised proportion of community-acquired pneumonia with influenza virus detection among children (n=2222) and adults (n=2259) hospitalized with community-acquired pneumonia.

	Estimate (95% CI)			
	No restriction of timing of NP/OP specimen collection (n=2222 children, 2259 adults)		Restriction of timing for NP/OP specimen collection within 28 days of illness onset (n=2151 children, 2142 adults)	
	Observed proportions of any positive influenza test result from the EPIC Study [157, 158]	Multiple imputation estimate: Revised proportion of any positive influenza test result	Observed proportions of any positive influenza test result from the EPIC Study	Multiple imputation estimate: Revised proportion of any positive influenza test result
Children	6.7% (5.7%, 7.7%)	11.1% (9.5%, 12.7%)	6.6% (5.6%, 7.6%)	10.9% (9.1%, 12.6%)
Adults	5.8% (4.8%, 6.8%)	7.9% (6.3%, 9.6%)	6.0% (5.0%, 7.0%)	8.1% (6.6%, 9.5%)

Supplementary Table S3.1. Characteristics of Hospitalized Children with Community-acquired Pneumonia: Comparison of Children with and without available influenza serology data

	Children with available paired serology data (N=961) No. (%)	Children without available paired serology data (N=1261) No. (%)
Age groups		
<2 years	359 (37.4%)	621 (49.2%)
2-4 years	241 (25.1%)	318 (25.2%)
5-9 years	217 (22.6%)	191 (15.2%)
10-17 years	144 (15.0%)	131 (10.4%)
Sex		
Male	546 (56.8%)	680 (53.9%)
Female	415 (43.2%)	581 (46.1%)
Race/Ethnicity		
Non-Hispanic White	439 (45.7%)	433 (34.3%)
Non-Hispanic Black	235 (24.5%)	530 (42.0%)
Hispanic	213 (22.2%)	201 (15.9%)
Other	74 (7.7%)	97 (7.7%)
Underlying Medical Conditions		
None reported	484 (50.4%)	606 (48.1%)
Asthma	314 (32.7%)	429 (34.0%)
Congenital heart disease	80 (8.3%)	79 (6.3%)
Preterm birth in children under 2 years old	64 (6.7%)	141 (11.2%)
Immunosuppression	18 (1.9%)	15 (1.2%)
Heart failure	10 (1.0%)	12 (1.0%)
Chronic kidney disease	8 (0.8%)	18 (1.4%)
Cancer	6 (0.6%)	3 (0.2%)
Diabetes mellitus	4 (0.4%)	3 (0.2%)
Chronic liver disease	1 (0.1%)	5 (0.4%)
Hospital Indicators		
Disease severity	216 (22.5%)	250 (19.8%)
Intensive care unit admission	214 (22.3%)	249 (19.8%)
Mechanical ventilation	64 (6.7%)	85 (6.7%)
Death	0 (0%)	3 (0.2%)

Supplementary Table S3.2. Latent class analysis estimations of a revised proportion of community-acquired pneumonia associated with influenza A/B viruses as well as sensitivities of PCR and serology among children (n=2222) and adults (n=2259) hospitalized with community-acquired pneumonia.

	Estimate (95% CI)	
	Children (n=2222)	Adults (n=2259)
Observed proportions of any positive detection from the EPIC Study	6.7% (5.7%, 7.7%)	5.8% (4.8%, 6.8%)
LCA estimate		
Revised proportion	9.5% (4.2%, 14.4%)	5.9% (4.6%, 8.8%)
Sensitivity of PCR	34.3% (22.2%, 82.1%)	83.1% (51.3%, 100%)
Sensitivity of Serology	87.9% (65.2%, 100%)	67.8% (48.2%, 84.4%)
LCA estimate when restricting timing of NP/OP specimen collection within 28 days of illness onset	(n=2151)	(n=2142)
Revised proportion	4.1% (2.3%, 8.4%)	5.8% (4.7%, 9.0%)
Sensitivity of PCR	86.8% (38.9%, 100%)	85.5% (52.4%, 100%)
Sensitivity of Serology	80.7% (58.1%, 97.1%)	70.6% (51.5%, 86.8%)
LCA estimate when restricting timing of NP/OP specimen collection within 7 days of illness onset	(n=1765)	(n=1600)
Revised proportion	4.4% (3.1%, 6.3%)	4.3% (2.1%, 6%)
Sensitivity of PCR	88.2% (62.3%, 100%)	100.0% (53.2%, 100%)

Supplementary Table S3.3. Characteristics of Hospitalized Adults with Community-acquired Pneumonia: Comparison of Adults with and without available influenza serology data

	Adults with available paired serology data (N=846) No. (%)	Adults without available paired serology data (N=1413) No. (%)
Age groups		
18-44 years	184 (21.7%)	325 (23.0%)
45-64 years	384 (45.4%)	561 (39.7%)
65+ years	278 (32.9%)	527 (37.3%)
Sex		
Male	428 (50.6%)	676 (47.8%)
Female	418 (49.4%)	737 (52.2%)
Race/Ethnicity		
Non-Hispanic White	390 (46.1%)	664 (47.0%)
Non-Hispanic Black	327 (38.6%)	547 (38.7%)
Hispanic	96 (11.4%)	142 (10.1%)
Other	33 (3.9%)	60 (4.2%)
Underlying Medical Conditions		
None reported	170 (20.1%)	323 (22.9%)
Coronary artery disease	274 (32.4%)	389 (27.5%)
Asthma	227 (26.8%)	357 (25.3%)
Diabetes mellitus	222 (26.2%)	362 (25.6%)
Chronic obstructive pulmonary disease	187 (22.1%)	333 (23.6%)
Cancer	162 (19.2%)	243 (17.2%)
Heart failure	160 (18.9%)	270 (19.1%)
Immunosuppression	150 (17.7%)	218 (15.4%)
Chronic kidney disease	148 (17.5%)	208 (14.7%)
Chronic liver disease	53 (6.3%)	73 (5.2%)
Hospital Indicators		
Disease severity	181 (21.4%)	324 (22.9%)
Intensive care unit admission	173 (20.5%)	309 (21.9%)
Mechanical ventilation	33 (3.9%)	84 (5.9%)
Death	4 (0.5%)	45 (3.2%)

Supplementary Table S3.4. Characteristics of Hospitalized Adults with Community-acquired Pneumonia: Comparison of Adults with and without available procalcitonin data

	Adults with available procalcitonin data (N=1339) No. (%)	Adults without available procalcitonin data (N=920) No. (%)
Age groups		
18-44 years	305 (22.8%)	204 (22.2%)
45-64 years	579 (43.2%)	366 (39.8%)
65+ years	455 (34.0%)	350 (38.0%)
Sex		
Male	666 (49.7%)	438 (47.6%)
Female	673 (50.3%)	482 (52.4%)
Race/Ethnicity		
Non-Hispanic White	504 (37.6%)	550 (59.8%)
Non-Hispanic Black	568 (42.4%)	306 (33.3%)
Hispanic	199 (14.9%)	39 (4.2%)
Other	68 (5.1%)	25 (2.7%)
Underlying Medical Conditions		
None reported	322 (24.1%)	171 (18.6%)
Coronary artery disease	426 (31.8%)	237 (25.8%)
Asthma	336 (25.1%)	248 (27.0%)
Diabetes mellitus	325 (24.3%)	259 (28.2%)
Chronic obstructive pulmonary disease	243 (18.2%)	277 (30.1%)
Heart failure	241 (18.0%)	189 (20.5%)
Cancer	236 (17.6%)	169 (18.4%)
Chronic kidney disease	209 (15.6%)	147 (16.0%)
Immunosuppression	200 (14.9%)	168 (18.3%)
Chronic liver disease	58 (4.3%)	68 (7.4%)
Hospital Indicators		
Disease severity	291 (21.7%)	214 (23.3%)
Intensive care unit admission	288 (21.5%)	194 (21.1%)
Mechanical ventilation	65 (4.9%)	52 (5.7%)
Death	23 (1.7%)	26 (2.8%)

CHAPTER 4. Prediction Models and Scores for Influenza-Associated Pneumonia among Children and Adults Hospitalized with Community-acquired Pneumonia

Abstract

Background: Early identification of influenza-associated pneumonia is crucial because of increased risk of severe outcomes and need for timely initiation of antiviral therapy. However, it is difficult to clinically discern influenza-associated pneumonia from other causes of pneumonia. A prediction score using readily available clinical factors on admission could help identify influenza-associated pneumonia earlier.

Methods: We analyzed data from an active population-based surveillance study for community-acquired pneumonia requiring hospitalization among children (<18 years old) and adults. Influenza testing included both polymerase chain reaction (PCR) and serology; although 56.8-62.5% of serology results were missing due to lack of convalescent serum. Influenza-associated pneumonia was defined two ways: one definition used observed PCR and serology results, and the other used multiply imputed data that accounted for missing serology results. Multivariable models were developed using factors to inform prediction scores. Point values were assigned based on the predictors' coefficients in the models; an individual's score was the sum of point values based on observed characteristics.

Results: Among 2222 children and 2259 adults, 5.8-6.7% patients had influenza-associated pneumonia based on the observed data; 7.3-10.5% patients had the outcome based on multiply imputed data. Significant predictors included age (<2 years old: adjusted odds ratios (aORs) from both definitions=0.51-0.58; 2-4 years old: aORs=0.46;

5-9 years old: aORs=0.84-1.07) and influenza season (aORs=3.06) for children and, for adults, any underlying medical conditions (aORs=0.52-0.53), leukocytosis (aORs=0.42-0.55), cough (aORs=3.74-17.56), abdominal pain (aORs=1.56-2.21), and influenza season (aORs=3.29-4.44). The discrimination was poor for children and good for adults (c-statistics ranged from 0.64-0.65 and 0.72-0.77, respectively); all models had good calibration (p-values for goodness of fit ranged from 0.15-0.99). The prediction scores had high negative predictive values (90-99%) and, overall, low positive predictive values (0-29%).

Conclusions: Despite identifying significant factors, our scores reflect difficulty in predicting influenza-associated pneumonia and differentiating it from other causes of pneumonia.

Introduction

Pneumonia is a known complication of seasonal [65, 105-108] and pandemic influenza [109-113] virus infection. Early diagnosis and identification of influenza-associated pneumonia is crucial because of increased risk of severe outcomes [107, 113, 119] and need for timely initiation of antiviral therapy [107, 109, 120]. However, clinical features of influenza-associated pneumonia are difficult to distinguish from other causes of community-acquired pneumonia (CAP) [65, 108, 122]. Additionally, not all patients with suspected influenza, including those with pneumonia, get tested for influenza virus infection [65]. Because testing practice is clinically driven and behaviors may shift, it is possible that patients with influenza-associated pneumonia may be missed if radiographs and/or appropriate diagnostic testing are not performed. It is possible that meaningful combinations of symptoms and clinical signs could be useful for early identification of influenza-associated pneumonia and could be determined through the development of a prediction model and a corresponding score.

The Centers for Disease Control and Prevention (CDC) Etiology of Pneumonia in the Community (EPIC) study systematically tested all enrolled patients for influenza and other respiratory pathogens and collected data on >20 symptoms and clinical signs through a multi-center active surveillance study of the incidence and etiology of CAP among hospitalized U.S. adults and children [157, 158]. Used the EPIC study data, the goal of the study was to develop and evaluate prediction models and corresponding scores for early identification of influenza-associated pneumonia among children and adults hospitalized with CAP.

Materials and Methods

Enrollment, Laboratory Methods, and Definitions

The CDC EPIC study was a prospective, multi-center, population-based, active surveillance study, that has previously been described in detail elsewhere [157, 158]. Briefly, from January 1, 2010 to June 30, 2012, children <18 years of age were enrolled in three pediatric hospitals in Memphis, TN, Nashville, TN, and Salt Lake City, UT, and adults were enrolled in three hospitals in Chicago, IL, and two hospitals in Nashville, TN. Patients were eligible if they were admitted to a study hospital, resided within the hospital catchment area, had evidence of acute respiratory infection, and radiographically-confirmed pneumonia within 72 hours of admission. Clinical, demographic, and epidemiologic information were collected through interviews and medical chart review. Blood, urine, and respiratory specimens for diagnostic testing using multiple modalities were obtained from enrolled patients. Informed consent was obtained and the study protocol was approved by the institutional review boards at each institution and the CDC.

Real-time PCR was performed on naso-/oropharyngeal (NP/OP) swabs for detection of influenza viruses; a positive PCR test was defined as a cycle threshold value of <40 for either influenza A or B virus. Influenza serology utilized both hemagglutination inhibition (HI) and microneutralization (MN) assays to define positive results for available paired acute and convalescent serum specimens [157, 158]. For both influenza A and B viruses, HI assay was performed; specimens positive for influenza B virus were further tested using the MN assay to improve specificity [143] . A positive serologic test was defined as a ≥4-fold rise in agent-specific IgG antibody titer between

paired acute and convalescent sera. If the influenza serology results indicated seroconversion when the vaccine was administered (based on self/caregiver report or vaccine verification) within two weeks before acute-phase serum collection, or between acute-phase and convalescent-phase serum collections, results were deemed inconclusive and were considered as missing serology results for this analysis.

A bacterial pathogen was defined as being present if *Chlamydophila pneumoniae* or *Mycoplasma pneumoniae* was detected in the NP/OP swab by a PCR assay or if bacteria (e.g., *Staphylococcus aureus, Haemophilus influenzae*, *S. pneumoniae,* and *S. pyogenes*) were detected in blood, endotracheal aspirate, or bronchoalveolar-lavage specimen by culture or, in pleural fluid, by culture or a PCR. For children, a bacterial pathogen was also defined as being present if bacteria were detected in whole blood by PCR for *S. pneumoniae or S. pyogenes.* For adults, a bacterial pathogen was also defined as being present if *S. pneumoniae* or *L. pneumophila* was detected by the urine antigen test or if bacteria were detected in high-quality sputum by culture. A summary variable of any bacterial detections was created.

We categorized both influenza PCR and serology test results dichotomously as either positive or negative using definitions described above. In the EPIC study and for these analyses, influenza virus was considered to be detected if there was a positive result from PCR or serology. Patients were included in this analysis if they met final radiographic criteria for CAP based on a study radiologist review and had samples available for both bacterial and viral detection.

Statistical Analysis

Outcome

The outcome for this analysis was influenza-associated pneumonia. For our primary analysis, influenza-associated pneumonia was defined based on results from multiple imputation that accounted for missing serology results. Missing influenza serology results were imputed in children and adults separately. Multiple imputation uses variables among patients with the available observed data to impute data for patients with missing data via a multivariable regression model [145, 146, 160]. The missing data were assumed to be missing at random [145, 146] through preliminary analyses described elsewhere. We then applied a modeling strategy to identify variables using available data for imputation of missing serology data; details are described in the supplementary appendix (Tables S4.1-S4.2). Our final multiple imputation model for children included age, sex, any bacterial detections, influenza PCR result, and influenza season. For adults, our final model included age, sex, any bacterial detections, influenza PCR result, and self-reported abdominal symptoms. Using these final variables, we applied multiple imputation to impute missing serology results. Twenty imputed datasets were created [161]. We combined the observed and imputed results to create a dichotomous outcome of influenza-associated pneumonia (hereafter referred to as 'MI-based outcome'). For a secondary analysis, influenza-associated pneumonia was based on observed PCR and serology study results (hereafter referred to as 'observed outcome'). For both definitions, a patient was classified as having influenza-associated pneumonia if PCR or serology results were positive, and a patient was classified as negative otherwise.

Covariates

We determined whether independent factors readily available early during hospitalization were associated with influenza-associated pneumonia. We considered

symptoms, clinical signs, white blood cell (WBC) count, underlying medical conditions, demographic characteristics, influenza season, receipt of influenza vaccine, and characteristics on chest radiograph. Most of these data were collected through medical chart review. Clinical signs were assessed by a physical examination, and symptoms were either self-reported or caregiver-reported. Influenza vaccination history was collected during the patient interview and medical charts were further reviewed to verify vaccination status. Vaccine receipt included the monovalent influenza A(H1N1)pdm09 vaccine (2009-2010 influenza season) or trivalent inactivated or live attenuated seasonal influenza vaccines (2010-2011 and 2011-2012 influenza seasons). Influenza season was defined as October 1 through April 30 for each study year.

We categorized age as <2 years, 2-4 years, 5-9 years, and 10-17 years for children and 18-49 years, 50-64 years, and ≥65 years for adults. WBC count was categorized as leukopenia (WBC <5,500/mm^3) or leukocytosis (WBC >15,000/mm^3) for children <5 years old. For those ≥5 years old, WBC count was categorized as leukopenia (WBC <3,000/mm^3) or leukocytosis (WBC >11,000/mm^3). Normal WBC count was the reference for leukopenia and leukocytosis. Presence of any underlying medical condition included asthma; chronic obstructive pulmonary disease (adults only); congenital heart disease (children only); coronary artery disease; pre-term birth (defined as gestational age <37 weeks at birth in children <2 years old); diabetes mellitus; chronic kidney disease; chronic liver disease; immunosuppression; any cancer (excluding skin cancers); neurological disorders; and chromosomal disorders.

Modeling Strategy and Developing Prediction Scores

Separate prediction models were constructed for children and adults. We conducted bivariate and multivariable analyses to assess which covariates were associated with influenza-associated pneumonia using logistic regression; we first used the MI-based outcome definition and then repeated the assessment using the observed outcome. Covariates with $p<0.20$ were eligible for inclusion in the multivariable prediction model. Biological plausibility was also considered in selecting covariates for potential inclusion in the final model. We constructed the multivariable model using two approaches. First, we included all eligible covariates as well as the interactions only between the demographic characteristics that were eligible for inclusion in a logistic regression model. We then eliminated covariates one at a time, starting with the covariate with the least non-statistically significant, until all remaining covariates had $p<0.05$ or significantly increased the -2 log-likelihood of the model. In the second approach, we added a covariate one at a time, starting with the covariate with the lowest p-value that was <0.05, until all included covariates had $p<0.05$ or significantly increased the -2 log-likelihood of the model.

We initially excluded specific covariates that may not have been readily available on admission or routinely performed, including chest radiography and verified receipt of the influenza vaccine. After the initial models were constructed, we added each of these covariates to assess whether it improved the predictive performance of the model; a given covariate was ultimately added if there was a meaningful increase in discrimination based on qualitative judgment.

We then developed prediction scores based on the final multivariable models for children and adults. A point value was assigned to each predictor in the model by

rounding its coefficient to the nearest integer. An individual's score was calculated by adding the corresponding point value for each of the observed characteristics. For consecutive cut-offs of the summed scores, the sensitivity, specificity, and positive and negative predictive values were calculated.

Evaluation of Model Performance and Internal Validation

We evaluated the model performance using the measures of discrimination and calibration. Discrimination indicates how well the model distinguishes between patients with and without the outcome of interest. It is measured by the c-statistic, the area under the receiver operating characteristic (ROC) curve. C-statistic values range from 0.5 (no discrimination) to 1.0 (perfect discrimination) [177]. Calibration indicates how well the model fits the data. It was assessed using the Hosmer-Lemeshow goodness of fit test; a p-value <0.05 was used to indicate that the model did not fit the data well.

Prediction models usually perform well in the original population, but can fail to reliably predict the outcome in another population; this phenomenon describes over-fitting, which leads to optimism in a model's performance [178]. This optimism can be assessed using bootstrapping methods [148, 179]. We performed an internal validation of the prediction models using bootstrapping to evaluate the potential degree of over-fitting. Bootstrapping draws a random sample with replacement within the original sample [147, 148]; these samples are the same size as the original sample. Three hundred bootstrap samples were drawn [147]. The average difference in the c-statistic between the bootstrap sample and the original sample indicates the optimism [147, 148].

Sensitivity Analyses

We performed a sensitivity analysis in which we provided an alternative definition for the influenza season to that used in the main analysis. Details about these sensitivity analyses are presented in the supplementary appendix.

Results

Children

Of 2636 enrolled children, 2358 (89.4%) met the final radiographic criteria for CAP, among whom 2222 (94.2%) had samples available for both bacterial and viral diagnostic tests. Median age was 2 years (interquartile range [IQR] 1-6); 50.9% of children had an underlying medical condition (Table 4.1). Among all 2222 children, 233 (10.5%) and 149 (6.7%) had influenza-associated pneumonia using the MI-based outcome and observed outcome definitions, respectively.

Among children, we identified clinical factors that were eligible for inclusion in the multivariable models using the MI-based outcome and the observed outcome definitions of influenza-associated pneumonia (Table 4.2). In the multivariable model using the MI-based outcome for children, clinician-reported confusion and influenza season remained associated with higher odds of influenza-associated pneumonia. Leukocytosis was associated with lower odds of the outcome, compared with normal WBC count. Additionally, children <2 years and 2-4 years old had decreased odds of influenza-associated pneumonia, compared with children 10-17 years old (Table 4.3). For the observed outcome definition of influenza-associated pneumonia, only influenza season remained significantly associated with higher odds of the outcome in the

multivariable model. Children <2 years old and 2-4 years old had decreased odds of influenza-associated pneumonia, compared with children aged 10-17 years (Table 4.3).

For the prediction score using the MI-based outcome, the following variables with associated point values were included: age <2 years old (subtract 1 point), 2-4 years old (subtract 1 point), 5-9 years old (add 0 points), clinician-reported confusion (add 1 point), leukopenia (add 1 point), leukocytosis (subtract 1 point), and hospitalization during the influenza season (add 1 point). This influenza-associated pneumonia prediction score ranged from 0 to 3; as the score values increased, the sensitivity decreased and the specificity increased. Using the MI-based outcome, the positive predictive values (PPVs) increased from 10% to 100% and the negative predictive values (NPVs) ranged from 90-91% as the score went from 0 to 3 (Table 4.4). In contrast, the prediction score using the observed outcome ranged from 0 to 1 as influenza season was the only factor that was associated with higher odds of the outcome. For a score value of 1, the sensitivity was 38% and the specificity was 80%. The PPVs ranged from 7-12% and the NPV was 95% as the score went from 0 to 1 (Table 4.4).

The models using the MI-based outcome and the observed outcome had c-statistics of 0.65 and 0.64, respectively, indicating poor discrimination between children with and without influenza-associated pneumonia. Both models also indicated good calibration, as the p-values for the goodness of fit test were 0.68 for the MI-based outcome and 0.99 for the observed outcome. Using the MI-based outcome, the c-statistic from the models of 300 bootstrap samples ranged from 0.61-0.70; there was an estimated average optimism of -0.01 in the original data. Using the observed outcome, the c-

statistic from models of 300 bootstrap samples ranged from 0.60-0.69; there was an estimated average optimism of -0.01.

We also examined the inclusion of additional characteristics to the multivariable models to determine whether they improved the model performance. The verified receipt of the influenza vaccine was associated with decreased odds of influenza-associated pneumonia using both the MI-based outcome (aOR=0.62, 95% CI: 0.42-0.92) and the observed outcome (aOR=0.62, 95% CI: 0.41-0.96). Adding this factor to both models increased the model discrimination by 0.01 but did not provide further meaningful discrimination; this variable was not retained in either of the prediction scores. Results from the sensitivity analyses are presented in the supplementary appendix.

Adults

Of 2488 enrolled adults, 2320 (93.2%) met the final radiographic criteria for CAP, among whom 2259 (97.3%) had specimens available for bacterial and viral diagnostic testing. Median age was 57 years (IQR 46-71); 78.2% of adults had an underlying medical condition (Table 4.1). Among all 2259 adults, 165 (7.3%) and 132 (5.8%) had influenza-associated pneumonia using the MI-based outcome and observed outcome definitions, respectively.

Among adults, we identified clinical factors that were eligible for inclusion in the multivariable models using the MI-based outcome and the observed outcome definitions of influenza-associated pneumonia (Table 4.5). In the final multivariable model using the MI-based outcome, cough, abdominal pain, and influenza season were significantly associated with influenza-associated pneumonia. Presence of any underlying medical conditions and leukocytosis were associated with decreased odds of influenza-associated

pneumonia (Table 4.6). In contrast, using the observed outcome, factors significantly associated with influenza-associated pneumonia were leukopenia, cough, abdominal pain, diarrhea, and influenza season. Leukocytosis was associated with decreased odds of influenza-associated pneumonia (Table 4.6).

Using the MI-based outcome, the prediction score ranged from 0 to 4. As the score value increased from 0 to 4, the PPVs increased from 7% to 29%, and the NPVs ranged between 93-96% (Table 4.7). The observed outcome score ranged from 0 to 7, with the PPVs increasing from 0 to 18% and NPVs ranging from 94-99% (Table 4.7).

The models using the MI-based outcome and the observed outcome had c-statistics of 0.72 and 0.77, respectively, indicating good discrimination. Both models also had good calibration, with p-values for the goodness of fit test=0.15 for MI-based outcome and 0.51 for observed outcome. Using the MI-based outcome, the c-statistic from models of 300 bootstrap samples ranged from 0.67-0.78; there was an estimated average optimism of -0.01 in the original data. Using the observed outcome, the c-statistic from models of 300 bootstrap samples ranged from 0.71-0.83; there was an estimated average optimism of 0.01.

Additional characteristics were considered for inclusion in the multivariable models to determine whether they improved the model performance. Using the MI-based outcome, pleural effusion was associated with decreased odds of influenza-associated pneumonia (aOR=0.45, 95% CI: 0.29-0.69); the c-statistic increased by 0.02. Using the observed outcome, pleural effusion (aOR=0.45, 95% CI: 0.29-0.69) and other infiltrates (aOR=1.66, 95% CI: 1.14-2.42) were associated with influenza-associated pneumonia. Including both characteristics in this model increased the c-statistic by 0.02. None of

these characteristics were retained in the prediction scores because they did not provide further meaningful discrimination for influenza-associated pneumonia using either definition. Results from the sensitivity analyses are presented in the supplementary appendix.

Discussion

We developed and evaluated prediction models and scores for influenza-associated pneumonia, using prospectively collected clinical and microbiological data from patients hospitalized with CAP enrolled in the EPIC study; to our knowledge, these are the first such models and scores developed among hospitalized children. Using two outcome definitions of influenza-associated pneumonia, age and influenza season were significant predictors among children; leukocytosis, presence of any underlying medical conditions, cough, abdominal pain, and influenza season were significant predictors among adults. These multivariable models had poor discrimination for children and good discrimination for adults. Moreover, the overall low PPVs from the prediction scores for children and adults may indicate difficulty in predicting influenza-associated pneumonia, especially in children. However, significant risk factors could be considered by physicians to inform influenza testing for early identification of influenza-associated pneumonia, if a child or adult had a hospitalization during the influenza season or if an adult had a cough or abdominal pain.

Among children hospitalized with CAP, few significant predictors (age and influenza season) were identified using both definitions of influenza-associated pneumonia. Influenza season was the only independent risk factor. Additionally,

children aged <2 years and aged 2-4 years had decreased odds of influenza-associated pneumonia compared with children aged 10-17 years. In one study of influenza hospitalizations, patients <2 years old had lower odds of pneumonia compared with those aged ≥50 years old [107], whereas another study indicated that patients 6-23 months old and 2-4 years old were more likely to have pneumonia than children 5-17 years old [121]. These studies compared influenza-positive patients with and without pneumonia; our study differed because we compared influenza-associated pneumonia to CAP due to other etiologies. Across these studies, few predictors of influenza-associated pneumonia were identified; there were also differences in predictors identified and in the direction of these associations. Identifying predictors using either observed or imputed definitions may be difficult for influenza-associated pneumonia in children, as the clinical features of influenza-associated pneumonia appear to be similar to other etiologies of community-acquired pneumonia among hospitalized children.

These pediatric prediction scores also reflect the difficulty in predicting influenza-associated pneumonia in children. First, the models using both outcome definitions had poor discrimination between children with and without influenza-associated pneumonia; this likely reflects the difficulty in differentiating CAP due to other pathogens, as 25% of all children had co-detections. Further, 50.6% and 63.7% of imputed or observed influenza-associated pneumonia outcomes had another pathogen detected, respectively. Additionally, there were overall low PPVs from the scores for both outcomes. However, for the MI-based outcome, higher PPVs (67-100%) were observed for having at least two of the following features: leukopenia, clinician-reported confusion, and hospitalization during the influenza season; however, 9 children (0.4%) had at least two of these factors.

Predicting influenza-associated pneumonia is difficult in children, as reflected by the poor discrimination and lower PPVs.

Among adults hospitalized with CAP, several predictors of influenza-associated pneumonia were identified using both definitions. Presence of any underlying medical conditions and leukocytosis were associated with decreased odds of influenza-associated pneumonia, which were also observed in other studies [126, 180]. Additionally, we identified cough, abdominal pain, and influenza season as risk factors, whereas null associations of cough and abdominal pain were found in these studies between patients with and without influenza-associated pneumonia. Diarrhea has also been shown to be a risk factor for influenza-associated pneumonia [107, 126]; this was only identified as a risk factor using the observed outcome definition. Of note, one of these studies included children and adult patients who were influenza-positive and compared characteristics between those with and without pneumonia. These several factors identified using both outcome definitions may provide meaningful combinations in predicting influenza-associated pneumonia in adults.

The multivariable models had good discrimination between adults with and without influenza-associated pneumonia, which is partially due to the identification of several predictors. Another possible reason for the good discrimination is that only 5% of adults had co-detections, which may have allowed for better differentiation between those with influenza-associated pneumonia and with CAP due to other etiologies. However, the prediction scores using both outcome definitions had low PPVs, which likely reflects the overall lack of specificity of CAP clinical features in adults. Additionally, because cough, abdominal pain, and influenza season had positive point

values in these scores, physicians could consider these factors to guide influenza testing for patients with suspected influenza-associated pneumonia. Despite good discrimination using both definitions, predicting influenza-associated pneumonia may be difficult in adults as reflected by lower PPVs.

To our knowledge, no formal prediction score has been developed and published for influenza-associated pneumonia in children, although one is available for hospitalized adults. In a study among hospitalized adults, a multivariable model and score was created for distinguishing between influenza A(H1N1)pdm09 pneumonia from interpandemic CAP [180]. Significant risk factors were age ≤65 years, white cell count $\leq 12x10^9/l$, bilateral radiographic change, absence of confusion, and temperature ≥38°C. This multivariable model had good discrimination, though no measures of statistical evaluation were reported for the prediction score. Selection bias is possible due to clinically-driven testing and a shift in clinical behaviors, as interpandemic CAP patients and influenza A(H1N1)pdm09 patients were enrolled during different time periods. Among children, a multivariable model and risk score were developed for predicting hospitalization for influenza virus infection [181], though this study was not able to predict complications due to influenza infection, including pneumonia. Significant risk factors for hospitalization were high-risk medical condition, respiratory distress on examination, radiographic evidence of pneumonia, and influenza B infection; however, no measures of statistical evaluation were reported. The PPVs for this score ranged from 74-98%, though the corresponding point values only included 4-27% of the population. From our prediction scores, we addressed a gap for children by developing the first published prediction score for influenza-associated pneumonia; additionally, we

overcame limitations from a published model of influenza-associated pneumonia in adults.

Influenza testing and prescription of antiviral treatment are underutilized in hospitals, and thus a useful prediction score could improve use of both diagnostics and treatment. Among prospective surveillance studies, 16-29% of patients with a culture/PCR-confirmed influenza received a clinical diagnosis of influenza [182, 183]; as influenza virus infections could have been missed without appropriate testing, there is a need for more comprehensive influenza testing by physicians. Similarly, antiviral prescriptions have also been underutilized among influenza hospitalizations. Using population-based surveillance data, the proportion of hospitalized children with laboratory-confirmed influenza virus infection who received antiviral treatment increased from 37-48% during 2003-2008 [106] to 79% during the 2009 H1N1 pandemic and then decreased to 56% during 2010-2011 [184]. In adults, antiviral treatment increased from 51-57% during 2005-2008 to 82% during the 2009 pandemic [185] and then decreased to 77% during 2010-2011 [184]. These studies illustrate the gaps between clinical practice and the CDC recommendations for influenza antiviral therapy as soon as possible for all persons with suspected or confirmed influenza requiring hospitalization [122]. Thus, a meaningful prediction score may inform whom clinicians should test and treat, rather than solely relying on clinical judgment.

There are limitations to these analyses. First, more than 50% of enrolled children and adults were missing serology results. Imputing these data assumed they were missing at random, which we believed was reasonable based on our analyses. Second, the inclusion criteria for the EPIC study could have obscured determining whether any of

these factors were predictive of influenza-associated pneumonia because enrolled patients were required to have evidence of acute infection (fever, chills, or abnormal WBC count) and evidence of acute respiratory infection (new cough, sputum production, chest pain, dyspnea, tachypnea, abnormal lung examination, or respiratory failure). Nevertheless, some clinical factors were identified as predictors. Third, co-detections may have also obscured the prediction of CAP due to influenza virus infection, as 25% of children and 5% of adults had more than one pathogen detected. Fourth, symptoms were self or caregiver-reported, which introduces potential misclassification.

In conclusion, significant factors were identified for prediction of influenza-associated pneumonia based on observed and imputed PCR and serology results, and prediction scores with these factors were developed among children and adults hospitalized with CAP. Despite using a breadth of readily available factors, our scores may not sufficiently predict influenza-associated pneumonia, particularly in children. However, due to the underutilization of influenza testing and prescription of antiviral treatment in hospitals, significant factors from these scores in adults may be considered by physicians for influenza testing in managing patients with suspected influenza-associated pneumonia. As influenza season was a strong risk factor for children and adults, influenza testing should continue to be encouraged during these calendar months. The difficulty of predicting influenza-associated pneumonia, especially for children, underscores the importance of developing meaningful tools for early diagnosis and identification.

Supplementary Appendix

Additional Methods for Multiple Imputation

We compared demographic and clinical characteristics between patients with and without available serology data. Because no meaningful differences were found (Tables S4.1-S4.2), we felt the assumption of serology data missing at random was plausible.

Using both bivariate and multivariable analyses, we identified independent variables that could be used to impute missing serology results as either positive or negative. First, we constructed "*a priori*" models for children and adults using age, sex, any bacterial detection, and influenza PCR result chosen *a priori* based on biological plausibility. We hypothesized that influenza PCR results would be positively correlated with influenza serology results, given the concordant results from both methods. Additionally, we hypothesized that any bacterial detection would be negatively correlated with influenza serology results, as patients with positive influenza serology results were less likely to have any bacteria detections compared to patients with negative influenza serology results.

Second, we assessed whether additional variables, including cough, fever, diarrhea, sore throat, abdominal pain, myalgia, chills, headache, influenza season, chest indrawing/retraction, underlying medical conditions, radiographic findings (e.g., consolidation, pleural effusion, and other infiltrates), ICU admission, and pneumonia severity (defined as having experienced at least one of the following: ICU admission, mechanical ventilation, acute respiratory distress syndrome, shock, or death), were significant in bivariate analyses. For the variables that were significant in the bivariate analyses, each variable was added one at a time to the *a priori* model to assess whether it

was significant in the multivariable model. Model fit was evaluated using the Hosmer-Lemeshow goodness of fit test; a p-value <0.05 was used to indicate that the model did not fit the data well. Additionally, we compared the area under the curve between the *a priori* model with and without the potential variable. A non *a priori* variable was used to impute serology data if it was significant in both the bivariate and multivariable analyses and the area under the curve was higher for the *a priori* model with this variable relative to the model without it. Potential variables were assessed for children and adults separately.

Methods for Sensitivity Analysis: Alternative definition for influenza season using surveillance data

We used data from CDC's U.S. outpatient influenza-like illness (ILI) surveillance network (ILINet) for the influenza seasons corresponding to the EPIC study years. We obtained the percentage of visits for ILI for each calendar week by influenza season and by region that corresponded to each study city. Defining influenza season was based on the weeks when the percentage of visits for ILI was above the regional baseline for a given influenza season; thus, this definition could have varied by region and by calendar year depending on the influenza circulation. Hospitalizations that occurred during these weeks were defined as being in the influenza season, and hospitalizations that occurred any time otherwise were defined as not being in the influenza season. We applied this alternative definition for influenza season in the main and secondary analyses.

Results for Sensitivity Analysis: Alternative definition for influenza season using surveillance data

Children

In this sensitivity analysis, we alternatively defined influenza season using CDC's ILI-Net surveillance data on influenza hospitalizations within the primary and secondary analyses. The multivariable models, including the covariates and their coefficients, using both outcome definitions remained the same for children. As a result, the point value for each covariate did not change, based on the definition of influenza season (data not shown).

Based on these models that incorporated the alternative definition for influenza season, we created prediction scores for the MI-based outcome and the observed outcome. For the prediction score using the MI-based outcome, the following variables with corresponding point values were included: age <2 years old (subtract 1 point), 2-4 years old (subtract 1 point), 5-9 years old (add 0 points), clinician-reported confusion (add 1 point), leukopenia (add 1 point), leukocytosis (subtract 1 point), and hospitalization during the influenza season (add 1 point). As the score values went from 0 to 3, the PPVs increased from 29% to 100% and the NPVs were 90%. In contrast, the prediction score using the observed outcome ranged from 0 to 1, as influenza season was the only factor that was associated with higher odds of the outcome. For a score value of 1, the sensitivity was 11% and the specificity was 96%. The PPVs increased from 7% to 18% and the NPV was 94%. From the prediction scores using both definitions, the PPVs were slightly higher and the NPVs were similar than those from the main and secondary analyses.

The models using the MI-based outcome and the observed outcome had c-statistics of 0.62 and 0.62, respectively, and indicated poor discrimination between children with and without influenza-associated pneumonia; these c-statistics were 0.02-

0.03 lower than those in the main and secondary analyses. Both models also indicated good calibration, as the p-values for the goodness of fit test for the MI-based outcome and for the observed outcome were 0.50 and 0.69, respectively. Using the MI-based outcome, the c-statistic from the models of 300 bootstrap samples ranged from 0.58-0.69; there was an estimated average optimism of -0.01 in the original data. Using the observed outcome, the c-statistic from models of 300 bootstrap samples ranged from 0.58-0.712; there was an estimated average optimism of -0.01.

We also examined whether additional characteristics improved the model performance. Receipt of the influenza vaccine was associated with decreased odds of influenza-associated pneumonia (aOR=0.60, 95% CI: 0.41-0.89). Adding this variable to the model increased the model discrimination by 0.01 but did not provide meaningful discrimination; this variable was not retained in the model.

Adults

The multivariable models, including the covariates and their coefficients, using both outcome definitions remained the same for adults. As a result, the point value for each covariate did not change, based on the definition of influenza season (data not shown).

Using the MI-based outcome, the prediction score ranged from 0 to 4. As the score value increased from 0 to 4, the PPV increased from 7% to 38%, and the NPVs ranged from 93-96%. The observed outcome score ranged from 0 to 7, with the PPVs increasing from 6% to 33% and NPV ranging from 94-99%. From both prediction scores, the PPVs were slightly higher and the NPVs were similar than those from the primary and secondary analyses.

The models using the MI-based outcome and the observed outcome had c-statistics of 0.74 and 0.78, respectively, and indicated good discrimination; these were 0.01-0.02 higher than those from the main and secondary analyses. Both models also had good calibration (p-value for the goodness of fit test=0.39 for MI-based outcome and 0.31 for observed outcome). Using the MI-based outcome, the c-statistic from models of 300 bootstrap samples ranged from 0.67-0.79; there was an estimated average optimism of 0.01 in the original data. Using the observed outcome, the c-statistic from models of 300 bootstrap samples ranged from 0.71-0.84; there was an estimated average optimism of -0.0006.

Additional characteristics were considered for inclusion in the multivariable models to determine whether they improved the model performance. Using the MI-based outcome, pleural effusion (aOR=0.48, 95% CI: 0.31-0.74) and other infilrates (aOR=1.45, 95% CI: 1.03-2.03) were associated with influenza-associated pneumonia; the c-statistic increased by 0.01. Using the observed outcome, pleural effusion (aOR=0.44, 95% CI: 0.27-0.73) and other infiltrates (aOR=1.80, 95% CI: 1.23-2.63) were associated with influenza-associated pneumonia. Including both characteristics in this model increased the c-statistic by 0.02, though they did not provide further meaningful discrimination. These characteristics were not retained in either prediction score.

Table 4.1. Demographic and clinical characteristics of children and adults hospitalized with community-acquired pneumonia

	Children (n=2222) No. (%)	Adults (n=2259) No. (%)
Age groups		
<2 years	980 (44.0%)	---
2-4 years	559 (25.2%)	---
5-9 years	408 (18.4%)	---
10-17 years	275 (12.4%)	---
18-44 years	---	509 (22.5%)
45-64 years	---	945 (41.8%)
65+ years	---	805 (35.6%)
Sex		
Male	1226 (55.2%)	1104 (48.9%)
Race/Ethnicity		
Non-Hispanic White	872 (39.2%)	1054 (46.7%)
Non-Hispanic Black	765 (34.4%)	874 (38.7%)
Hispanic	414 (18.6%)	238 (10.5%)
Other	171 (7.8%)	93 (4.1%)
Underlying Medical Conditions*		
None reported	1090 (49.1%)	493 (21.8%)
Asthma	743 (33.4%)	584 (25.9%)
Preterm birth in children under 2 years old	205 (9.2%)	---
Congenital heart disease	159 (7.2%)	---
Immunosuppression	33 (1.5%)	368 (15.8%)
Chronic kidney disease	26 (1.2%)	356 (15.8%)
Heart failure	22 (1%)	430 (19.0%)
Cancer	9 (0.4%)	405 (17.9%)
Diabetes mellitus	7 (0.3%)	584 (25.9%)
Chronic liver disease	6 (0.3%)	126 (5.6%)
Coronary artery disease	---	663 (29.4%)
Chronic obstructive pulmonary disease (COPD)	---	520 (23.0%)
Study City		
Chicago	---	1507 (66.7%)
Memphis	842 (37.9%)	---
Nashville	600 (27%)	752 (33.3%)
Salt Lake City	780 (35.1%)	---

*Underlying medical conditions included asthma, chronic obstructive pulmonary disease (adults only); congenital heart disease (children only), coronary artery disease; pre-term birth (defined as gestational age <37 weeks at birth in children under 2 years old); diabetes mellitus; chronic kidney disease; chronic liver disease; immunosuppression; any cancer (excluding skin cancers); neurological disorders (including seizure, cerebral palsy, scoliosis); and chromosomal disorders (including Down's syndrome).

Table 4.2. Predictors that were eligible for inclusion in multivariable models for influenza-associated pneumonia, using both outcome definitions based on multiple imputation and based on observed results from PCR and serology, among 2222 children hospitalized with community-acquired pneumonia

Type of Predictor	Predictor	MI-based outcome	Observed outcome
Symptoms	Cough	X	
	Anorexia/loss of appetite		X
	Chest indrawing/retraction	X	X*
	Chills	X*	X*
	Sore throat	X*	
	Muscle aches	X*	X*
	Difficult to wake or rouse	X*	
	Altered mental status	X*	X
	Headache	X*	X*
	Conjunctivitis		X
	Chest pain	X	X
	Abdominal pain	X	
Clinical Signs	Altered mental status	X*	X
	Chest indrawing		X
	Wheezing		X
	Dullness to percussion	X	
	Egophony	X	
Lab Finding	WBC count	X	
Underlying Medical Conditions	Presence of any underlying medical condition	X	X
	Cardiovascular disease	X	X
Demographic Characteristics	Age	X*	X*
	Sex	X	

*$p<0.05$

Table 4.3. Predictors of influenza-associated pneumonia, using definitions based on PCR and serology results from multiple imputation and based on observed results from these tests, among 2222 children hospitalized with community-acquired pneumonia.

	Definition based on PCR and serology results from multiple imputation			Definition based on observed PCR and serology results		
	Adjusted Odds Ratios (aOR)	95% Confidence Interval (CI)	Prediction Score Value*	aOR	95% CI	Prediction Score Value*
Age group:						
<2 years old	0.58	0.37-0.91	-1	0.51	0.32-0.81	-1
2-4 years old	0.46	0.27-0.76	-1	0.46	0.27-0.79	-1
5-9 years old	1.07	0.67-1.73	0	0.84	0.50-1.41	0
10-17 years old	1.00	(ref)	0	1.00	(ref)	0
Clinician-reported confusion	2.25	1.08-4.70	1	---	---	---
WBC count:						
Leukopenia	1.66	0.90-3.04	1	---	---	---
Normal	1.00	(ref)	0	---	---	---
Leukocytosis	0.69	0.50-0.96	-1	---	---	---
Influenza Season	3.06	1.98-4.73	1	3.06	1.85-5.06	1

*Prediction score value was determined by rounding the beta coefficient to the nearest integer.

Table 4.4. Prediction scores for influenza-associated pneumonia, using definitions based on PCR and serology results from multiple imputation and based on observed results from these tests, among 2222 children hospitalized with community-acquired pneumonia.

Prediction Score Value	n (%)	Sensitivity*	Specificity*	Positive Predictive Value* (PPV)	Negative Predictive Value* (NPV)
Definition based on PCR and serology results from multiple imputation					
≤0	1915 (86.2%)	100%	0%	10%	N/A
1	298 (13.4%)	24%	87%	18%	91%
2	8 (0.4%)	3%	100%	67%	90%
3	1 (0.1%)	0%	100%	100%	90%
Definition based on observed PCR and serology results					
≤0	1749 (78.7%)	100%	0%	7%	N/A
1	473 (21.3%)	38%	80%	12%	95%

*For influenza-associated pneumonia, using each cut-off at the corresponding value.

Table 4.5. Predictors that were eligible for inclusion in multivariable models for influenza-associated pneumonia, using both outcome definitions based on multiple imputation and based on observed results from PCR and serology, among 2259 adults hospitalized with community-acquired pneumonia.

Type of Predictor	Predictor	MI-based outcome	Observed outcome
Symptoms	Cough	X*	X*
	Fever	X*	X*
	Diarrhea	X*	X*
	Sore throat	X	X*
	Headache	X	X
	Cough with sputum	X	
	Chest pain	X	X
	Abdominal pain	X*	X*
Clinical Signs	Altered mental status		X
	Egophony	X	X*
Lab Finding	WBC count	X*	X*
Underlying Medical Conditions	Presence of any underlying medical condition	X*	X*
	Cardiovascular disease	X*	X*
	Neurological disease	X	X
Demographic Characteristics	Race/ethnicity	X	
	Age	X*	X*

*$p<0.05$

Table 4.6. Predictors of influenza-associated pneumonia, using definitions based on PCR and serology results from multiple imputation and based on observed results from these tests, among 2259 adults hospitalized with community-acquired pneumonia.

	Definition based on PCR and serology results from multiple imputation			Definition based on observed PCR and serology results		
	aOR	95% CI	Prediction Score Value	aOR	95% CI	Prediction Score Value
WBC count: Leukopenia	2.53	0.97-6.63	1	2.96	1.11-7.88	1
Normal	1	(ref)	0	1.00	(ref)	0
Leukocytosis	0.55	0.40-0.77	-1	0.42	0.28-0.62	-1
Presence of any underlying medical conditions	0.53	0.37-0.76	-1	0.52	0.35-0.76	-1
Cough	3.74	1.63-8.60	1	17.56	2.44-126.65	3
Abdominal pain	2.21	1.55-3.14	1	1.56	1.03-2.38	1
Diarrhea	---	---	---	1.57	1.04-2.37	1
Influenza Season	3.29	2.09-5.18	1	4.44	2.52-7.82	1

Table 4.7. Prediction scores for influenza-associated pneumonia, using definitions based on PCR and serology results from multiple imputation and based on observed results from these tests, among 2259 adults hospitalized with community-acquired pneumonia.

Prediction Score Value	n (%)	Sensitivity	Specificity	PPV	NPV
Definition based on PCR and serology results from multiple imputation					
≤0	1184 (52.4%)	100%	0%	7%	N/A
1	768 (34.0%)	75%	55%	11%	96%
2	276 (12.2%)	36%	88%	20%	95%
3	30 (1.3%)	5%	99%	29%	93%
4	1 (0.04%)	0%	100%	0%	93%
Definition based on observed PCR and serology results					
≤0	237 (10.5%)	100%	0%	6%	N/A
1	222 (9.8%)	98%	11%	6%	99%
2	592 (26.2%)	95%	21%	7%	99%
3	752 (33.3%)	83%	48%	9%	98%
4	351 (15.5%)	53%	82%	15%	97%
5	94 (4.2%)	17%	96%	22%	95%
6	10 (0.4%)	2%	100%	18%	94%
7	1 (0.1%)	0%	100%	0%	94%

Supplementary Table S4.1. Characteristics of Hospitalized Children with Community-acquired Pneumonia: Comparison of Children with and without available influenza serology data

	Children with available paired serology data (N=961) No. (%)	Children without available paired serology data (N=1261) No. (%)
Age groups		
<2 years	359 (37.4%)	621 (49.2%)
2-4 years	241 (25.1%)	318 (25.2%)
5-9 years	217 (22.6%)	191 (15.2%)
10-17 years	144 (15.0%)	131 (10.4%)
Sex		
Male	546 (56.8%)	680 (53.9%)
Female	415 (43.2%)	581 (46.1%)
Race/Ethnicity		
Non-Hispanic White	439 (45.7%)	433 (34.3%)
Non-Hispanic Black	235 (24.5%)	530 (42.0%)
Hispanic	213 (22.2%)	201 (15.9%)
Other	74 (7.7%)	97 (7.7%)
Underlying Medical Conditions		
None reported	484 (50.4%)	606 (48.1%)
Asthma	314 (32.7%)	429 (34.0%)
Congenital heart disease	80 (8.3%)	79 (6.3%)
Preterm birth in children under 2 years old	64 (6.7%)	141 (11.2%)
Immunosuppression	18 (1.9%)	15 (1.2%)
Heart failure	10 (1.0%)	12 (1.0%)
Chronic kidney disease	8 (0.8%)	18 (1.4%)
Cancer	6 (0.6%)	3 (0.2%)
Diabetes mellitus	4 (0.4%)	3 (0.2%)
Chronic liver disease	1 (0.1%)	5 (0.4%)
Hospital Indicators		
Disease severity	216 (22.5%)	250 (19.8%)
Intensive care unit admission	214 (22.3%)	249 (19.8%)
Mechanical ventilation	64 (6.7%)	85 (6.7%)
Death	0 (0%)	3 (0.2%)

Supplementary Table S4.2. Characteristics of Hospitalized Adults with Community-acquired Pneumonia: Comparison of Adults with and without available influenza serology data

	Adults with available paired serology data (N=846) No. (%)	Adults without available paired serology data (N=1413) No. (%)
Age groups		
18-44 years	184 (21.7%)	325 (23.0%)
45-64 years	384 (45.4%)	561 (39.7%)
65+ years	278 (32.9%)	527 (37.3%)
Sex		
Male	428 (50.6%)	676 (47.8%)
Female	418 (49.4%)	737 (52.2%)
Race/Ethnicity		
Non-Hispanic White	390 (46.1%)	664 (47.0%)
Non-Hispanic Black	327 (38.6%)	547 (38.7%)
Hispanic	96 (11.4%)	142 (10.1%)
Other	33 (3.9%)	60 (4.2%)
Underlying Medical Conditions		
None reported	170 (20.1%)	323 (22.9%)
Coronary artery disease	274 (32.4%)	389 (27.5%)
Asthma	227 (26.8%)	357 (25.3%)
Diabetes mellitus	222 (26.2%)	362 (25.6%)
Chronic obstructive pulmonary disease	187 (22.1%)	333 (23.6%)
Cancer	162 (19.2%)	243 (17.2%)
Heart failure	160 (18.9%)	270 (19.1%)
Immunosuppression	150 (17.7%)	218 (15.4%)
Chronic kidney disease	148 (17.5%)	208 (14.7%)
Chronic liver disease	53 (6.3%)	73 (5.2%)
Hospital Indicators		
Disease severity	181 (21.4%)	324 (22.9%)
Intensive care unit admission	173 (20.5%)	309 (21.9%)
Mechanical ventilation	33 (3.9%)	84 (5.9%)
Death	4 (0.5%)	45 (3.2%)

CHAPTER 5. Conclusions and Future Directions

Overview of Findings

The overall objective of this dissertation was to account for misclassification of pneumonia etiology from imperfect detections of influenza and non-influenza respiratory viruses, which was translated into two goals. The first goal was to estimate revised proportions of CAP due to adenovirus, human metapneumovirus, parainfluenza virus, respiratory syncytial virus, and influenza A/B virus. The second goal was to develop and evaluate prediction models and scores for influenza-associated pneumonia. For both goals, we compared the findings using the observed data from the EPIC study to the findings after accounting for potential misclassification. The objective and goals of this dissertation were achieved through three research aims.

In the first aim, we investigated potential misclassification from both missing diagnostic test results and imperfect test characteristics on burden estimates of pneumonia due to non-influenza viruses. We estimated the revised proportions of CAP due to adenovirus, human metapneumovirus, parainfluenza virus, and respiratory syncytial virus among children and adults hospitalized with CAP, accounting for missing serology results using multiple imputation. Multiple imputation estimates of these revised proportions were 0.8-3.2% higher in absolute differences and 1.1-3.0 times higher in relative terms than those estimated from the EPIC study among children and adults. One reason for these higher estimates is that we imputed missing serology results for more than half of enrolled children and adults. Moreover, it is likely that serology detects viruses that are not captured by PCR.

In the second aim, we investigated potential misclassification from both missing diagnostic test results and imperfect test characteristics on estimates of pneumonia due to influenza viruses. By accounting for missing serology results, we estimated the revised proportion of pneumonia due to influenza virus among children and adults hospitalized with CAP. From multiple imputation, this revised proportion was 2.1-4.4% higher in absolute differences and 1.4-1.7 times higher in relative terms than that directly estimated from the observed EPIC study data among children and adults. From the first and second aims, our results illustrated the potential underestimation of virus-specific pneumonia burden when serology results are either not included or missing.

In the third aim, we developed, evaluated, and validated prediction models and scores of influenza-associated pneumonia. To aid early identification and diagnosis of this outcome, we used clinical and demographic factors that are readily available at clinical presentation. We utilized two outcome definitions for influenza-associated pneumonia with and without accounting for missing serology results. The MI-based outcome was based on the revised proportion of influenza virus detections that accounted for missing serology results in the second aim. The observed outcome was defined using only observed results from PCR and serology and did not account for missing serology results. Among children, few significant predictors were identified in the models using both the MI-based outcome and observed outcome definitions; these models had good but weaker discrimination and good calibration. The prediction score for the MI-based outcome had positive predictive values (PPVs) increasing up to 100%, though <1% of children had scores that corresponded to higher PPVs. The score for the observed outcome had lower PPVs (7-12%). Predicting influenza-associated pneumonia is more

difficult in children, as reflected by identifying few predictors, poor discrimination, and lower PPVs. In contrast, among adults, several significant predictors were identified using both outcome definitions, and these models had good discrimination and calibration. Despite identifying several predictors among adults, the prediction scores had lower PPVs (0-29%) and may also indicate difficulty in predicting this outcome.

Overall, the findings from the three aims were consistent with our expectations. We anticipated that the revised proportions of CAP with respiratory viruses detected would be higher than the observed estimates from the EPIC study after accounting for missing serology results. Because missing serology results are imputed based on data from patients with available results, these available results serve as a lower bound of the proportion of patients with positive serology results. Thus, these revised proportions from imputing missing serology results cannot be lower than the observed proportions from the EPIC study. The only scenario in which the revised proportions could be equal to the observed proportions from the EPIC study is if imputing missing serology results did not increase the proportion of patients with positive serology results; however, when this proportion does increase, revised estimates will be higher under the believe-the-positive interpretation. Moreover, these higher estimates likely indicate that serology may detect viruses not captured by PCR. Among 955 (43.0%) children and 833 (36.9%) adults who had both PCR and serology results available, 0.2-7.4% had PCR-positive and serology-negative results for adenovirus, HMPV, parainfluenza virus, and RSV. Additionally, 3.0-6.8% had PCR-positive and serology-negative results for influenza virus among 870 (39.2%) children and 741 (32.8%) adults who had both PCR and serology results available. We also anticipated some difficulty in predicting influenza-

associated pneumonia among children and adults, despite the breadth of symptoms and clinical signs available as well as systematic influenza testing of all enrolled patients. One reason for this was that clinical features of pneumonia due to influenza virus are similar to those of CAP due to other etiologies. Additionally, the prediction of CAP due to influenza virus could have been obscured, as 25% of children and 5% of adults had more than one pathogen detected.

Overall Strengths and Limitations of Data and Analyses

There were strengths in utilizing data from the EPIC study to achieve these dissertation goals. One strength was the systematic specimen collection, prospective testing for multiple pathogens, and systematic performance of radiographs for all enrolled patients. Because diagnostic testing was not clinically driven, there may also be less bias in estimating revised proportions of respiratory virus detections and assessing whether radiographic characteristics predict influenza-associated pneumonia. Another strength is the breadth of symptoms and clinical signs available, which allows for a more complete assessment of potential predictors of influenza-associated pneumonia.

There were also limitations that broadly affected these analyses. More than 50% of enrolled children and adults were missing serology results for influenza and non-influenza respiratory viruses. Imputing these data assumed they were missing at random based on our analyses, which we believed was reasonable. Further, estimates of virus-specific pneumonia burdens from multiple imputation could have changed if other independent variables had been used to impute missing serology results. This concern may be minimal because results from another diagnostic test (PCR) were available and

were a strong predictor of serology results. Additionally, 25% of children and 5% of adults had more than one pathogen detected; this contributed to the difficulty in identifying a latent class labeled for a specific respiratory virus and also could have obscured the prediction of CAP due to influenza virus. These limitations primarily pertain to the analyses. However, a broader challenge remains in ascertaining infection status based on the interpretation of diagnostic test results. In particular, specimens from the upper respiratory tract (e.g., from a NP/OP swab) may not be reflective of the lower respiratory tract. Moreover, the detection of viruses in the NP/OP swabs is not necessarily indicative of the cause of CAP and may represent resolving infection rather than acute infection or an infection limited to the upper tract and not the lower tract. In this context, timing between illness onset and NP/OP specimen collection is important; a resolving infection could be reflected in a positive PCR result because PCR can detect non-viable genetic material, which may occur with a longer duration. Additionally, serology measures the virus-specific antibody response and does not detect the presence of a respiratory virus.

Implications and Future Directions

This dissertation had novel applications of statistical methods to address its objective and gaps in the literature. Overall, this was the first investigation into the potential misclassification of pneumonia etiology attributable to respiratory viruses from missing diagnostic test results and from imperfect test characteristics, especially in a hospitalized population. Specifically, to our knowledge, this was the first time that revised viral-specific pneumonia burdens have been estimated by applying multiple

imputation to account for missing diagnostic test results. This dissertation also included a novel application of LCA in estimating revised proportions of CAP with specific respiratory viruses detected, though there were potentially significant limitations. Finally, this dissertation addressed another gap in the literature by developing formal prediction models and scores for influenza-associated pneumonia for hospitalized children. Unlike other related models in the literature, our models and scores were also evaluated and validated to measure discrimination, calibration, and optimism that may inform their potential utility in a clinical setting; however, further validation may still be needed.

As estimates of virus-specific pneumonia burden may have been underestimated, further elucidation of diagnostic test characteristics may be warranted. Describing these characteristics could discern the likelihood of false-positive and false-negative results, though timing and quality of specimen collection should be considered. Within the literature, PCR and serology are considered to have high but not perfect specificity. The "believe the positive" approach may be appropriately applied if all tests (in this context, PCR and serology) have high but not perfect specificity, as combining imperfect tests will decrease the specificity from any one imperfect measure. Ultimately, because of the increase in sensitivity by combining PCR and serology results, we believe that applying the "believe the positive" approach was appropriate in this application. Because of the likely underestimated virus-specific pneumonia burdens, our findings support the development of more sensitive PCR methods for clinical purposes, which could inform population-based estimates, clinical guidance, and public health policy. Alternatively, if a third accurate diagnostic test became available for the detection of respiratory viruses,

LCA may be a more viable analytic option to potentially elucidate characteristics of available diagnostic tests.

In the third aim of this dissertation, we experienced difficulty in predicting influenza-associated pneumonia in children and adults hospitalized with CAP. We assessed a breadth of symptoms and clinical signs that were readily available at clinical presentation as potential predictors of influenza-associated pneumonia. Further exploration of other factors (e.g., vital signs or biomarkers) that could predict influenza-associated pneumonia may be warranted, though some factors may only be available in a subset of the patients. Despite the robust clinical data available, these scores may not be sufficiently predictive. For each score value, positive likelihood ratios were calculated for the scores in children and adults; higher ratio values indicate a larger increase in the likelihood of disease. Among children, the positive likelihood ratios were 1.9 and 17.1 for score values of 1 and 2, respectively, for both the MI-based and observed outcomes; however, 9 children (0.4%) had a score value of 2. Among adults, the positive likelihood ratios ranged from 1.1-5.2 with score values of 1-7 using the MI-based and observed outcomes; these ratios indicated a small to moderate increase in the likelihood of disease. Additional validation of these prediction scores developed for children and adults may still be warranted in other inpatient populations, especially in settings where diagnostic testing and radiographs are systematically performed on all patients. Even if these scores are sufficiently validated, these scores may not sufficiently predict the outcome to inform influenza testing and guide appropriate antimicrobial therapy in settings where not all patients with suspected influenza-associated pneumonia are tested for influenza and/or have a radiograph performed.

REFERENCES

1. Heron M. National Vital Statistics Report. Deaths: Leading Causes for 2010. Centers for Disease Control and Prevention. National Center for Health Statistics., December 20, 2013 December 20, 2013. Report No.
2. Number and Rate of Discharges by First-listed Diagnostic Categories, 2010. Centers for Disease Control and Prevention, National Center for Health Statistics, 2010.
3. Griffin MR, Zhu Y, Moore MR, Whitney CG, Grijalva CG. U.S. hospitalizations for pneumonia after a decade of pneumococcal vaccination. N Engl J Med. 2013;369(2):155-63. Epub 2013/07/12. doi: 10.1056/NEJMoa1209165. PubMed PMID: 23841730.
4. Alcon A, Fabregas N, Torres A. Pathophysiology of pneumonia. Clinics in chest medicine. 2005;26(1):39-46. Epub 2005/04/02. doi: 10.1016/j.ccm.2004.10.013. PubMed PMID: 15802164.
5. Wunderink RG, Waterer GW. Clinical practice. Community-acquired pneumonia. N Engl J Med. 2014;370(6):543-51. Epub 2014/02/07. doi: 10.1056/NEJMcp1214869. PubMed PMID: 24499212.
6. Mani C, Murray D. Acute Pneumonia and Its Complications. In: Long S, editor. Principles and Practice of Pediatric Infectious Diseases. Fourth ed. New York: Elsevier; 2012.
7. McIntosh K. Community-acquired pneumonia in children. N Engl J Med. 2002;346(6):429-37. Epub 2002/02/08. doi: 10.1056/NEJMra011994. PubMed PMID: 11832532.
8. Edelstein P, Cianciotto N. Legionella. In: Mandell GL, Bennett JE, Dolin R, editors. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. Philadelphia, PA: Churchill Livingstone/Elsevier; 2010. p. 2969-84.
9. Falsey AR, Walsh EE. Viral pneumonia in older adults. Clin Infect Dis. 2006;42(4):518-24. Epub 2006/01/20. doi: 10.1086/499955. PubMed PMID: 16421796.
10. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007;44 Suppl 2:S27-72. Epub 2007/02/06. doi: 10.1086/511159. PubMed PMID: 17278083.
11. World Health Organization. Technical bases for the WHO recommendations on the management of pneumonia in children at first-level health facilities. Geneva, Switzerland: World Health Organization, 1991.
12. Madhi SA, De Wals P, Grijalva CG, Grimwood K, Grossman R, Ishiwada N, et al. The burden of childhood pneumonia in the developed world: a review of the literature. Pediatr Infect Dis J. 2013;32(3):e119-27. Epub 2012/10/27. doi: 10.1097/INF.0b013e3182784b26. PubMed PMID: 23099423.
13. Cherian T, Mulholland EK, Carlin JB, Ostensen H, Amin R, de Campo M, et al. Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. Bull World Health Organ. 2005;83(5):353-9. Epub 2005/06/25. doi: /S0042-96862005000500011. PubMed PMID: 15976876; PubMed Central PMCID: PMC2626240.

14. World Health Organization Pneumonia Vaccine Trial Investigators' Group. Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children. October 2001.
15. Bradley JS, Byington CL, Shah SS, Alverson B, Carter ER, Harrison C, et al. The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. Clin Infect Dis. 2011;53(7):e25-76. Epub 2011/09/02. doi: 10.1093/cid/cir531. PubMed PMID: 21880587.
16. Marston BJ, Plouffe JF, File TM, Jr., Hackman BA, Salstrom SJ, Lipman HB, et al. Incidence of community-acquired pneumonia requiring hospitalization. Results of a population-based active surveillance Study in Ohio. The Community-Based Pneumonia Incidence Study Group. Arch Intern Med. 1997;157(15):1709-18. Epub 1997/08/11. PubMed PMID: 9250232.
17. Michelow IC, Olsen K, Lozano J, Rollins NK, Duffy LB, Ziegler T, et al. Epidemiology and clinical characteristics of community-acquired pneumonia in hospitalized children. Pediatrics. 2004;113(4):701-7. Epub 2004/04/03. PubMed PMID: 15060215.
18. Pavia AT. Viral infections of the lower respiratory tract: old viruses, new viruses, and the role of diagnosis. Clin Infect Dis. 2011;52 Suppl 4:S284-9. Epub 2011/04/06. doi: 10.1093/cid/cir043. PubMed PMID: 21460286; PubMed Central PMCID: PMC3106235.
19. Murdoch DR, O'Brien KL, Driscoll AJ, Karron RA, Bhat N, Pneumonia Methods Working G, et al. Laboratory methods for determining pneumonia etiology in children. Clin Infect Dis. 2012;54 Suppl 2:S146-52. Epub 2012/04/17. doi: 10.1093/cid/cir1073. PubMed PMID: 22403229.
20. Jartti T, Lehtinen P, Vuorinen T, Koskenvuo M, Ruuskanen O. Persistence of rhinovirus and enterovirus RNA after acute respiratory illness in children. Journal of medical virology. 2004;72(4):695-9. doi: 10.1002/jmv.20027. PubMed PMID: 14981776.
21. Hammitt LL, Murdoch DR, Scott JA, Driscoll A, Karron RA, Levine OS, et al. Specimen collection for the diagnosis of pediatric pneumonia. Clin Infect Dis. 2012;54 Suppl 2:S132-9. Epub 2012/04/17. doi: 10.1093/cid/cir1068. PubMed PMID: 22403227; PubMed Central PMCID: PMC3693496.
22. Bartlett JG. Diagnostic tests for agents of community-acquired pneumonia. Clin Infect Dis. 2011;52 Suppl 4:S296-304. Epub 2011/04/06. doi: 10.1093/cid/cir045. PubMed PMID: 21460288.
23. Bartlett JG, Finegold SM. Bacteriology of expectorated sputum with quantitative culture and wash technique compared to transtracheal aspirates. The American review of respiratory disease. 1978;117(6):1019-27. PubMed PMID: 352206.
24. Bartlett RC. Medical microbiology: quality, cost and clinical relevance. New York, N.Y.: John Wiley & Sons, Inc.; 1974.
25. Pollock HM, Hawkins EL, Bonner JR, Sparkman T, Bass JB, Jr. Diagnosis of bacterial pulmonary infections with quantitative protected catheter cultures obtained during bronchoscopy. J Clin Microbiol. 1983;17(2):255-9. PubMed PMID: 6339545; PubMed Central PMCID: PMC272617.
26. Avni T, Mansur N, Leibovici L, Paul M. PCR using blood for diagnosis of invasive pneumococcal disease: systematic review and meta-analysis. J Clin Microbiol.

2010;48(2):489-96. doi: 10.1128/JCM.01636-09. PubMed PMID: 20007385; PubMed Central PMCID: PMC2815606.
27. Kee C, Palladino S, Kay I, Pryce TM, Murray R, Rello J, et al. Feasibility of real-time polymerase chain reaction in whole blood to identify Streptococcus pneumoniae in patients with community-acquired pneumonia. Diagn Microbiol Infect Dis. 2008;61(1):72-5. doi: 10.1016/j.diagmicrobio.2007.12.011. PubMed PMID: 18221851.
28. Peters RP, de Boer RF, Schuurman T, Gierveld S, Kooistra-Smid M, van Agtmael MA, et al. Streptococcus pneumoniae DNA load in blood as a marker of infection in patients with community-acquired pneumonia. J Clin Microbiol. 2009;47(10):3308-12. doi: 10.1128/JCM.01071-09. PubMed PMID: 19675218; PubMed Central PMCID: PMC2756912.
29. Genne D, Siegrist HH, Lienhard R. Enhancing the etiologic diagnosis of community-acquired pneumonia in adults using the urinary antigen assay (Binax NOW). International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases. 2006;10(2):124-8. doi: 10.1016/j.ijid.2005.03.006. PubMed PMID: 16290014.
30. Gutierrez F, Masia M, Rodriguez JC, Ayelo A, Soldan B, Cebrian L, et al. Evaluation of the immunochromatographic Binax NOW assay for detection of Streptococcus pneumoniae urinary antigen in a prospective study of community-acquired pneumonia in Spain. Clin Infect Dis. 2003;36(3):286-92. doi: 10.1086/345852. PubMed PMID: 12539069.
31. Murdoch DR, Laing RT, Mills GD, Karalus NC, Town GI, Mirrett S, et al. Evaluation of a rapid immunochromatographic test for detection of Streptococcus pneumoniae antigen in urine samples from adults with community-acquired pneumonia. J Clin Microbiol. 2001;39(10):3495-8. doi: 10.1128/JCM.39.10.3495-3498.2001. PubMed PMID: 11574562; PubMed Central PMCID: PMC88378.
32. Roson B, Fernandez-Sabe N, Carratala J, Verdaguer R, Dorca J, Manresa F, et al. Contribution of a urinary antigen assay (Binax NOW) to the early diagnosis of pneumococcal pneumonia. Clin Infect Dis. 2004;38(2):222-6. doi: 10.1086/380639. PubMed PMID: 14699454.
33. Smith MD, Sheppard CL, Hogan A, Harrison TG, Dance DA, Derrington P, et al. Diagnosis of Streptococcus pneumoniae infections in adults with bacteremia and community-acquired pneumonia: clinical comparison of pneumococcal PCR and urinary antigen detection. J Clin Microbiol. 2009;47(4):1046-9. doi: 10.1128/JCM.01480-08. PubMed PMID: 19225103; PubMed Central PMCID: PMC2668348.
34. Dominguez J, Gali N, Blanco S, Pedroso P, Prat C, Matas L, et al. Detection of Streptococcus pneumoniae antigen by a rapid immunochromatographic assay in urine samples. Chest. 2001;119(1):243-9. PubMed PMID: 11157611.
35. Murdoch DR, Laing RT, Cook JM. The NOW S. pneumoniae urinary antigen test positivity rate 6 weeks after pneumonia onset and among patients with COPD. Clin Infect Dis. 2003;37(1):153-4. doi: 10.1086/375610. PubMed PMID: 12830428.
36. Dowell SF, Garman RL, Liu G, Levine OS, Yang YH. Evaluation of Binax NOW, an assay for the detection of pneumococcal antigen in urine samples, performed among pediatric patients. Clin Infect Dis. 2001;32(5):824-5. doi: 10.1086/319205. PubMed PMID: 11229853.

37. Lahti E, Peltola V, Waris M, Virkki R, Rantakokko-Jalava K, Jalava J, et al. Induced sputum in the diagnosis of childhood community-acquired pneumonia. Thorax. 2009;64(3):252-7. doi: 10.1136/thx.2008.099051. PubMed PMID: 19052043.
38. Musher DM, Montoya R, Wanahita A. Diagnostic value of microscopic examination of Gram-stained sputum and sputum cultures in patients with bacteremic pneumococcal pneumonia. Clin Infect Dis. 2004;39(2):165-9. doi: 10.1086/421497. PubMed PMID: 15307023.
39. Andreo F, Dominguez J, Ruiz-Manzano J, Prat C, Blanco S, Lores L, et al. Usefulness of pneumococcal antigen detection in pleural fluid samples by immunochromatographic assay for diagnosis of pneumococcal pneumonia. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2006;12(7):682-4. doi: 10.1111/j.1469-0691.2006.01484.x. PubMed PMID: 16774569.
40. Falguera M, Lopez A, Nogues A, Porcel JM, Rubio-Caballero M. Evaluation of the polymerase chain reaction method for detection of Streptococcus pneumoniae DNA in pleural fluid samples. Chest. 2002;122(6):2212-6. PubMed PMID: 12475865.
41. Lahti E, Mertsola J, Kontiokari T, Eerola E, Ruuskanen O, Jalava J. Pneumolysin polymerase chain reaction for diagnosis of pneumococcal pneumonia and empyema in children. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology. 2006;25(12):783-9. doi: 10.1007/s10096-006-0225-9. PubMed PMID: 17089094.
42. Casado Flores J, Nieto Moro M, Berron S, Jimenez R, Casal J. Usefulness of pneumococcal antigen detection in pleural effusion for the rapid diagnosis of infection by Streptococcus pneumoniae. European journal of pediatrics. 2010;169(5):581-4. doi: 10.1007/s00431-009-1077-y. PubMed PMID: 19806363.
43. Blaschke AJ, Heyrend C, Byington CL, Obando I, Vazquez-Barba I, Doby EH, et al. Molecular analysis improves pathogen identification and epidemiologic study of pediatric parapneumonic empyema. Pediatr Infect Dis J. 2011;30(4):289-94. doi: 10.1097/INF.0b013e3182002d14. PubMed PMID: 21057372; PubMed Central PMCID: PMC3053443.
44. Azzari C, Moriondo M, Indolfi G, Cortimiglia M, Canessa C, Becciolini L, et al. Realtime PCR is more sensitive than multiplex PCR for diagnosis and serotyping in children with culture negative pneumococcal invasive disease. PLoS One. 2010;5(2):e9282. doi: 10.1371/journal.pone.0009282. PubMed PMID: 20174571; PubMed Central PMCID: PMC2824814.
45. Carvalho Mda G, Tondella ML, McCaustland K, Weidlich L, McGee L, Mayer LW, et al. Evaluation and improvement of real-time PCR assays targeting lytA, ply, and psaA genes for detection of pneumococcal DNA. J Clin Microbiol. 2007;45(8):2460-6. doi: 10.1128/JCM.02498-06. PubMed PMID: 17537936; PubMed Central PMCID: PMC1951257.
46. Azzari C, Cortimiglia M, Moriondo M, Canessa C, Lippi F, Ghiori F, et al. Pneumococcal DNA is not detectable in the blood of healthy carrier children by real-time PCR targeting the lytA gene. J Med Microbiol. 2011;60(Pt 6):710-4. doi: 10.1099/jmm.0.028357-0. PubMed PMID: 21349984; PubMed Central PMCID: PMC3167920.

47. van Haeften R, Palladino S, Kay I, Keil T, Heath C, Waterer GW. A quantitative LightCycler PCR to detect Streptococcus pneumoniae in blood and CSF. Diagn Microbiol Infect Dis. 2003;47(2):407-14. PubMed PMID: 14522514.
48. Zhang Y, Isaacman DJ, Wadowsky RM, Rydquist-White J, Post JC, Ehrlich GD. Detection of Streptococcus pneumoniae in whole blood by PCR. J Clin Microbiol. 1995;33(3):596-601. PubMed PMID: 7751363; PubMed Central PMCID: PMC227996.
49. Rudolph KM, Parkinson AJ, Black CM, Mayer LW. Evaluation of polymerase chain reaction for diagnosis of pneumococcal pneumonia. J Clin Microbiol. 1993;31(10):2661-6. PubMed PMID: 8253962; PubMed Central PMCID: PMC265966.
50. Rello J, Lisboa T, Lujan M, Gallego M, Kee C, Kay I, et al. Severity of pneumococcal pneumonia associated with genomic bacterial load. Chest. 2009;136(3):832-40. doi: 10.1378/chest.09-0258. PubMed PMID: 19433527.
51. Werno AM, Anderson TP, Murdoch DR. Association between pneumococcal load and disease severity in adults with pneumonia. J Med Microbiol. 2012;61(Pt 8):1129-35. doi: 10.1099/jmm.0.044107-0. PubMed PMID: 22499777.
52. Anh DD, Kilgore PE, Slack MP, Nyambat B, Tho le H, Yoshida LM, et al. Surveillance of pneumococcal-associated disease among hospitalized children in Khanh Hoa Province, Vietnam. Clin Infect Dis. 2009;48 Suppl 2:S57-64. doi: 10.1086/596483. PubMed PMID: 19191620.
53. Rajam G, Anderton JM, Carlone GM, Sampson JS, Ades EW. Pneumococcal surface adhesin A (PsaA): a review. Critical reviews in microbiology. 2008;34(3-4):131-42. doi: 10.1080/10408410802275352. PubMed PMID: 18728990.
54. Scott JA, Obiero J, Hall AJ, Marsh K. Validation of immunoglobulin G enzyme-linked immunosorbent assay for antibodies to pneumococcal surface adhesin A in the diagnosis of pneumococcal pneumonia among adults in Kenya. J Infect Dis. 2002;186(2):220-6. doi: 10.1086/341205. PubMed PMID: 12134258.
55. Rapola S, Jantti V, Haikala R, Syrjanen R, Carlone GM, Sampson JS, et al. Natural development of antibodies to pneumococcal surface protein A, pneumococcal surface adhesin A, and pneumolysin in relation to pneumococcal carriage and acute otitis media. J Infect Dis. 2000;182(4):1146-52. doi: 10.1086/315822. PubMed PMID: 10979911.
56. Scott JA, Mlacha Z, Nyiro J, Njenga S, Lewa P, Obiero J, et al. Diagnosis of invasive pneumococcal disease among children in Kenya with enzyme-linked immunosorbent assay for immunoglobulin G antibodies to pneumococcal surface adhesin A. Clinical and diagnostic laboratory immunology. 2005;12(10):1195-201. doi: 10.1128/CDLI.12.10.1195-1201.2005. PubMed PMID: 16210483; PubMed Central PMCID: PMC1247825.
57. Tharpe JA, Russell H, Leinonen M, Plikaytis BD, Breiman RF, Carlone GM, et al. Comparison of a pneumococcal common protein (PsaA) antibody ELISA and a PsaA immune complex ELISA for detection of pneumococcal serum antibody. Pathobiology : journal of immunopathology, molecular and cellular biology. 1998;66(2):77-83. PubMed PMID: 9645631.
58. Korppi M, Leinonen M, Ruuskanen O. Pneumococcal serology in children's respiratory infections. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology. 2008;27(3):167-75. doi: 10.1007/s10096-007-0436-8. PubMed PMID: 18087733.

59. Butler JC, Bosshardt SC, Phelan M, Moroney SM, Tondella ML, Farley MM, et al. Classical and latent class analysis evaluation of sputum polymerase chain reaction and urine antigen testing for diagnosis of pneumococcal pneumonia in adults. J Infect Dis. 2003;187(9):1416-23. Epub 2003/04/30. doi: 10.1086/374623. PubMed PMID: 12717623.
60. Sinclair A, Xie X, Teltscher M, Dendukuri N. Systematic review and meta-analysis of a urine-based pneumococcal antigen test for diagnosis of community-acquired pneumonia caused by Streptococcus pneumoniae. J Clin Microbiol. 2013;51(7):2303-10. Epub 2013/05/17. doi: 10.1128/JCM.00137-13. PubMed PMID: 23678060; PubMed Central PMCID: PMC3697702.
61. Said MA, Johnson HL, Nonyane BA, Deloria-Knoll M, O'Brien KL, Team AAPBS, et al. Estimating the burden of pneumococcal pneumonia among adults: a systematic review and meta-analysis of diagnostic techniques. PLoS One. 2013;8(4):e60273. Epub 2013/04/09. doi: 10.1371/journal.pone.0060273. PubMed PMID: 23565216; PubMed Central PMCID: PMC3615022.
62. Niederman MS. Biological markers to determine eligibility in trials for community-acquired pneumonia: a focus on procalcitonin. Clin Infect Dis. 2008;47 Suppl 3:S127-32. doi: 10.1086/591393. PubMed PMID: 18986278.
63. Simon L, Gauvin F, Amre DK, Saint-Louis P, Lacroix J. Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. Clin Infect Dis. 2004;39(2):206-17. doi: 10.1086/421997. PubMed PMID: 15307030.
64. Muller B, Christ-Crain M, Nylen ES, Snider R, Becker KL. Limits to the use of the procalcitonin level as a diagnostic marker. Clin Infect Dis. 2004;39(12):1867-8. doi: 10.1086/426148. PubMed PMID: 15578415.
65. Harper SA, Bradley JS, Englund JA, File TM, Gravenstein S, Hayden FG, et al. Seasonal influenza in adults and children--diagnosis, treatment, chemoprophylaxis, and institutional outbreak management: clinical practice guidelines of the Infectious Diseases Society of America. Clin Infect Dis. 2009;48(8):1003-32. doi: 10.1086/598513. PubMed PMID: 19281331.
66. Templeton KE, Scheltinga SA, Beersma MF, Kroes AC, Claas EC. Rapid and sensitive method using multiplex real-time PCR for diagnosis of infections by influenza a and influenza B viruses, respiratory syncytial virus, and parainfluenza viruses 1, 2, 3, and 4. J Clin Microbiol. 2004;42(4):1564-9. PubMed PMID: 15071005; PubMed Central PMCID: PMC387552.
67. van Elden LJ, van Kraaij MG, Nijhuis M, Hendriksen KA, Dekker AW, Rozenberg-Arska M, et al. Polymerase chain reaction is more sensitive than viral culture and antigen testing for the detection of respiratory viruses in adults with hematological cancer and pneumonia. Clin Infect Dis. 2002;34(2):177-83. doi: 10.1086/338238. PubMed PMID: 11740705.
68. Falsey AR, Formica MA, Walsh EE. Diagnosis of respiratory syncytial virus infection: comparison of reverse transcription-PCR to viral culture and serology in adults with respiratory illness. J Clin Microbiol. 2002;40(3):817-20. PubMed PMID: 11880399; PubMed Central PMCID: PMC120281.
69. Gavin PJ, Thomson Jr RB, Jr. Review of Rapid Diagnostic Tests for Influenza. Clinical and Applied Immunology Reviews. 2003;4:151-72.

70. Cazacu AC, Chung SE, Greer J, Demmler GJ. Comparison of the directigen flu A+B membrane enzyme immunoassay with viral culture for rapid detection of influenza A and B viruses in respiratory specimens. J Clin Microbiol. 2004;42(8):3707-10. doi: 10.1128/JCM.42.8.3707-3710.2004. PubMed PMID: 15297520; PubMed Central PMCID: PMC497654.
71. Chan MC, Lee N, Ngai KL, Leung TF, Chan PK. Clinical and virologic factors associated with reduced sensitivity of rapid influenza diagnostic tests in hospitalized elderly patients and young children. J Clin Microbiol. 2014;52(2):497-501. doi: 10.1128/JCM.02316-13. PubMed PMID: 24478479; PubMed Central PMCID: PMC3911332.
72. Ruest A, Michaud S, Deslandes S, Frost EH. Comparison of the Directigen flu A+B test, the QuickVue influenza test, and clinical case definition to viral culture and reverse transcription-PCR for rapid diagnosis of influenza virus infection. J Clin Microbiol. 2003;41(8):3487-93. PubMed PMID: 12904343; PubMed Central PMCID: PMC179849.
73. Uyeki TM, Prasad R, Vukotich C, Stebbins S, Rinaldo CR, Ferng YH, et al. Low sensitivity of rapid diagnostic test for influenza. Clin Infect Dis. 2009;48(9):e89-92. doi: 10.1086/597828. PubMed PMID: 19323628.
74. Ruuskanen O, Lahti E, Jennings LC, Murdoch DR. Viral pneumonia. Lancet. 2011;377(9773):1264-75. doi: 10.1016/S0140-6736(10)61459-6. PubMed PMID: 21435708.
75. Oosterheert JJ, van Loon AM, Schuurman R, Hoepelman AI, Hak E, Thijsen S, et al. Impact of rapid detection of viral and atypical bacterial pathogens by real-time polymerase chain reaction for patients with lower respiratory tract infection. Clin Infect Dis. 2005;41(10):1438-44. doi: 10.1086/497134. PubMed PMID: 16231254.
76. Templeton KE, Scheltinga SA, van den Eeden WC, Graffelman AW, van den Broek PJ, Claas EC. Improved diagnosis of the etiology of community-acquired pneumonia with real-time polymerase chain reaction. Clin Infect Dis. 2005;41(3):345-51. doi: 10.1086/431588. PubMed PMID: 16007532.
77. Weinberg GA, Erdman DD, Edwards KM, Hall CB, Walker FJ, Griffin MR, et al. Superiority of reverse-transcription polymerase chain reaction to conventional viral culture in the diagnosis of acute respiratory tract infections in children. J Infect Dis. 2004;189(4):706-10. doi: 10.1086/381456. PubMed PMID: 14767825.
78. Iwasenko JM, Cretikos M, Paterson DL, Gibb R, Webb SA, Smith DW, et al. Enhanced diagnosis of pandemic (H1N1) 2009 influenza infection using molecular and serological testing in intensive care unit patients with suspected influenza. Clin Infect Dis. 2010;51(1):70-2. doi: 10.1086/653610. PubMed PMID: 20482372.
79. Suess T, Buchholz U, Dupke S, Grunow R, an der Heiden M, Heider A, et al. Shedding and transmission of novel influenza virus A/H1N1 infection in households--Germany, 2009. Am J Epidemiol. 2010;171(11):1157-64. doi: 10.1093/aje/kwq071. PubMed PMID: 20439308.
80. Feikin DR, Njenga MK, Bigogo G, Aura B, Gikunju S, Balish A, et al. Additional diagnostic yield of adding serology to PCR in diagnosing viral acute respiratory infections in Kenyan patients 5 years of age and older. Clin Vaccine Immunol. 2013;20(1):113-4. doi: 10.1128/CVI.00325-12. PubMed PMID: 23114699; PubMed Central PMCID: PMC3535781.

81. Sawatwong P, Chittaganpitch M, Hall H, Peruski LF, Xu X, Baggett HC, et al. Serology as an adjunct to polymerase chain reaction assays for surveillance of acute respiratory virus infections. Clin Infect Dis. 2012;54(3):445-6. doi: 10.1093/cid/cir710. PubMed PMID: 22057703.
82. Veguilla V, Hancock K, Schiffer J, Gargiullo P, Lu X, Aranio D, et al. Sensitivity and specificity of serologic assays for detection of human infection with 2009 pandemic H1N1 virus in U.S. populations. J Clin Microbiol. 2011;49(6):2210-5. doi: 10.1128/JCM.00229-11. PubMed PMID: 21471339; PubMed Central PMCID: PMC3122722.
83. Papenburg J, Baz M, Hamelin ME, Rheaume C, Carbonneau J, Ouakki M, et al. Evaluation of serological diagnostic methods for the 2009 pandemic influenza A (H1N1) virus. Clin Vaccine Immunol. 2011;18(3):520-2. doi: 10.1128/CVI.00449-10. PubMed PMID: 21228145; PubMed Central PMCID: PMC3067392.
84. Ieven MM, Loens K. Should Serology be Abolished in Favor of PCR for the Diagnosis of Mycoplasma pneumoniae Infections? Current Pediatric Review. 2013;9:304-13.
85. She RC, Thurber A, Hymas WC, Stevenson J, Langer J, Litwin CM, et al. Limited utility of culture for Mycoplasma pneumoniae and Chlamydophila pneumoniae for diagnosis of respiratory tract infections. J Clin Microbiol. 2010;48(9):3380-2. doi: 10.1128/JCM.00321-10. PubMed PMID: 20610673; PubMed Central PMCID: PMC2937689.
86. Helbig JH, Uldum SA, Luck PC, Harrison TG. Detection of Legionella pneumophila antigen in urine samples by the BinaxNOW immunochromatographic assay and comparison with both Binax Legionella Urinary Enzyme Immunoassay (EIA) and Biotest Legionella Urin Antigen EIA. J Med Microbiol. 2001;50(6):509-16. PubMed PMID: 11393288.
87. Nilsson AC, Bjorkman P, Persson K. Polymerase chain reaction is superior to serology for the diagnosis of acute Mycoplasma pneumoniae infection and reveals a high rate of persistent infection. BMC microbiology. 2008;8:93. doi: 10.1186/1471-2180-8-93. PubMed PMID: 18547431; PubMed Central PMCID: PMC2446399.
88. Templeton KE, Scheltinga SA, Graffelman AW, Van Schie JM, Crielaard JW, Sillekens P, et al. Comparison and evaluation of real-time PCR, real-time nucleic acid sequence-based amplification, conventional PCR, and serology for diagnosis of Mycoplasma pneumoniae. J Clin Microbiol. 2003;41(9):4366-71. PubMed PMID: 12958270; PubMed Central PMCID: PMC193789.
89. Thurman KA, Walter ND, Schwartz SB, Mitchell SL, Dillon MT, Baughman AL, et al. Comparison of laboratory diagnostic procedures for detection of Mycoplasma pneumoniae in community outbreaks. Clin Infect Dis. 2009;48(9):1244-9. doi: 10.1086/597775. PubMed PMID: 19331586.
90. Dowell SF, Peeling RW, Boman J, Carlone GM, Fields BS, Guarner J, et al. Standardizing Chlamydia pneumoniae assays: recommendations from the Centers for Disease Control and Prevention (USA) and the Laboratory Centre for Disease Control (Canada). Clin Infect Dis. 2001;33(4):492-503. doi: 10.1086/322632. PubMed PMID: 11462186.
91. Miyashita N, Akaike H, Teranishi H, Kawai Y, Ouchi K, Kato T, et al. Chlamydophila pneumoniae serology: cross-reaction with Mycoplasma pneumoniae

infection. Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy. 2013;19(2):256-60. doi: 10.1007/s10156-012-0494-4. PubMed PMID: 23065148.
92. Beersma MF, Dirven K, van Dam AP, Templeton KE, Claas EC, Goossens H. Evaluation of 12 commercial tests and the complement fixation test for Mycoplasma pneumoniae-specific immunoglobulin G (IgG) and IgM antibodies, with PCR used as the "gold standard". J Clin Microbiol. 2005;43(5):2277-85. doi: 10.1128/JCM.43.5.2277-2285.2005. PubMed PMID: 15872256; PubMed Central PMCID: PMC1153783.
93. Klugman KP, Madhi SA, Albrich WC. Novel approaches to the identification of Streptococcus pneumoniae as the cause of community-acquired pneumonia. Clin Infect Dis. 2008;47 Suppl 3:S202-6. Epub 2008/11/18. doi: 10.1086/591405. PubMed PMID: 18986290.
94. McGeer AJ. Diagnostic testing or empirical therapy for patients hospitalized with suspected influenza: what to do? Clin Infect Dis. 2009;48 Suppl 1:S14-9. doi: 10.1086/591852. PubMed PMID: 19067610.
95. Grijalva CG, Nuorti JP, Arbogast PG, Martin SW, Edwards KM, Griffin MR. Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis. Lancet. 2007;369(9568):1179-86. Epub 2007/04/10. doi: 10.1016/S0140-6736(07)60564-9. PubMed PMID: 17416262.
96. Hamano-Hasegawa K, Morozumi M, Nakayama E, Chiba N, Murayama SY, Takayanagi R, et al. Comprehensive detection of causative pathogens using real-time PCR to diagnose pediatric community-acquired pneumonia. Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy. 2008;14(6):424-32. Epub 2008/12/18. doi: 10.1007/s10156-008-0648-6. PubMed PMID: 19089556.
97. Heiskanen-Kosma T, Korppi M, Jokinen C, Kurki S, Heiskanen L, Juvonen H, et al. Etiology of childhood pneumonia: serologic results of a prospective, population-based study. Pediatr Infect Dis J. 1998;17(11):986-91. Epub 1998/12/16. PubMed PMID: 9849979.
98. Juven T, Mertsola J, Waris M, Leinonen M, Meurman O, Roivainen M, et al. Etiology of community-acquired pneumonia in 254 hospitalized children. Pediatr Infect Dis J. 2000;19(4):293-8. Epub 2000/04/27. PubMed PMID: 10783017.
99. MacNeil JR, Cohn AC, Farley M, Mair R, Baumbach J, Bennett N, et al. Current epidemiology and trends in invasive Haemophilus influenzae disease--United States, 1989-2008. Clin Infect Dis. 2011;53(12):1230-6. Epub 2011/11/15. doi: 10.1093/cid/cir735. PubMed PMID: 22080119.
100. Techasaensiri C, Messina AF, Katz K, Ahmad N, Huang R, McCracken GH, Jr. Epidemiology and evolution of invasive pneumococcal disease caused by multidrug resistant serotypes of 19A in the 8 years after implementation of pneumococcal conjugate vaccine immunization in Dallas, Texas. Pediatr Infect Dis J. 2010;29(4):294-300. Epub 2009/12/02. doi: 10.1097/INF.0b013e3181c2a229. PubMed PMID: 19949357.
101. Walter ND, Taylor TH, Shay DK, Thompson WW, Brammer L, Dowell SF, et al. Influenza circulation and the burden of invasive pneumococcal pneumonia during a non-pandemic period in the United States. Clin Infect Dis. 2010;50(2):175-83. Epub 2009/12/18. doi: 10.1086/649208. PubMed PMID: 20014948.

102. McCullers JA. Insights into the interaction between influenza virus and pneumococcus. Clin Microbiol Rev. 2006;19(3):571-82. Epub 2006/07/19. doi: 10.1128/CMR.00058-05. PubMed PMID: 16847087; PubMed Central PMCID: PMC1539103.
103. Ruiz-Gonzalez A, Falguera M, Nogues A, Rubio-Caballero M. Is Streptococcus pneumoniae the leading cause of pneumonia of unknown etiology? A microbiologic study of lung aspirates in consecutive patients with community-acquired pneumonia. The American journal of medicine. 1999;106(4):385-90. PubMed PMID: 10225239.
104. Brown JS. Community-acquired pneumonia. Clinical medicine. 2012;12(6):538-43. Epub 2013/01/25. PubMed PMID: 23342408.
105. Dao CN, Kamimoto L, Nowell M, Reingold A, Gershman K, Meek J, et al. Adult hospitalizations for laboratory-positive influenza during the 2005-2006 through 2007-2008 seasons in the United States. J Infect Dis. 2010;202(6):881-8. doi: 10.1086/655904. PubMed PMID: 20677944.
106. Dawood FS, Fiore A, Kamimoto L, Bramley A, Reingold A, Gershman K, et al. Burden of seasonal influenza hospitalization in children, United States, 2003 to 2008. J Pediatr. 2010;157(5):808-14. doi: 10.1016/j.jpeds.2010.05.012. PubMed PMID: 20580018.
107. Jain S, Benoit SR, Skarbinski J, Bramley AM, Finelli L, Pandemic Influenza AVHIT. Influenza-associated pneumonia among hospitalized patients with 2009 pandemic influenza A (H1N1) virus--United States, 2009. Clin Infect Dis. 2012;54(9):1221-9. doi: 10.1093/cid/cis197. PubMed PMID: 22437239.
108. Schrag SJ, Shay DK, Gershman K, Thomas A, Craig AS, Schaffner W, et al. Multistate surveillance for laboratory-confirmed, influenza-associated hospitalizations in children: 2003-2004. Pediatr Infect Dis J. 2006;25(5):395-400. doi: 10.1097/01.inf.0000214988.81379.71. PubMed PMID: 16645501.
109. Jain S, Kamimoto L, Bramley AM, Schmitz AM, Benoit SR, Louie J, et al. Hospitalized patients with 2009 H1N1 influenza in the United States, April-June 2009. N Engl J Med. 2009;361(20):1935-44. doi: 10.1056/NEJMoa0906695. PubMed PMID: 19815859.
110. Lee N, Chan PK, Lui GC, Wong BC, Sin WW, Choi KW, et al. Complications and outcomes of pandemic 2009 Influenza A (H1N1) virus infection in hospitalized adults: how do they differ from those in seasonal influenza? J Infect Dis. 2011;203(12):1739-47. doi: 10.1093/infdis/jir187. PubMed PMID: 21606532.
111. Louria DB, Blumenfeld HL, Ellis JT, Kilbourne ED, Rogers DE. Studies on influenza in the pandemic of 1957-1958. II. Pulmonary complications of influenza. J Clin Invest. 1959;38(1 Part 2):213-65. doi: 10.1172/JCI103791. PubMed PMID: 13620784; PubMed Central PMCID: PMC444127.
112. Morens DM, Taubenberger JK, Fauci AS. Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. J Infect Dis. 2008;198(7):962-70. doi: 10.1086/591708. PubMed PMID: 18710327; PubMed Central PMCID: PMC2599911.
113. Reed C, Chaves SS, Perez A, D'Mello T, Daily Kirley P, Aragon D, et al. Complications among adults hospitalized with influenza: a comparison of seasonal influenza and the 2009 H1N1 pandemic. Clin Infect Dis. 2014;59(2):166-74. doi: 10.1093/cid/ciu285. PubMed PMID: 24785230.

114. Thompson WW, Shay DK, Weintraub E, Brammer L, Bridges CB, Cox NJ, et al. Influenza-associated hospitalizations in the United States. JAMA : the journal of the American Medical Association. 2004;292(11):1333-40. doi: 10.1001/jama.292.11.1333. PubMed PMID: 15367555.
115. Jain S, Self W, Wunderink RG, al. e. Incidence and Etiology of Community-Acquired Pneumonia Hospitalizations among U.S. Adults. In Press.
116. Jain S, Williams DJ, Arnold SR, Ampofo K, al. e. Incidence and Etiology of Community-Acquired Pneumonia Requiring Hospitalization among U.S. Children. New England Journal of Medicine. In press.
117. Dawood FS, Chaves SS, Perez A, Reingold A, Meek J, Farley MM, et al. Complications and associated bacterial coinfections among children hospitalized with seasonal or pandemic influenza, United States, 2003-2010. J Infect Dis. 2014;209(5):686-94. doi: 10.1093/infdis/jit473. PubMed PMID: 23986545.
118. Chaves SS, Aragon D, Bennett N, Cooper T, D'Mello T, Farley M, et al. Patients hospitalized with laboratory-confirmed influenza during the 2010-2011 influenza season: exploring disease severity by virus type and subtype. J Infect Dis. 2013;208(8):1305-14. doi: 10.1093/infdis/jit316. PubMed PMID: 23863950.
119. Louie JK, Acosta M, Winter K, Jean C, Gavali S, Schechter R, et al. Factors associated with death or hospitalization due to pandemic 2009 influenza A(H1N1) infection in California. JAMA : the journal of the American Medical Association. 2009;302(17):1896-902. doi: 10.1001/jama.2009.1583. PubMed PMID: 19887665.
120. Skarbinski J, Jain S, Bramley A, Lee EJ, Huang J, Kirschke D, et al. Hospitalized patients with 2009 pandemic influenza A (H1N1) virus infection in the United States--September-October 2009. Clin Infect Dis. 2011;52 Suppl 1:S50-9. doi: 10.1093/cid/ciq021. PubMed PMID: 21342900.
121. Dawood FS, Fiore A, Kamimoto L, Nowell M, Reingold A, Gershman K, et al. Influenza-associated pneumonia in children hospitalized with laboratory-confirmed influenza, 2003-2008. Pediatr Infect Dis J. 2010;29(7):585-90. PubMed PMID: 20589966.
122. Fiore A, Fry A, Shay D, Gubareva L, Bresee J, Uyeki T. Antivital Agents for the Treatment and Chemoprophylaxis of Influenza: Recommendations of the Advisory Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly Report. 2011;60(1):1-18.
123. Lee N, Chan PK, Hui DS, Rainer TH, Wong E, Choi KW, et al. Viral loads and duration of viral shedding in adult patients hospitalized with influenza. J Infect Dis. 2009;200(4):492-500. doi: 10.1086/600383. PubMed PMID: 19591575.
124. Zarychanski R, Stuart TL, Kumar A, Doucette S, Elliott L, Kettner J, et al. Correlates of severe disease in patients with 2009 pandemic influenza (H1N1) virus infection. CMAJ. 2010;182(3):257-64. doi: 10.1503/cmaj.091884. PubMed PMID: 20093297; PubMed Central PMCID: PMC2826467.
125. Gupta V, Dawood FS, Rai SK, Broor S, Wigh R, Mishra AC, et al. Validity of clinical case definitions for influenza surveillance among hospitalized patients: results from a rural community in North India. Influenza and other respiratory viruses. 2013;7(3):321-9. doi: 10.1111/j.1750-2659.2012.00401.x. PubMed PMID: 22804843.
126. Bjarnason A, Thorleifsdottir G, Love A, Gudnason JF, Asgeirsson H, Hallgrimsson KL, et al. Severity of influenza A 2009 (H1N1) pneumonia is

underestimated by routine prediction rules. Results from a prospective, population-based study. PLoS One. 2012;7(10):e46816. doi: 10.1371/journal.pone.0046816. PubMed PMID: 23071646; PubMed Central PMCID: PMC3469650.
127. Tuyisenge L, Ndimubanzi CP, Ndayisaba G, Muganga N, Menten J, Boelaert M, et al. Evaluation of latent class analysis and decision thresholds to guide the diagnosis of pediatric tuberculosis in a Rwandan reference hospital. Pediatr Infect Dis J. 2010;29(2):e11-8. doi: 10.1097/INF.0b013e3181c61ddb. PubMed PMID: 19935116.
128. Trenz RC, Scherer M, Duncan A, Harrell PT, Moleko AG, Latimer WW. Latent class analysis of polysubstance use, sexual risk behaviors, and infectious disease among South African drug users. Drug and alcohol dependence. 2013;132(3):441-8. doi: 10.1016/j.drugalcdep.2013.03.004. PubMed PMID: 23562370.
129. Muma JB, Toft N, Oloya J, Lund A, Nielsen K, Samui K, et al. Evaluation of three serological tests for brucellosis in naturally infected cattle using latent class analysis. Veterinary microbiology. 2007;125(1-2):187-92. doi: 10.1016/j.vetmic.2007.05.012. PubMed PMID: 17590540.
130. Jokinen J, Scott JA. Estimating the proportion of pneumonia attributable to pneumococcus in Kenyan adults: latent class analysis. Epidemiology. 2010;21(5):719-25. doi: 10.1097/EDE.0b013e3181e4c4d5. PubMed PMID: 20562627; PubMed Central PMCID: PMC2923075.
131. Hartnack S, Budke CM, Craig PS, Jiamin Q, Boufana B, Campos-Ponce M, et al. Latent-class methods to evaluate diagnostics tests for Echinococcus infections in dogs. PLoS neglected tropical diseases. 2013;7(2):e2068. doi: 10.1371/journal.pntd.0002068. PubMed PMID: 23459420; PubMed Central PMCID: PMC3573084.
132. Magidson J, Vermunt J. Latent Class Models Statistical Innovations; http://www.statisticalinnovations.com/technicalsupport/lcmodels2.pdf [cited 2014 February 16]. Available from: http://www.statisticalinnovations.com/technicalsupport/lcmodels2.pdf.
133. McCutcheon AL. Latent class analysis. Quantitative applications in the social sciences. 64: SAGE Publications, Inc; 1987. p. 96.
134. Clogg CC. Latent Class Models. In: Arminger G CC, Sobel ME, editor. Handbook of Statistical Modeling for the Social and Behavioral Sciences. New York: Plenum Press; 1995. p. 592.
135. Muthen B. Beyond SEM: General latent variable modeling. Behaviormetrika. 2002;29(1):81-119.
136. Thompson D. Performing Latent Class Analysis Using the CATMOD Procedure. SAS Users Group International; March 26-29, 2006; San Francisco, CA2006.
137. Bartholomew D, Knott M, I M. Latent Variable Models and Factor Analysis: A Unified Approach. 3rd ed: Wiley; 2011.
138. Vermunt JK MJ. Latent class analysis. In: Lewis-Beck M, Bryman A, Liao T, editors. Encyclopedia of social science research methods. Thousand Oaks, CA: SAGE Publications, Inc. ; 2004. p. 550-4.
139. Vermunt J, Magdison J. Latent Class Analysis. Encyclopedia of Social Science Research Methods: Sage Publications; 2003.
140. Christ-Crain M, Jaccard-Stolz D, Bingisser R, Gencay MM, Huber PR, Tamm M, et al. Effect of procalcitonin-guided treatment on antibiotic use and outcome in lower respiratory tract infections: cluster-randomised, single-blinded intervention trial. Lancet.

2004;363(9409):600-7. doi: 10.1016/S0140-6736(04)15591-8. PubMed PMID: 14987884.
141. Christ-Crain M, Muller B. Biomarkers in respiratory tract infections: diagnostic guides to antibiotic prescription, prognostic markers and mediators. The European respiratory journal. 2007;30(3):556-73. doi: 10.1183/09031936.00166106. PubMed PMID: 17766633.
142. Gilbert DN. Procalcitonin as a biomarker in respiratory tract infection. Clin Infect Dis. 2011;52 Suppl 4:S346-50. doi: 10.1093/cid/cir050. PubMed PMID: 21460294.
143. World Health Organization. Global Influenza Surveillance Network. Manual for the laboratory diagnosis and virological surveillance of influenza. 2011.
144. Papenburg J, Baz M, Hamelin ME, Rheaume C, Carbonneau J, Ouakki M, et al. Household transmission of the 2009 pandemic A/H1N1 influenza virus: elevated laboratory-confirmed secondary attack rates and evidence of asymptomatic infections. Clin Infect Dis. 2010;51(9):1033-41. doi: 10.1086/656582. PubMed PMID: 20887206.
145. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. Journal of clinical epidemiology. 2006;59(10):1087-91. doi: 10.1016/j.jclinepi.2006.01.014. PubMed PMID: 16980149.
146. Allison PD. Missing Data. University of Pennsylvania: SAGE Publications, Inc; 2001.
147. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4. PubMed PMID: 8668867.
148. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. Journal of clinical epidemiology. 2001;54(8):774-81. PubMed PMID: 11470385.
149. Peterson LR. Molecular laboratory tests for the diagnosis of respiratory tract infection due to Staphylococcus aureus. Clin Infect Dis. 2011;52 Suppl 4:S361-6. Epub 2011/04/06. doi: 10.1093/cid/cir052. PubMed PMID: 21460297.
150. Plouffe JF. Importance of atypical pathogens of community-acquired pneumonia. Clin Infect Dis. 2000;31 Suppl 2:S35-9. Epub 2000/09/13. doi: 10.1086/314058. PubMed PMID: 10984326.
151. Blaschke AJ. Interpreting assays for the detection of Streptococcus pneumoniae. Clin Infect Dis. 2011;52 Suppl 4:S331-7. Epub 2011/04/06. doi: 10.1093/cid/cir048. PubMed PMID: 21460292; PubMed Central PMCID: PMC3069982.
152. Bartlett JG. Diagnostic test for etiologic agents of community-acquired pneumonia. Infectious disease clinics of North America. 2004;18(4):809-27. Epub 2004/11/24. doi: 10.1016/j.idc.2004.08.002. PubMed PMID: 15555826.
153. Nolte FS. Molecular diagnostics for detection of bacterial and viral pathogens in community-acquired pneumonia. Clin Infect Dis. 2008;47 Suppl 3:S123-6. Epub 2008/11/18. doi: 10.1086/591392. PubMed PMID: 18986277.
154. Pfuntner A, Wier LM, Stocks C. Most frequent conditions in U.S. hospitals, 2011. HCUP Statistical Brief #162. Rockville, MD: Agency for Healthcare Research and Quality, 2013.

155. Yu H, Wier LM, Elixhauser A. Hospital stays for children, 2009. HCUP Statistical Brief #118. Rockville, MD: Agency for Healthcare Research and Quality, 2011.
156. Singh K, Vasoo S, Stevens J, Schreckenberger P, Trenholme G. Pitfalls in diagnosis of pandemic (novel) A/H1N1 2009 influenza. J Clin Microbiol. 2010;48(4):1501-3. doi: 10.1128/JCM.02483-09. PubMed PMID: 20164266; PubMed Central PMCID: PMCPMC2849562.
157. Jain S, Self WH, Wunderink RG, Fakhran S, Balk R, Bramley AM, et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. N Engl J Med. 2015;373(5):415-27. doi: 10.1056/NEJMoa1500245. PubMed PMID: 26172429.
158. Jain S, Williams DJ, Arnold SR, Ampofo K, Bramley AM, Reed C, et al. Community-acquired pneumonia requiring hospitalization among U.S. children. N Engl J Med. 2015;372(9):835-45. doi: 10.1056/NEJMoa1405870. PubMed PMID: 25714161.
159. Marshall RJ. The Predictive Value of Simple Rules for Combining Two Diagnostic Tests. Biometrics. 1989;45(4):1213-22.
160. Little RJA. Pattern-Mixture Models for Multivariate Incomplete Data. Journal of American Statistical Association. March 1993;88(421):125-34.
161. Rubin DB. Multiple Imputation for Nonresponse in Surveys. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1987. 287 p.
162. Baughman AL, Bisgard KM, Cortese MM, Thompson WW, Sanden GN, Strebel PM. Utility of composite reference standards and latent class analysis in evaluating the clinical accuracy of diagnostic tests for pertussis. Clin Vaccine Immunol. 2008;15(1):106-14. doi: 10.1128/CVI.00223-07. PubMed PMID: 17989336; PubMed Central PMCID: PMC2223853.
163. Dorny P, Phiri IK, Vercruysse J, Gabriel S, Willingham AL, 3rd, Brandt J, et al. A Bayesian approach for estimating values for prevalence and diagnostic test characteristics of porcine cysticercosis. Int J Parasitol. 2004;34(5):569-76. doi: 10.1016/j.ijpara.2003.11.014. PubMed PMID: 15064121.
164. Miller WC. Can we do better than discrepant analysis for new diagnostic test evaluation? Clin Infect Dis. 1998;27(5):1186-93. PubMed PMID: 9827267.
165. Naaktgeboren CA, Bertens LC, van Smeden M, de Groot JA, Moons KG, Reitsma JB. Value of composite reference standards in diagnostic research. BMJ. 2013;347:f5605. doi: 10.1136/bmj.f5605. PubMed PMID: 24162938.
166. Schiller I, van Smeden M, Hadgu A, Libman M, Reitsma JB, Dendukuri N. Bias due to composite reference standards in diagnostic accuracy studies. Stat Med. 2015. doi: 10.1002/sim.6803. PubMed PMID: 26555849.
167. Macaskill P, Walter SD, Irwig L, Franco EL. Assessing the gain in diagnostic performance when combining two diagnostic tests. Stat Med. 2002;21(17):2527-46. doi: 10.1002/sim.1227. PubMed PMID: 12205697.
168. Fan J, Henrickson KJ. Rapid diagnosis of human parainfluenza virus type 1 infection by quantitative reverse transcription-PCR-enzyme hybridization assay. J Clin Microbiol. 1996;34(8):1914-7. PubMed PMID: 8818880; PubMed Central PMCID: PMC229152.
169. Boivin G, Cote S, Dery P, De Serres G, Bergeron MG. Multiplex real-time PCR assay for detection of influenza and human respiratory syncytial viruses. J Clin

Microbiol. 2004;42(1):45-51. PubMed PMID: 14715730; PubMed Central PMCID: PMC321675.
170. Freymuth F, Vabret A, Galateau-Salle F, Ferey J, Eugene G, Petitjean J, et al. Detection of respiratory syncytial virus, parainfluenzavirus 3, adenovirus and rhinovirus sequences in respiratory tract of infants by polymerase chain reaction and hybridization. Clinical and diagnostic virology. 1997;8(1):31-40. PubMed PMID: 9248656.
171. Dean N, Raftery AE. Latent Class Analysis Variable Selection. Ann Inst Stat Math. 2010;62(1):11-35. doi: 10.1007/s10463-009-0258-9. PubMed PMID: 20827439; PubMed Central PMCID: PMCPMC2934856.
172. Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, Leach S, et al. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. Am J Epidemiol. 2008;167(7):775-85. doi: 10.1093/aje/kwm375. PubMed PMID: 18230677.
173. Hayden FG, Fritz R, Lobo MC, Alvord W, Strober W, Straus SE. Local and systemic cytokine responses during experimental human influenza A virus infection. Relation to symptom formation and host defense. J Clin Invest. 1998;101(3):643-9. doi: 10.1172/JCI1355. PubMed PMID: 9449698; PubMed Central PMCID: PMCPMC508608.
174. Cowling BJ, Chan KH, Fang VJ, Lau LL, So HC, Fung RO, et al. Comparative epidemiology of pandemic and seasonal influenza A in households. N Engl J Med. 2010;362(23):2175-84. doi: 10.1056/NEJMoa0911530. PubMed PMID: 20558368; PubMed Central PMCID: PMCPMC4070281.
175. Lau LL, Cowling BJ, Fang VJ, Chan KH, Lau EH, Lipsitch M, et al. Viral shedding and clinical illness in naturally acquired influenza virus infections. J Infect Dis. 2010;201(10):1509-16. doi: 10.1086/652241. PubMed PMID: 20377412; PubMed Central PMCID: PMCPMC3060408.
176. Giannella M, Alonso M, Garcia de Viedma D, Lopez Roa P, Catalan P, Padilla B, et al. Prolonged viral shedding in pandemic influenza A(H1N1): clinical significance and viral load analysis in hospitalized patients. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2011;17(8):1160-5. doi: 10.1111/j.1469-0691.2010.03399.x. PubMed PMID: 20946412.
177. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. Third edition. ed. xvi, 500 pages p.
178. Steyerberg EW. Clinical prediction models : a practical approach to development, validation, and updating. xxviii, 497 pages p.
179. Schumacher M, Hollander N, Sauerbrei W. Resampling and cross-validation techniques: a tool to reduce bias caused by model building? Stat Med. 1997;16(24):2813-27. PubMed PMID: 9483716.
180. Bewick T, Myles P, Greenwood S, Nguyen-Van-Tam JS, Brett SJ, Semple MG, et al. Clinical and laboratory features distinguishing pandemic H1N1 influenza-related pneumonia from interpandemic community-acquired pneumonia in adults. Thorax. 2011;66(3):247-52. doi: 10.1136/thx.2010.151522. PubMed PMID: 21252388; PubMed Central PMCID: PMC3047189.
181. Bender JM, Ampofo K, Gesteland P, Stoddard GJ, Nelson D, Byington CL, et al. Development and validation of a risk score for predicting hospitalization in children with

influenza virus infection. Pediatric emergency care. 2009;25(6):369-75. doi: 10.1097/PEC.0b013e3181a792a9. PubMed PMID: 19458562.
182. Peters TR, Suerken CK, Snively BM, Winslow JE, Nadkarni MD, Kribbs SB, et al. Influenza testing, diagnosis, and treatment in the emergency department in 2009-2010 and 2010-2011. Acad Emerg Med. 2013;20(8):786-94. doi: 10.1111/acem.12175. PubMed PMID: 24033621; PubMed Central PMCID: PMCPMC3775287.
183. Poehling KA, Edwards KM, Weinberg GA, Szilagyi P, Staat MA, Iwane MK, et al. The underrecognized burden of influenza in young children. N Engl J Med. 2006;355(1):31-40. doi: 10.1056/NEJMoa054869. PubMed PMID: 16822994.
184. Garg S, Chaves SS, Perez A, D'Mello T, Gershman K, Meek J, et al. Reduced influenza antiviral treatment among children and adults hospitalized with laboratory-confirmed influenza infection in the year after the 2009 pandemic. Clin Infect Dis. 2012;55(3):e18-21. doi: 10.1093/cid/cis442. PubMed PMID: 22543024.
185. Doshi S, Kamimoto L, Finelli L, Perez A, Reingold A, Gershman K, et al. Description of antiviral treatment among adults hospitalized with influenza before and during the 2009 pandemic: United States, 2005-2009. J Infect Dis. 2011;204(12):1848-56. doi: 10.1093/infdis/jir648. PubMed PMID: 22013219.