



REVIEW

Best Practices for Identifying Hospitalized Lower Respiratory Tract Infections Using Administrative Data: A Systematic Literature Review of Validation Studies

Germaine Hanquet · Christian Theilacker · Jeffrey Vietri · Ingrid Sepúlveda-Pachón ·
Sonia Menon · Bradford Gessner · Elizabeth Begier

Received: December 19, 2023 / Accepted: February 22, 2024
© Pfizer Inc. 2024

ABSTRACT

Introduction: Estimating the burden of lower respiratory tract infections (LRTIs) increasingly relies on administrative databases using International Classification of Diseases (ICD) codes, but no standard methodology exists. We defined best practices for ICD-based algorithms that estimate LRTI incidence in adults.

Methods: We conducted a systematic review of validation studies assessing the use of ICD code-based algorithms to identify hospitalized LRTIs in adults, published in Medline, EMBASE,

and LILACS between January 1996 and January 2022, according to PRISMA guidelines. We assessed sensitivity, specificity, and other accuracy measures of different algorithms.

Results: We included 26 publications that used a variety of ICD code-based algorithms and gold standard criteria, and 18 reported sensitivity and/or specificity. Sensitivity was below 80% in 72% (38/53) of algorithms and specificity exceeded 90% in 77% (37/48). Algorithms for all-cause LRTI ($n=18$) that included only pneumonia codes in primary position ($n=3$) had specificity greater than 90% but low sensitivity (55–72%). Sensitivity increased by 5–15%, with minimal loss in specificity, with the addition of primary codes for severe pneumonia (e.g. sepsis) while pneumonia codes were in secondary position, and by 13% with codes from LRTI-related infections (e.g. viral) or other respiratory diseases (e.g. empyema). Sensitivity increased by 8% when pneumonia codes were in any position, but specificity was not reported. In hospital-acquired pneumonia and pneumococcal-specific pneumonia, algorithms containing only nosocomial- or pathogen-specific ICD codes had poor sensitivity, which improved when broader pneumonia codes were added, in particular codes for unspecified organisms.

Conclusion: Our systematic review highlights that most ICD code-based algorithms are relatively specific, but miss a substantial number of hospitalized LRTI adult cases. Best practices

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40121-024-00949-8>.

G. Hanquet · I. Sepúlveda-Pachón · S. Menon
P95 Epidemiology and Pharmacovigilance, Koning
Leopold III Laan 1, 3001 Louvain, Belgium

C. Theilacker
Pfizer Inc., Linkstrasse 10, 10785 Berlin, Germany

J. Vietri · B. Gessner
Pfizer Inc., 500 Arcola Rd, Collegeville, PA 19426,
USA

E. Begier (✉)
Scientific Affairs, Older Adult RSV Vaccine Program,
Global Medical Development Scientific and Clinical
Affairs, Pfizer Vaccines, Pfizer Inc., 9 Riverwalk,
Citywest Business Campus, Dublin 24, Dublin,
Ireland
e-mail: Elizabeth.Begier@pfizer.com

to estimate LRTI incidence in this population include the use of all pneumonia ICD codes for any LRTI outcome and, to a lesser extent, those for other LRTI-related infections or respiratory diseases.

Keywords: Administrative data; ICD; Incidence; Lower respiratory tract infection; Pneumonia; Sensitivity; Specificity; Systematic review; Validation

Key Summary Points

Lower respiratory tract infection (LRTI) burden estimates increasingly rely on administrative databases using International Classification of Diseases (ICD) codes, but no standard methodology exists.

Our systematic review of validation studies showed that most ICD code-based algorithms miss a substantial number of hospitalized LRTI cases in adults, whether community- or hospital-acquired cases.

ICD-based algorithms containing only nosocomial- or pathogen-specific ICD codes to identify hospital-acquired and pneumococcal pneumonia have poor sensitivity.

ICD-based algorithms should include all pneumonia ICD codes, in particular those for unspecified organisms, to estimate the incidence of all-cause hospitalized LRTI and hospital-acquired pneumonia in adults.

INTRODUCTION

Lower respiratory tract infections (LRTIs) are among the top five leading causes of morbidity and mortality worldwide. The most frequent LRTIs are community-acquired pneumonia (CAP) and hospital-acquired pneumonia (HAP), which differ in etiology, treatment, and mortality [1, 2]. Pneumonia signs and symptoms vary while diagnostic tests, including chest X-ray, are

non-specific, making clinical diagnosis challenging particularly in adults [3]. The identification of pneumonia cases is even more complex in studies based on real-world databases. LRTIs can be caused by a variety of viral, bacterial, and fungal etiologies. However, in a high proportion of LRTI no pathogen is detected (e.g. up to 62% of hospitalized adult CAP cases) [4], usually because of limitations of traditional microbiological methods such as blood and sputum cultures, prior antibiotic use, and single specimen testing [5–7].

Estimating LRTI incidence is a cornerstone in assessing the epidemiologic burden of disease and, subsequently, the potential impact of vaccination programs targeting respiratory pathogens. The most reliable incidence data are generated by population-based cohort studies following patients with LRTI in a clearly defined population and using clinical records for case identification. Although such designs remain the gold standard in assessing epidemiological trends of LRTI, they are time-consuming and difficult to generalize across clinical settings. As an alternative, epidemiological studies increasingly rely on administrative databases based on International Classification of Diseases (ICD) codes, which are used primarily for administrative or financial purposes, such as hospital financing. However, study methodologies vary, and their validity to identify LRTI cases and, therefore, to estimate LRTI burden in general, and CAP and HAP in particular, are largely unknown. Furthermore, existing literature on the LRTI burden focuses primarily on children rather than among adults.

To address this gap, we conducted a systematic review to summarize the accuracy of ICD code-based algorithms for identifying community and hospital-acquired LRTI cases and propose best practices for such analyses.

METHODS

The study protocol was registered on PROSPERO (International Prospective Register of Systematic Reviews) with registration number CRD42022299634. This article is based on

previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors.

Study Selection and Data Extraction

A systematic search was conducted in Medline, EMBASE, and LILACS databases for studies published between January 1996 and January 2022 that validated the use of ICD code-based algorithms for the identification of hospitalized LRTIs (including CAP) in adults ≥ 18 years of age. Studies specific to COVID-19-related LRTIs were excluded. Details of the selection process, inclusion and exclusion criteria, as well as the search strategy are provided in Tables S1 and S2. The selection of studies was carried out in two stages. First, title and abstract screening was performed by two reviewers, and discrepancies were resolved by a third reviewer. Full-text screening was conducted by one reviewer and 10% of non-selected articles were checked by a second reviewer. Hand searching was performed for the references of included articles and previous SLRs identified, and grey literature was searched on OpenGrey [8]. Besides study and algorithm characteristics, we extracted the following validation measures: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and positive/negative likelihood ratio (LR+ and LR-), when available. When LRs were not available, they were calculated as follows: $LR+ = \text{sensitivity} / (1 - \text{specificity})$ and $LR- = (1 - \text{sensitivity}) / \text{specificity}$. The risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) [9], which was tailored to the review. Table S3 outlines the components of each domain.

Data Analysis

Data from each study were summarized descriptively, by main outcome. When validation measures were not provided and numbers were available, we calculated the missing indicators, in particular LR+ and LR-. Pooled analysis was not considered because we sought to establish best practices for this type of analysis rather than quantify incidence.

Best Practices

The identification of best practices for establishing ICD-based algorithms to estimate hospitalized LRTI incidence was based on the sensitivity, specificity, LR+, and LR- of the algorithms, and not on predictive values to avoid biased statistics from varying disease prevalence [10]. We compared values from studies testing different algorithms and retrieved information from the discussion sections of included papers to collect lessons learned regarding ICD algorithms from individual studies to derive best practices.

RESULTS

Study Description

Our database search identified 1697 unique references and 838 studies via hand searching, of which 26 studies were eligible and contributed to the final analysis (Fig. 1). These were conducted in ten high-income countries (Table 1), including 14 (56%) from the USA [11–24]. Six (23%) studies included a separate analysis for patients with comorbidities [14, 18, 22, 23, 25, 26], including chronic obstructive pulmonary disease and immune suppression. Overall, 24 studies reported on all-cause LRTI, including 17 on any or community-acquired LRTI (two on LRTI only, five on any pneumonia, nine on CAP, and one on empyema) [11, 12, 14–16, 18–22, 25, 27–32] and seven on HAP [23, 24, 26, 33–36] (Table S4). Three studies involved pathogen-specific LRTI (one covered both, Table 3) [13, 17, 32].

Among the 53 algorithms, the reference standards used to confirm LRTI were based on either the review of medical files (63% of algorithms), including clinical, laboratory and/or radiological data, or on the diagnosis established by the treating physician (Table S4). Ten studies used reference standards based on explicit clinical criteria.

Overall, 22 of the 26 included studies were graded as having a risk of bias: 16 had one uncertainty, and 11 had one high risk of bias in at least one domain (see Table S5 and Fig. S1).

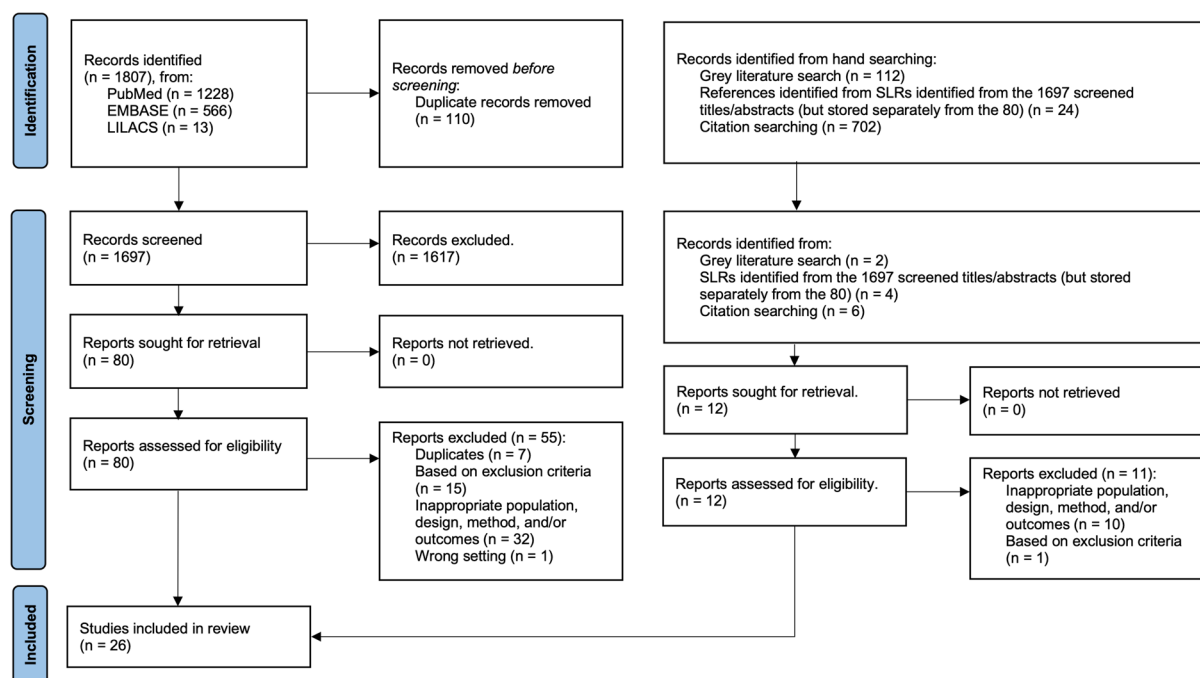


Fig. 1 PRISMA diagram

Characteristics and Accuracy of ICD Algorithms

The algorithms included ICD-9 ($n = 17$ studies) or ICD-10 codes ($n = 9$), alone or in combination with additional criteria, such as length of stay ($n = 2$) or free text search ($n = 2$) such as natural language processing. All studies, except three HAP studies, included ICD codes from the “classical” *Pneumonia and influenza* ICD group (ICD-9 480–488 or ICD-10 J10–J18), named here *classical* pneumonia codes. These were the only codes in seven studies, while ten studies included codes for pneumonia due to specific pathogens (ICD-9 001–139 or ICD-10 from A and B groups) and/or other respiratory codes (other codes from ICD-9 460–519 codes or ICD-10 J group than those of pneumonia above). Some algorithms included ($n = 7$, five CAP and two HAP studies) or excluded ($n = 2$ CAP studies) “Aspiration pneumonia” codes (ICD-9 507 or ICD-10 J69). Pneumonia classical codes were required to be in primary position only ($n = 7$), in any position ($n = 12$), or position not stated

($n = 4$). Three studies added codes of disease severity (e.g. respiratory failure and sepsis) in primary position when pneumonia codes were in secondary position. HAP studies included additional criteria such as specific codes for nosocomial infection and/or pneumonia codes not present at admission, see below.

Among the 26 studies, eight reported only PPV and/or NPV and were excluded from the analysis on algorithm performance (see “Methods”), and data are in Table S6. The remaining 18 studies reported sensitivity and/or specificity and the performances of their algorithms are described below by clinical outcome. These include 16 studies on all-cause LRTI (six on HAP only) and two on pathogen-specific LRTI.

Any or Community-Acquired LRTI, All Causes

Among the ten studies on non-HAP all-cause LRTI (i.e. excluding those focusing on only HAP), two, three, and five studies involved any LRTI, any pneumonia, and any CAP respectively (Table 1). The 18 algorithms

Table 1 Characteristics and validation measures of ICD algorithms to identify LRTI in adults (excluding those specific to hospital-acquired infection)

Country (study)	Patient population	Algorithm and coding position*	Pneumonia codes	Aspiration pneumonia	Other infection codes (that are not pneumonia codes)	Other codes of respiratory system	Reference standard	Se (%)	Sp (%)	LR+	LR-
Any LRTI											
ICD-10											
Denmark (Henriksen, 2014)	≥ 15 years	ICD-10, any position	J12–18, without influenza and some missing	Yes (J69)	Yes, bacterial and fungal pneumonia, opportunistic LRTI, tuberculosis	Yes, bronchitis, lung abscess, COPD with LRTI, pyothorax	Chart review (clinical, radiological and lab)	71	92	8	0.32
Thailand (Ratanaumpawan, 2016)	≥ 15 years	ICD-10, any position	J12–18, without influenza	No	Yes, bacterial, viral, fungal, parasitic pneumonia and tuberculosis	No	Chart review (clinical and lab)	57	94	9	0.46
Pneumonia overall											
ICD-9											
USA (Aronsky, 2005)	≥ 18 years	ICD-9 codes in primary	480–487, without infectious diseases classified elsewhere (484)	No	No	No	Chart review (clinical and radiological) and recoding	55	99	61	0.46
		ICD-9 codes in primary	480–487	Yes (507)	Yes, bacterial, viral, fungal, toxoplasmosis pneumonia	Yes, empyema and other unspecified pneumonia-related		68	99	68	0.32
		ICD-9 primary, or secondary with codes of pneumonia severity as primary	480–487, without infectious diseases classified elsewhere (484)	Yes (507)	No**	No**		70	99	63	0.31
USA (Ahmed, 2014)	≥ 18 years, patients at risk of ARDS	ICD-9 codes, position not stated	480–486, without influenza	No	No	No	Chart review and concordance between three algorithms	77	98	39	0.23
ICD-10											
Australia (Skull, 2008)	≥ 65 years	ICD-10, any position	J10–18	No	No	No	Radiological data	89	62	2	0.17
							Chart review	98	97	32	0.02
							Chart review and radiological data	98	64	3	0.03

Table 1 continued

Country (study)	Patient population	Algorithm and coding position*	Pneumonia codes	Aspiration pneumonia	Other infection codes (that are not pneumonia codes)	Other codes of respiratory system	Reference standard	Se (%)	Sp (%)	LR+	LR-
Community-acquired pneumonia											
ICD-9											
USA (Whittle, 1997)	≥ 18 years, excluding trauma and elective surgery patients	ICD-9, primary ICD-9 primary or secondary with severity as primary, or "pneumonia" in patient classification	480-487 480-487	Yes (507)	Yes, bacterial, viral, fungal, toxoplasmosis pneumonia	Yes, empyema and other unspecified pneumonia-related	Chart review (clinical and radiological)	84	86	6	0.19
Singapore (Mukhopadhyay, 2017)	≥ 21 years	ICD-9, position not stated + text mining for pneumonia in admission and in X-ray report	480-487	No	No	No	Diagnosis of reviewing physician based on chart review	85	78	4	0.20
Netherlands (van de Gard, 2007)	≥ 18 years	ICD-9, primary	481-486, without viral pneumonia and influenza	No	No	No	Chart review (clinical, laboratory and radiological)	72			
		ICD-9, any position	481-486, without viral pneumonia and influenza	No	No	No	Chart review (clinical, laboratory and radiological)	80			
USA (Yu, 2011)	18-64 years	ICD-9, primary	480-487	Yes (507)	No	No	Diagnosis by treating physician	63	93	9	0.40
	≥ 65 years	ICD-9, primary			No			65	85	4	0.41
	18-64 years	ICD-9 pneumonia codes primary OR LOS ≤ 3 days without 507 OR LOS > 3 days with ARF in primary	480-487	Excluded (507)	No	No**	Diagnosis by treating physician	81	82	5	0.23
	≥ 65 years	ICD-9 primary OR LOS ≤ 3 days without 507 OR LOS > 3 days with ARF in primary OR no pneumonia codes in primary nor 507.0 in any field, with LOS ≤ 9 days without an operation procedure code	480-487	Excluded (507)	No	No**	Diagnosis by treating physician	89	63	2	0.17

Table 1 continued

Country (study)	Patient population	Algorithm and coding position*	Pneumonia codes	Aspiration pneumonia	Other infection codes (that are not pneumonia codes)	Other codes of respiratory system	Reference standard	Se (%)	Sp (%)	LR+	LR-
USA (Jones, 2018)	≥ 18 years seen at ED	ICD-9 primary, or secondary with RF or sepsis in primary, ED	481–486, without viral pneumonia and influenza	No	No**	No**	Diagnosis or treatment by treating physician	44	100	∞	0.56
	≥ 18 years hospitalized	ICD-9 primary, or secondary with RF or sepsis in primary, discharge			No**	No**	Diagnosis or treatment by treating physician	67	99	67	0.33
	≥ 18 years seen at ED	ICD-9 primary, or secondary with sepsis or RF in primary in ED + natural language processing	481–486, without viral pneumonia and influenza	No	No**	No**	Diagnosis or treatment by treating physician	89	98	45	0.11

ARDS acute respiratory distress syndrome, ARF acute respiratory failure (518.81), COPD chronic obstructive pulmonary disease, ED emergency department, ICD International Classification of Diseases, LOS length of stay, LR+ likelihood ratio positive, LR- likelihood ratio negative, LRTI lower respiratory tract infection, RF respiratory failure, Se sensitivity, Sp specificity

*Coding position refers in most cases to “Classical pneumonia codes position”, but sometimes to other pneumonia-related codes

**However, some codes (e.g. sepsis or respiratory failure) from this group are included as severity indicator

did not differ across outcomes and included either ICD-9 ($n=15$) or ICD-10 ($n=3$) codes. Sensitivity was at least 80% in around 10/18 of the algorithms reporting it, while specificity was above 90% in 9/16 of them. LR+ was high in three-quarters of the algorithms (13/17, $\text{LR+} \geq 5$) and LR- was low in one-third of them (6/17, ≤ 0.20).

The three algorithms based on classical pneumonia codes in primary position only (including or not aspiration pneumonia) yielded a low sensitivity (range 55–72%) and high specificity (>93%), with varying LR+ and LR- [11, 21, 32]. Sensitivity increased with minimal loss in specificity in algorithms that included codes of pneumonia severity (sepsis or respiratory failure) in the primary position when pneumonia was in a secondary position, by 15% compared to the above algorithm [11], and by 5% compared to the above algorithm combined with codes for other pathogens and respiratory codes [15]. Sensitivity increased by 13% when infection or other respiratory codes (such as empyema, pleurisy, or lung abscess) were added to pneumonia codes in primary position only, while specificity remained high [11]. In the four algorithms with pneumonia ICD codes in any position, sensitivity (57–98%), specificity (62–97%), LR+ (2–32), and LR- (0.02–0.46) varied [27–29, 32]; and no major change was observed when other infection or respiratory codes were added [27]. Sensitivity (85–89%) was high and LR- was below 0.2 when text search (text mining or natural language processing) was added to ICD codes, while specificity (78–98%) and LR+ varied [19, 31]. More complex algorithms, using predictors identified through analysis (such as length of hospital stay), reached a high sensitivity (81–89%) and lower LR- (around 0.2) but a lower specificity (63–82%) and LR+ (2–5) [21].

The influence of the reference standard used for case confirmation has been illustrated in one study, in which confirmation by radiological data only led to a drop in both sensitivity (from 98% to 89%) and specificity (from 97% to 62%) compared to chart reviews of medical files for the same algorithm [29]. The type of patients also had an impact on the performance values of algorithms when different groups were included, with higher sensitivity

and lower specificity in older adults (≥ 65 years) compared to younger adults (18–64 years) [21, 29], and higher sensitivity in hospitalized patients compared to those seen at emergency departments [19].

Hospital-Acquired Pneumonia, All Causes

Among the six studies (nine algorithms) on HAP (Table 2), five involved any HAP (two with ICD-9 and three with ICD-10) and one based on ICD-9 included ventilator-associated pneumonia (VAP) only [23, 24, 26, 33–35]. Six algorithms included specific ICD codes for HAP or VAP [24, 34, 35], and five included classical pneumonia codes (without HAP-specific codes in three algorithms) [23, 24, 26, 33]. All nine algorithms had ICD codes in secondary or any position (or not stated), and six required the ICD codes to be not present on admission [24, 26, 34]. The three algorithms that used HAP/CAP codes alone with the present on admission criteria presented a very low sensitivity ($\leq 25\%$), high specificity ($\geq 98\%$), high LR+ (range 83–233) and poor LR- (0.75–0.77) [24, 34]. When classical pneumonia codes were added to the specific VAP code and present on admission criteria, sensitivity increased to 61% and LR- improved (0.42–0.47) while specificity slightly declined (83–93%) and LR+ dropped to 4–9 [24]. The three algorithms including only classical pneumonia codes displayed a higher sensitivity (35–100%) than specific HAP/VAP codes, high specificity (99–100%), high LR+ (44–333) and varying LR- (0.00–0.65) [23, 26, 33]. Algorithms showed similar performance in patients with continuous invasive mechanical ventilation as in the total patient population [24].

Pathogen-Specific LRTI

All three studies assessing pathogen-specific LRTI included pneumococcal pneumonia, one study covered ten other pathogens [13], and they included 26 different algorithms (Table 3), after excluding data for which the reference standard included possible or probable cases [17]. Pathogen-specific ICD-9 codes were included in all

Table 2 Characteristics and validation measures of ICD algorithms to identify hospital-acquired pneumonia in adults

Country (study)	Patient population	Algorithm and coding position*	Pneumonia codes	Aspiration pneumonia	Other infection codes (that are not pneumonia codes)	Other codes of respiratory system	Reference standard	Sc (%)	Sp (%)	LR+	LR-
Any hospital-acquired pneumonia											
ICD-9											
Canada (Quan, 2004)	≥ 18 years	ICD-9, any position (up to I6)	480–486, without influenza	Yes (507)	Yes, bacterial, viral, fungal, parasitic pneumonia and tuberculosis	Yes, empyema and other unspecified pneumonia-related	Chart review using recoding	35	99	58	0.65
USA (Romano, 2002)	≥ 18 years with elective lumbar discectomy	ICD-9, position not specified	480–487	Yes (507)	No	Yes, empyema and other unspecified pneumonia-related	Chart review using recoding	100	100	333	0.00
ICD-10											
Germany (Azaouagh, 2008)	≥ 18 years with mixed comorbidities	ICD-10 codes not present on admission	J10–18		Viral and bacterial pneumonia specific codes	Other affections related to respiratory system	Chart review (clinical, radiological and laboratory data)	43	99	44	0.58
Switzerland (Wolfensberger, 2018)	≥ 18 years	ICD-10 code for HAP	Code for HAP (U69,00)	No	No	No	Chart review (clinical, radiology and lab—HELICS definition)	59	98	33	0.42
Germany (Maass, 2015)	≥ 65 years	ICD-10 codes not present on admission and nosocomial origin, any position	Code for HAP (U69,00, German version)	No	No	No	Chart review, according to CDC criteria	23	100	233	0.77

Table 2 continued

Country (study)	Patient population	Algorithm and coding position*	Pneumonia codes	Aspiration pneumonia	Other infection codes (that are not pneumonia codes)	Other codes of respiratory system	Reference standard	Se (%)	Sp (%)	LR+	LR-
Ventilator-associated pneumonia											
ICD-9											
USA (Cass, 2013)	≥ 18 years	ICD-9 for ventilator-associated pneumonia, primary, not present on admission (restricted)	None	No	No	Ventilator-associated pneumonia	As defined by hospital surveillance	25	100	83	0.75
	≥ 18 years	ICD-9 30 diagnosis codes + ventilator-associated pneumonia, not present on admission (expanded)	480–487, without infectious diseases classified elsewhere (484)	No	No	Ventilator-associated pneumonia	As defined by hospital surveillance	61	93	9	0.42
	≥ 18 years with continuous invasive mechanical ventilation	ICD-9 for ventilator-associated pneumonia, not present on admission (restricted)	None	No	No	Ventilator-associated pneumonia	As defined by hospital surveillance	25	99	21	0.76
	≥ 18 years with continuous invasive mechanical ventilation	ICD-9 30 diagnosis codes + ventilator-associated pneumonia, not present on admission (expanded)	480–487, without infectious diseases classified elsewhere (484)	No	No	Ventilator-associated pneumonia	As defined by hospital surveillance	61	83	4	0.47

CDC Centers for Disease Control and Prevention, *HELICS* Hospitals in Europe Link for Infection Control through surveillance, *HAP* hospital-acquired pneumonia, ICD International Classification of Diseases, *LR+* likelihood ratio positive, *LR-* likelihood ratio negative, *LRTI* lower respiratory tract infection, *Se* sensitivity, *Sp* specificity

*Coding position refers to in most cases to “Classical pneumonia codes position”, but sometimes to other pneumonia-related codes

algorithms, in primary ($n=3$ algorithms), in primary or secondary with severity codes ($n=12$), or any ($n=11$) position. Sepsis or bacteremia general codes were added in 16 algorithms. The reference standard was based on laboratory tests, with or without clinical and radiological criteria.

The performance of algorithms varied across specific pathogens [13], with sensitivity ranging from 14% for parainfluenza to 96% for influenza and specificity being always high ($\geq 98\%$). In the three studies involving pneumococcal pneumonia, sensitivity of the pneumococcal-specific code (ICD-9 481) alone was low to moderate (35–58%), while specificity was high (98–99%). In one study, sensitivity was 45% when the pneumococcal pneumonia code was in primary position and 58% when it was in any position. In the same study, sensitivity increased from 58% to 89% when the codes for pneumonia from organism unspecified were added (ICD-9 485–486, any position) [17]. However, specificity declined in the latter, from 98% to 45%. The addition of the acute respiratory failure code (518.81) did not improve performance [17].

Best Practices Learned from Included Studies

Distinguishing Community-Acquired from Hospital-Acquired Pneumonia

In the included studies, the distinction between CAP and HAP cases was performed in the inclusion of pneumonia-suspected cases, in the algorithms applied to these, and/or in the reference standard. In all 26 studies, including those not providing sensitivity and specificity, the case definitions for the inclusion of suspected HAP or CAP frequently included a time threshold, i.e. diagnosis or medical information being obtained or reported within or after 24 or 48 h after hospital admission. At the algorithm level, seven of the nine CAP studies (10 algorithms) included codes (pneumonia, or severity codes—with pneumonia as secondary) in primary position only [15, 16, 18–21, 32]. Other criteria to exclude HAP were antibiotic prescription within

72 h after admission [22], and the exclusion of patients with major trauma or elective surgery [15]. In HAP studies, the algorithm criteria most frequently used to exclude CAP are pneumonia ICD codes in secondary or in any position and/or not present on admission [24, 26, 34–36], and the use of specific HAP codes, such as U69.00 (elsewhere classified, hospital-acquired pneumonia) [34, 35], and 997.31 for VAP [24, 33, 36]. However, the performance of the differentiating criteria has not been evaluated.

Algorithms Including Code Position

In the five non-HAP LRTI studies comparing the sensitivity and/or specificity of algorithms [11, 15, 19, 21, 32], adding primary codes of pneumonia severity (sepsis or respiratory failure) when pneumonia was in secondary position or adding other respiratory codes to classical pneumonia codes improved sensitivity with minimal losses in specificity [11, 15]. The benefit of using pneumonia codes in primary or any position is not clear for CAP: sensitivity slightly improved in one study but other accuracy measures were not available [32], and another study made additional changes than coding position [11]. Adding text search in the medical file, such as natural language processing, doubled the sensitivity relative to classical pneumonia codes alone in one study, with limited impact on specificity [19]. The benefit of adding or excluding the aspiration pneumonia code is also unclear, as studies applied other changes in algorithms [11, 21]. The comparison of performance between ICD-9 and ICD-10 versions is not available.

In the single HAP study comparing algorithms, adding classical pneumonia codes improved the poor sensitivity of specific HAP code, with an ensuing slight decrease in specificity [24]. In the pathogen-specific CAP studies, including any position and adding other pneumonia codes to pathogen-specific codes increased sensitivity [17, 32], but specificity declined when codes for organism unspecified were included [17].

Table 3 Characteristics and validation measures of ICD algorithms to identify pathogen-specific CAP in adults

Country (study)	Population	Algorithm and coding position*	Pneumonia codes	Aspiration pneumonia	Other infection codes (that are not pneumonia codes)	Other codes of respiratory system	Reference standard	Pathogen	Se (%)	Sp (%)	LR+	LR-	
Pneumonia													
USA (Higgins, 2020)	≥ 18 years	ICD-9 codes in primary, OR secondary with ARF or sepsis in primary, discharge	482.41	No	Yes, sepsis and bacteremia codes	Yes, other unspecified pneumonia-related, ARF	Laboratory data (blood or respiratory culture, urinary antigen, or PCR)	Methicillin-sensitive <i>S. aureus</i>	14	100	142	0.9	
			482.42							49	99	82	0.5
			481, 482.30						<i>S. pneumoniae</i>	60	99	75	0.4
			482.1					<i>Pseudomonas</i>	47	100	156	0.5	
			482.82					<i>E. coli</i>	17	100			0.8
			482					<i>Klebsiella pneumoniae</i>	36	100	357	0.6	
			482.2					<i>H. influenzae</i>	43	100	428	0.6	
			483					<i>Mycoplasma pneumoniae</i>	54	100	268	0.5	
			482.84					<i>Legionella</i>	78	100	784	0.2	
			487.x, 488.x					Influenza	96	99	87	0.0	
			480.1					RSV	24	100			0.8
			480.2					Parainfluenza	14	100			0.9

Table 3 continued

Country (study)	Population	Algorithm and coding position*	Pneumonia codes	Aspiration pneumonia	Other infection codes (that are not pneumonia codes)	Other codes of respiratory system	Reference standard	Pathogen	Se (%)	Sp (%)	LR+	LR-	
Community-acquired pneumonia													
USA (Guevara, 1999)	≥ 18 years	ICD-9-CM code in primary	481	No	No	No	Clinical data, radiological and laboratory data (blood or pleural fluid)	<i>S. pneumoniae</i>	45	98	2.1	0.6	
		ICD-9-CM code in 1-5 position		No	No	No			58	98	2.3	0.4	
		ICD-9-CM code in primary	482.3	No	No	No	No			7	99	8	0.9
		ICD-9-CM code in 1-5 position		No	No	No	No			11	99	11	0.9
		ICD-9-CM code in first position	486	No	No	No	No			11	55	0	1.6
		ICD-9-CM code in 1-5 position		No	No	No	No			14	48	0	1.8
		ICD-9-CM codes in 1-5 position	481.00	No	No	Yes, pneumococcal septicemia	No			64	97	25	0.4
			481.00	No	No	Yes, pneumococcal and streptococcal septicemia	No			77	97	25	0.2
			481.00, 482.30	No	No		No			81	96	21	0.2
			481.00, 482.30	No	No	Yes, ARF*	Yes, ARF*			81	96	20	0.2
Netherlands (Van de Garde, 2007)	≥ 18 years	ICD-9 codes in primary or secondary admission and discharge	481, 486.00	No	No	No	Clinical, radiological and laboratory data (From sputum and blood, urinary antigen)	<i>S. pneumoniae</i>	35				
			481	No	No	No			18				
			482.x, 483.x	No	No	No		Pneumonia with other organism specified					
		485-486	No	No	No		Pneumonia, organism unspecified	63					

ARF acute respiratory failure (518.81), ICD International Classification of Diseases, CM clinical modification, LR+ likelihood ratio positive, LR- likelihood ratio negative, PCR polymerase chain reaction, RSV respiratory syncytial virus, Se sensitivity, Sp specificity

*Coding position refers to in most cases to "Classical pneumonia codes position", but sometimes to other pneumonia-related codes

DISCUSSION

In this systematic review of 26 validation studies, we found that the sensitivity of ICD code-based algorithms for identifying LRTI cases among hospitalized adults was below 80% in two-thirds of studies, while specificity exceeded 90% in the majority. For all-cause LRTI studies, algorithms that included only pneumonia codes in primary position had a particular low sensitivity, which increased (with minimal loss in specificity) with the addition of codes for severe pneumonia, other LRTI-related infections, other respiratory diseases and free text search. The influence of coding position on sensitivity and specificity was unclear because algorithms

differed in other characteristics that may have confounded this analysis. In HAP and pneumococcal-specific pneumonia, algorithms containing only nosocomial-specific or pathogen-specific ICD codes had poor sensitivity, which improved when broader pneumonia codes were added, in particular codes for unspecified organisms. The comparison of algorithms performance allowed us to derive some best practices that are summarized in Table 4.

Our review found that a majority of ICD-based algorithms reached a low sensitivity to identify LRTI cases among hospitalized adults. This finding is in line with two previous systematic reviews on hospitalized pneumonia, including one published during our review process [10], and one on hospital-acquired infections

Table 4 Summary of best practices for ICD-based algorithm to estimate the incidence of inpatient LRTI in adults, per clinical syndrome

Best practices	Evidence
Do not use specific ICD codes to determine incidence rates of:	3 studies [13, 17, 32]
<ul style="list-style-type: none"> • Pneumococcal LRTI • Hospital-acquired pneumonia 	5 studies [24, 26, 33–35]
Include pneumonia codes ICD-9 480–488 or ICD-10 J10-18 in primary or any position:	5 studies for primary position [11, 15, 19, 21, 32], 4 for any position [27–29, 32]
<ul style="list-style-type: none"> • Patients with severe pneumonia and underlying diseases will be better captured when pneumonia codes are in any position • Always include the “organism unspecified” pneumonia code (ICD-9 486 or ICD-10 J18.9) • Prefer to include <i>all</i> pneumonia and influenza codes 	3 studies [11, 15, 32] 4 studies [13, 21, 29, 32] 4 studies [15, 21, 29, 31]
Sensitivity improves compared to pneumonia codes in primary position, with limited changes to specificity when:	
<ul style="list-style-type: none"> • Codes for severe pneumonia (e.g. sepsis, acute respiratory failure, pneumothorax) are added in primary position, with pneumonia codes in secondary position • Other infection codes (e.g. pathogens causing pneumonia) and/or respiratory codes (e.g. empyema, lung abscess) are added • Pneumonia and influenza codes are in any position • Free text search for pneumonia terms are added, when feasible 	2 studies [11, 15] 1 study [11] for both infection or respiratory codes 1 study [32], but no data on specificity 1 study [19]
Unclear whether sensitivity would improve or not:	2 studies [11, 21]
<ul style="list-style-type: none"> • Code for aspiration pneumonia is added or excluded 	

ICD International Classification of Diseases, LRTI lower respiratory tract infection

that included pneumonia [37]. The determination of the ICD codes that would improve the sensitivity of an algorithm is complex. The classical pneumonia codes were included in all algorithms, except four HAP-specific algorithms, but achieved low sensitivities overall. The addition of pneumonia severity codes (sepsis or respiratory failure) in primary position when pneumonia was in secondary position increased sensitivity by 5–15% with minimal changes to specificity in two studies on any pneumonia or CAP comparing different algorithms [11, 15]. The addition of other infections and/or respiratory codes was only explored in one study [11], in which sensitivity increased by 13% with the addition of other specific infections and respiratory codes. The limited impact of adding those codes was explained by the few patients who were assigned such codes, as these tend to be underreported in administrative databases [11]. Four studies reported the frequency of specific codes within the pneumonia and influenza group and revealed that the majority of ICD codes are rarely reported in hospitalized adult patients [13, 21, 29, 32]. One exception is the code for “organism unspecified” pneumonia cases (ICD-9 486 or ICD-10 J18.9), which was reported in 51–65% and 39–92% of LRTI or pneumonia cases in primary or any position, respectively, likely due to lack of etiological diagnosis at the time of coding [13, 21, 29, 32]. Therefore, it is essential to include the “organism unspecified” pneumonia code in studies aiming to measure LRTI incidence rates.

The influence of code position on sensitivity and specificity is poorly documented. Only one study on all-cause CAP compared pneumonia codes in primary vs. any position and reported a mild (8%) increase in sensitivity, but specificity was not reported [32]. Interestingly, several studies stated that pneumonia codes in secondary position better identify patients with severe pneumonia and underlying diseases [11, 27, 32], as pneumonia may be reported in secondary position only when it occurs in patients with other diseases, such as congestive heart failure or chronic bronchitis [11]. While not a formal validation study, one recently presented analysis in the UK found that ICD code-based pneumonia incidence using the first five positions correlated

highly with estimates from a concurrent prospective study, but agreement decreases with age [38]. This suggests that algorithms aiming at deriving incidence rates of community-acquired LRTI cases in elderly patients should include pneumonia codes in any position, as this age group has a higher probability of presenting underlying diseases and severe pneumonia.

The distinction between CAP and HAP cases based on ICD codes is complex. In studies using the full medical file, the distinction is usually based on a time criterion, i.e. HAP is often defined as pneumonia with onset reported 24 or 48 h after hospital admission to be distinguished from CAP. Studies based on ICD databases only cannot apply this criterion because time after admission is not translated into ICD codes. There is also a lack of a specific code for HAP in ICD-9 and ICD-10, though some countries such as Germany and Switzerland introduced their own code. The ICD studies used different criteria to distinguish HAP from CAP, such as code position, specific HAP code, or pneumonia code not present at admission, but the performance of these criteria was not measured in these studies. Three studies provided some data on misclassification: 24% of all ICD-coded pneumonia cases were found to be HAP after clinical review [15], HAP represented 50% of misclassified cases in one CAP study [18], and CAP represented 25% of misclassified cases in one HAP study [35]. This highlights the need to introduce clear HAP codes in future ICD coding systems. The ICD-11 version, which has come into effect in 2022, includes an additional code for nosocomial origin (i.e. XB25) but it remains to be seen if it will be used in practice and how it will perform.

The added value of including or excluding ICD codes for aspiration pneumonia in algorithms is unclear. Aspiration, which is the inhalation of oropharyngeal or gastric content into the lower respiratory tract, may lead to aspiration pneumonitis, which is a chemical injury, and to aspiration pneumonia, which accounts for 5–15% of CAP cases [11]. Although there is some overlap between these syndromes, they are distinct clinical entities [39] but with similar microbiologic etiology [40]. Aspiration pneumonitis is often difficult to distinguish from CAP [39]. The seven algorithms including aspiration

pneumonia codes covered different types of LRTI (any LRTI, any pneumonia, CAP, and HAP), those excluding it were CAP algorithms, and the performance of including or excluding it has not been measured separately. Its use is also controversial in the literature [11, 21, 22]. On one hand, including aspiration pneumonia codes may add patients who have pneumonitis instead of lung infection. On the other hand, its exclusion in CAP studies risks eliminating true CAP cases in nursing home residents or other frail patients [11]. Studies excluding aspiration codes may thus exclude a particular spectrum of patients.

Our review highlights the lack of a standard for the reference used to confirm true positive and negative in cases detected by ICD codes, as illustrated by the wide variety of criteria and processes, i.e. manual review of medical charts, diagnosis of the treating physician, both with or without explicit clinical criteria and standard forms. Only ten out of 26 used explicit clinical criteria in the process. A previous SLR also pointed to the variety of criteria and, in particular, to the lack of reproducibility and reliability of using the physician diagnosis of LRTI noted in the medical chart by manual review [10]. This process may be prone to subjective interpretation based on unspecific symptoms and/or clinical, laboratory, and/or imaging findings. However, LRTIs can have atypical presentations in the frail and elderly development and such events may be missed by standardized clinical case definitions. Further, ICD code incidence studies primarily aim to capture diagnoses disease in line with current practice which involved physician judgment. Specifically for methodological studies, greater standardization of the reference process would be helpful for future validation studies of ICD code algorithms to allow a quantitative combination of results among studies.

Our review of 26 eligible studies and 18 studies with sensitivity and/or specificity measures is the first one to our knowledge to cover any LRTI, from community-acquired as well as hospital-acquired origin. Our added value is to compare performance of the different algorithms to inform best practices and recommendations for future protocols for ICD-based studies to derive

LRTI incidence rates. A main limitation is the lack of reproducibility and reliability of the clinical review used as reference standard to confirm pneumonia in a majority of validation studies. Standardization of criteria as a minimum, and prospective studies using explicit criteria to confirm LRTI, including laboratory testing to confirm pathogen-specific CAP, would address that limitation. The diversity of study characteristics and heterogeneity across patients (and their expected prevalence of pneumonia) also limited the inference of our results to other populations and settings. We also expect that our findings are biased by the influence of reimbursement rules, as many countries base their hospital financing on Diagnosis Related Groups which include ICD codes, but we found no data to estimate the extent of this influence. The poor quality of some studies (41% in high risk of bias in at least one domain) limits the robustness of our conclusions. Lastly, many studies have identified that respiratory pathogens play a causative role in broader group of cardiopulmonary conditions than LRTI (e.g. chronic obstructive pulmonary disease and chronic heart failure) [41], and the recommendations here are only intended to capture the subset of respiratory pathogen illness recognized as LRTI by the treating clinicians; the extended impact of these would be best captured by time-series modelling studies [42, 43].

CONCLUSION

Our systematic review highlights that many studies of ICD codes to detect LRTI cases among hospitalized adults, and HAP in particular, while being relatively specific, miss a substantial number of cases as shown by their poor sensitivity and can negatively impact burden of disease assessments and related assessments of the utility of preventive interventions. Best practices to estimate LRTI incidence in this population include the use of ICD codes from the pneumonia and influenza group for any subtype of LRTI, including HAP and pathogen-specific groups. The addition of codes for severe pneumonia, specific pathogens or other respiratory entities close to pneumonia may

improve sensitivity but would capture a limited number of additional cases. Algorithms targeting elderly patients should prefer the use of pneumonia codes in any position to better capture those with underlying diseases and/or severe pneumonia. Future validation studies of ICD algorithms should aim at a higher quality, in particular, the use of a more objective and evidence-based gold standard. Further comparison of algorithms and their performance based on new and higher quality studies may help to derive more robust recommendations to measure the LRTI burden based on ICD codes.

ACKNOWLEDGEMENTS

Medical Writing/Editorial Assistance. The authors would like to thank Omar Okasha for contribution to study protocol, literature search strategy and screening, quality check, analysis and summary tables, Wendy Hartig-Merkel for contribution to study protocol, full-text collection, literature screening, and data extraction, and Jennifer Judy for manuscript review. Editorial assistance in the preparation of this article was provided by Alejandra Gonzalez and Wendy Hartig-Merkel, which was funded by Pfizer.

Author Contributions. Conceptualization and methodology: Germaine Hanquet, Elizabeth Begier, Christian Theilacker, Jeffrey Vietri, and Ingrid Sepúlveda-Pachón; formal analysis: Ingrid Sepúlveda-Pachón, and Germaine Hanquet; writing—original draft preparation: Ingrid Sepúlveda-Pachón, Sonia Menon, and Germaine Hanquet; writing—review and editing: Germaine Hanquet, Elizabeth Begier, Jeffrey Vietri, Bradford Gessner and Christian Theilacker; funding acquisition: Elizabeth Begier.

Funding. This study was funded by Pfizer Inc., New York City, USA, and was a research collaboration between Pfizer Inc. and P95 Epidemiology & Pharmacovigilance, and as such has co-authors from both organizations, with all co-authors contributing across the different phases of the SLR, the drafting and revising of

the manuscript for important intellectual content, the decision to publish, and final approval. Pfizer is funding the journal's publication fee.

Data Availability. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interest. Germaine Hanquet and Ingrid Sepúlveda-Pachón are employees of P95 Epidemiology & Pharmacovigilance, which was contracted by Pfizer to conduct the research described in this manuscript and for manuscript development. Sonia Menon was an employee of P95 Epidemiology & Pharmacovigilance at the time of contributing to the research; her current affiliation is Epitech Research, Avenue de Nenuphars 32, Auderghem 1160, Belgium. Germaine Hanquet has received personal fees from MSD, Sanofi Pasteur, Janssens, SNB, and Pfizer as speaker at an international meeting and as a member of advisory boards, outside the scope of the submitted work. Elizabeth Begier, Christian Theilacker, Bradford Gessner and Jeffrey Vietri are employees of Pfizer Inc. and may hold stocks or stock options.

Ethical Approval. This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is

not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. GBD 2016 Lower Respiratory Infections Collaborators. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis*. 2016;18(11):1191–210.
2. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1459–544.
3. Ramirez JA, Wiemken TL, Peyrani P, et al. Adults hospitalized with pneumonia in the united states: incidence, epidemiology, and mortality. *Clin Infect Dis*. 2017;65(11):1806–12.
4. Jain S, Self WH, Wunderink RG, et al. Community-acquired pneumonia requiring hospitalization among US adults. *N Engl J Med*. 2015;373(5):415–27.
5. Kim PA-O, Deshpande A, Rothberg MB. Urinary antigen testing for respiratory infections: current perspectives on utility and limitations. *Infect Drug Resist*. 2022;2022(15):2219–28.
6. Ramirez JA-O, Carrico R, Wilde A, et al. Diagnosis of respiratory syncytial virus in adults substantially increases when adding sputum, saliva, and serology testing to nasopharyngeal swab RT-PCR. *Infect Dis Ther*. 2023;12(6):1593–603.
7. Onwuchekwa C, Moreo LM, Menon S, et al. Underascertainment of respiratory syncytial virus infection in adults due to diagnostic testing limitations: a systematic literature review and meta-analysis. *J Infect Dis*. 2023;228(2):173–84.
8. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
9. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–36.
10. Corrales-Medina VF, van Walraven C. Accuracy of administrative database algorithms for hospitalized pneumonia in adults: a systematic review. *J Gen Intern Med*. 2021;36(3):683–90.
11. Aronsky D, Haug PJ, Lagor C, Dean NC. Accuracy of administrative data for identifying patients with pneumonia. *Am J Med Qual*. 2016;20(6):319–28.
12. Ahmed A, Thongprayoon C, Pickering BW, et al. Towards prevention of acute syndromes: electronic identification of at-risk patients during hospital admission. *Appl Clin Inform*. 2014;5(1):58–72.
13. Higgins TL, Deshpande A, Zilberberg MD, et al. Assessment of the accuracy of using ICD-9 diagnosis codes to identify pneumonia etiology in patients hospitalized with pneumonia. *JAMA Netw Open*. 2020;3(7):e207750.
14. Kern DM, Davis J, Williams SA, et al. Validation of an administrative claims-based diagnostic code for pneumonia in a US-based commercially insured COPD population. *Int J Chron Obstruct Pulmon Dis*. 2015;10:1417–25.
15. Whittle J, Fine MJ, Joyce DZ, et al. Community-acquired pneumonia: can it be defined with claims data? *Am J Med Qual*. 1997;12(4):187–93.
16. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran’s affairs hospital discharge databases coded serious bacterial infections accurately. *J Clin Epidemiol*. 2007;60(4):397–409.
17. Guevara RE, Butler JC, Marston BJ, Plouffe JF, File TM Jr, Breiman RF. Accuracy of ICD-9-CM codes in detecting community-acquired pneumococcal pneumonia for incidence and vaccine efficacy studies. *Am J Epidemiol*. 1999;149(3):282–9.
18. Grijalva CG, Chung CP, Stein CM, et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. *Pharmacoepidemiol Drug Saf*. 2008;17(9):890–5.
19. Jones BE, South BR, Shao Y, et al. Development and validation of a natural language processing tool to identify patients treated for pneumonia across VA emergency departments. *Appl Clin Inform*. 2018;9(1):122–8.
20. Wiese AD, Griffin MR, Stein CM, et al. Validation of discharge diagnosis codes to identify serious infections among middle age and older adults. *BMJ Open*. 2018;8(6):e020857.
21. Yu O, Nelson JC, Bounds L, Jackson LA. Classification algorithms to improve the accuracy of identifying patients

- hospitalized with community-acquired pneumonia using administrative data. *Epidemiol Infect.* 2011;139(9):1296–306.
22. Rodriguez-Barradas MC, McGinnis KA, Akgün K, et al. Validation for using electronic health records to identify community acquired pneumonia hospitalization among people with and without HIV. *Pneumonia (Nathan)*. 2020;12:6.
 23. Romano P, Chan B, Schembri M, Rainwater J. Can administrative data be used to compare postoperative complication rates across hospitals? *Med Care.* 2002;40:856–67.
 24. Cass AL, Kelly JW, Probst JC, Addy CL, McKeown RE. Identification of device-associated infections utilizing administrative data. *Am J Infect Control.* 2013;41(12):1195–9.
 25. Holland-Bill L, Xu H, Sørensen HT, et al. Positive predictive value of primary inpatient discharge diagnoses of infection among cancer patients in the Danish National Registry of Patients. *Ann Epidemiol.* 2014;24(8):593-7.e1-18.
 26. Azaouagh A, Stausberg J. Frequency of hospital-acquired pneumonia—comparison between electronic and paper-based patient records. *Pneumologie.* 2008;62(5):273–8.
 27. Henriksen DP, Nielsen SL, Laursen CB, Hallas J, Pedersen C, Lassen AT. How well do discharge diagnoses identify hospitalised patients with community-acquired infections?—a validation study. *PLoS ONE.* 2014;9(3): e92891.
 28. Rattanaumpawan P, Wongkamhla T, Thamlikitkul V. Accuracy of ICD-10 coding system for identifying comorbidities and infectious conditions using data from a Thai university hospital administrative database. *J Med Assoc Thailand.* 2016;99(4):368–73.
 29. Skull SA, Andrews RM, Byrnes GB, et al. ICD-10 codes are a valid tool for identification of pneumonia in hospitalized patients aged > or = 65 years. *Epidemiol Infect.* 2008;136(2):232–40.
 30. Søgaaard M, Kornum JB, Schønheyder HC, Thomsen RW. Positive predictive value of the ICD-10 hospital diagnosis of pleural empyema in the Danish National Registry of Patients. *Clin Epidemiol.* 2011;3:85–9.
 31. Mukhopadhyay A, Maliapen M, Ong V, et al. Community-acquired pneumonia case validation in an anonymized electronic medical record-linked expert system. *Clin Infect Dis.* 2017;64:S141–4.
 32. van de Garde EMW, Oosterheert JJ, Bonten M, Kaplan RC, Leufkens HGM. International classification of diseases codes showed modest sensitivity for detecting community-acquired pneumonia. *J Clin Epidemiol.* 2007;60(8):834–8.
 33. Quan H, Parsons GA, Ghali WA. Assessing accuracy of diagnosis-type indicators for flagging complications in administrative data. *J Clin Epidemiol.* 2004;57(4):366–72.
 34. Maass C, Kuske S, Lessing C, Schrappe M. Are administrative data valid when measuring patient safety in hospitals? A comparison of data collection methods using a chart review and administrative data. *Int J Qual Health Care.* 2015;27(4):305–13.
 35. Wolfensberger A, Meier AH, Kuster SP, Mehra T, Meier MT, Sax H. Should International Classification of Diseases codes be used to survey hospital-acquired pneumonia? *J Hosp Infect.* 2018;99(1):81–4.
 36. Verelst S, Jacques J, Van den Heede K, et al. Validation of Hospital Administrative Dataset for adverse event screening. *Qual Saf Health Care.* 2010;19(5): e25.
 37. Redondo-González O, Tenías JM, Arias Á, Lucendo AJ. Validity and reliability of administrative coded data for the identification of hospital-acquired infections: an updated systematic review with meta-analysis and meta-regression analysis. *Health Serv Res.* 2018;53(3):1919–56.
 38. Campling JA, Begier E, Lahuerta M, et al. S111 Adult hospitalised community acquired pneumonia incidence in Bristol: comparison of retrospective ICD-10 based analysis and prospective study data. *Thorax.* 2022. <https://doi.org/10.1136/thorax-2022-BTSabstracts.117>.
 39. Marik PE. Aspiration pneumonitis and aspiration pneumonia. *N Engl J Med.* 2001;344(9):665–71.
 40. Marin-Corral J, Pascual-Guardia S, Amati F, et al. Aspiration risk factors, microbiology, and empiric antibiotics for patients hospitalized with community-acquired pneumonia. *Chest.* 2021;159(1):58–72.
 41. Falsey AR, Hennessey P, Formica MA, Cox C, Walsh EE. Respiratory syncytial virus infection in elderly and high-risk adults. *N Engl J Med.* 2005;352(17):1749–59.
 42. Sharp A, Minaji M, Panagiotopoulos N, Reeves R, Charlett A, Pebody R. Estimating the burden of adult hospital admissions due to RSV and other respiratory pathogens in England. *Influenza Other Respir Viruses.* 2022;16(1):125–31.

43. Cong B, Dighero I, Zhang T, Chung A, Nair H, Li Y. Understanding the age spectrum of respiratory syncytial virus associated hospitalisation and mortality burden based on statistical modelling methods: a systematic analysis. *BMC Med.* 2023;21(1):224.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.