

Positive predictive value of the ICD-10 hospital diagnosis of pleural empyema in the Danish National Registry of Patients

Mette Sogaard^{1,2}
Jette Brommann Kornum²
Henrik Carl Schønheyder¹
Reimar Wernich Thomsen²

¹Department of Clinical Microbiology, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark; ²Department of Clinical Epidemiology, Institute of Clinical Medicine, Aarhus University Hospital, Aarhus, Denmark

Objective: Health care databases are a valuable source for epidemiological research in respiratory diseases if diagnoses are valid. We validated the International Classification of Diseases, 10th revision (ICD-10) diagnosis of pleural empyema in the Danish National Registry of Patients (DNRP).

Methods: We randomly selected hospitalized patients registered in the DNRP with a discharge diagnosis of pleural empyema between 1995 and 2009 in the North Denmark Region. We retrieved and reviewed medical records and estimated the positive predictive value (PPV) of the empyema diagnosis. Analyses were stratified by study period, hospital type (referral versus district), department type (pulmonary medicine or thoracic surgery versus other), cause of empyema (medical, surgical, or traumatic), and age group. To assess changes over time, we included chi-square tests for linear trend.

Results: We retrieved the medical records of 224/225 sampled patients with empyema (99.6%). Of those, 182 were classified as being definite cases, and 21 were probable cases, yielding a PPV of 90.6% (95% confidence interval [CI]: 86.0–94.1). The PPV decreased from 95.7% in patients aged 15–39 years to 87.5% in patients aged 80 years and over but was uniformly high regardless of study period, hospital or department type, or cause of empyema.

Conclusion: Our finding of a high overall PPV indicated good agreement between ICD-10 codes for pleural empyema and medical records. Registry-based discharge codes may be a suitable source of data on pleural empyema for epidemiological research.

Keywords: pneumonia, positive predictive value, validation, International Classification of Disease Codes, empyema

Introduction

Pleural empyema is a serious infection characterized by the accumulation of pus in the pleural space.¹ In the United States, 65,000 patients suffer from empyema each year, with 30-day mortality up to 15% and an estimated hospital cost of \$500 million due to prolonged hospitalizations and multiple invasive procedures.^{2–5} The empyema usually develops as a complication of pneumonia (40%–60% of all empyemas), thoracic surgery (20%), or trauma (4%–10%).⁶ Less commonly, empyema can develop as a result of esophageal rupture and subdiaphragmatic spread.

Many of the studies addressing risk factors and outcome of empyema rely on hospital discharge diagnosis codes recorded in administrative health care databases.^{7–9} Use of these databases for epidemiological research can be a cost-effective way of collecting data on empyema.^{10–12} The large size of many databases offers the potential for precise estimates of effect and for studying rare exposures or outcomes.

Correspondence: Mette Sogaard
Department of Clinical Epidemiology,
Sdr. Skovvej 15, 9000 Aalborg,
Denmark
Tel +45 99 32 69 00
Fax +45 99 32 69 14
Email mette.soegaard@rn.dk

At the same time, the risk of some biases, such as recall bias and nonresponse, is reduced. On the other hand, the researchers have no control of data selection and quality, and the usability of these databases is therefore highly reliant on the validity of the registered data.^{10–12} Thus, to draw valid inferences from database studies of pleural empyema, the validity of the diagnosis needs to be quantified.^{11,13}

To our knowledge, no study has validated pleural empyema diagnosis in hospital registries. We therefore examined the positive predictive value (PPV) of empyema discharge diagnosis in the population-based Danish National Patient Registry (DNRP) using data from medical record reviews as the reference. In addition, we determined whether the validity of the empyema diagnosis varied by calendar period, hospital or department type, cause of empyema, or patient age.

Materials and methods

Denmark has 5.3 million inhabitants, and the National Health Service provides universal tax-supported health care. Since 1968, a unique 10-digit civil registry number (CPR number) has been assigned to all Danish residents at birth or emigration by the Danish Civil Registry System.¹⁴ In all Danish registries, patients are identified by their CPR-number. We therefore linked national data on hospital diagnoses and medical records using the CPR number and identified patients diagnosed with pleural empyema between 1995 and 2009 residing in the North Denmark Region (with a population of 579,741 inhabitants or ~10% of the population of Denmark in 2009¹⁵).

The DNRP

The DNRP tracks each hospital admission in Denmark since 1977, recording dates of admission and discharge, and up to 20 discharge diagnoses. The registry covers 99.4% of all discharge records from Danish hospitals.¹⁶ Diagnoses are coded by physicians using the Danish version of the International Classification of Diseases, 8th revision (ICD-8) (1977–1993) and 10th revision (ICD-10) (1994 onward). Since 1995, outpatient contacts and visits to the emergency room have been recorded in addition to the hospital stays.

Identification of coded pleural empyema cases

Using the DNRP, we identified all discharges between 1995 and 2009 associated with a primary or secondary diagnosis of empyema (ICD-10 codes J86.0 Pyothorax with fistula and J86.9 Pyothorax without fistula). We did not evaluate the coding of tuberculous empyema (ICD-10 code A15.6) because we consider these cases to constitute a distinct disease group

that is categorized separately in the ICD-10 system; that is, under A15 “Respiratory tuberculosis, bacteriologically and histologically confirmed”.¹⁷ Because we were interested in first incident episodes of empyema, we removed all hospitalizations for any recurrent episode of empyema during 1995–2009, and patients who had been hospitalized for empyema between 1977 and 1994 (ICD-8 code 510 Empyema pleurae). We also removed pediatric patients <15 years of age.

Medical record validation

To validate the diagnoses of empyema identified in the DNRP, we randomly selected 15 patients in each calendar year using the SAS RANUNI function (SAS Institute, Cary, NC). We then used CPR numbers to identify the corresponding medical records. We confined the review to the North Denmark Region, as data quality is considered uniform in the hospital discharge registries.¹³ The records were reviewed by one of two reviewers (MS or JBK), who assessed all the available information, including written radiology reports and results from laboratory tests. The actual imaging films were not reinterpreted, and the available information was in most cases identical with the information available to the physician at the time of discharge. All cases with an uncertain diagnosis based on the available information were discussed, and diagnoses were made according to consensus.

Based on the available information, we classified the empyema cases as either definite, probable, or no evidence of empyema. Definite empyema was defined by frank pus aspirated from the pleural space, a positive Gram stain/subculture for pathogenic microorganisms in the pleural fluid, and/or an autopsy diagnosis of empyema.^{1,6,18,19} Patients with clinical symptoms suggestive of empyema in association with compatible radiographic features (eg, pleural thickening, loculated and/or septated pleural effusions^{6,20}) who did not meet the criteria for definite empyema were classified as having probable empyema. We ruled out empyema when results of diagnostic tests were negative, or if no tests were performed because symptoms had disappeared, or were explained by another disease. We further categorized the most probable cause of the patients’ empyemas as: previous surgery (patients who had an operation involving the thoracic cavity before their empyema), trauma, esophageal rupture, or any medical cause (patients who did not have thoracic surgery, trauma, or esophageal rupture and were presumed to have a medical cause for their empyema, most frequently pneumonia).

Statistical analysis

Our study outcome was the PPV of an empyema diagnosis, defined as the proportion of patients with an empyema

diagnosis in the DNRP who had definite or probable empyema according to their medical record. We calculated the PPV with 95% confidence intervals (CIs). The PPV was calculated for the whole study population and stratified by calendar period (1995–1999, 2000–2004, and 2005–2009). This stratification was done to examine whether the validity differed over time and included a chi-square test for linear trend. In addition, PPVs were calculated according to hospital type (referral versus district), department type (pulmonary medicine or thoracic surgery versus other departments), cause of empyema, and age group (15–39, 40–64, 65–79, 80 years and older).

We analyzed the data with Stata Software (v. 11.1; Stata Corp. College Station, TX). The study was approved by the Danish Data Protection Agency (Journal no. 2010-41-4969).

Results

Description of study population

From 509 adult patients in Northern Denmark registered with empyema in the DNRP between 1995 and 2009, we randomly selected 225 cases for this validation study. We excluded one patient who was transferred to a hospital outside the region with a suspected empyema. Of the remaining 224 patients with available medical records, we classified 182 (81.3%) as being definite empyema cases. Twenty-one patients did not meet the criteria for definite empyema and were classified as probable cases. Among the remaining 21 patients whose empyema diagnosis was rejected, 19 patients had other pulmonary disorders; one had a pulmonary abscess, one had pulmonary tuberculosis, one had sarcoidosis, two had emphysema, eight had pneumonia, one had chylothorax, one had multiple rib fractures, one had chronic sequelae after previous pneumonia, and three had pulmonary cancer. The two remaining patients had severe muscle dystrophia and surgical site infection, respectively. Table 1 shows demographic and diagnostic characteristics of patients registered in the DNRP according to whether they had definite, probable, or no empyema. Characteristics of pleural fluid were left undocumented in the medical record in ten patients classified as having definite or probable empyema, three of which died shortly after admission. Among the 191 patients with definite or probable empyema who had their pleural fluid microbiologically evaluated, the proportion of positive samples was 61.8% (118/191). Seven of the 12 definite or probable empyema cases without microbiological evaluation died shortly after admission.

Validation of empyema diagnosis

The overall PPV including definite and probable empyema hospitalizations was 90.6% (95% CI: 86.0–94.1) and 81.3%

Table 1 Descriptive and diagnostic characteristics of the 224 patients with pleural empyema registered in the Danish National Registry of Patients, 1995–2009

Characteristic	Diagnosis of empyema		
	Definite, n (%)	Probable, n (%)	Rejected, n (%)
Total	182 (81.3)	21 (9.4)	21 (9.4)
Age group			
15–39	19 (10.4)	3 (14.3)	1 (4.8)
40–64	76 (41.8)	11 (52.4)	8 (38.1)
65–79	69 (37.9)	4 (19.1)	9 (42.9)
80 and older	18 (9.9)	3 (14.3)	3 (14.3)
Sex			
Male	125 (68.7)	16 (76.2)	14 (66.7)
Female	57 (31.1)	5 (23.8)	7 (33.3)
Chronic obstructive pulmonary disease	22 (12.1)	0	5 (23.8)
Bronchial asthma	5 (2.8)	1 (4.8)	4 (19.1)
Cause of empyema			
Medical	144 (79.1)	19 (90.5)	16 (76.2)
Surgical	24 (13.2)	0	4 (19.1)
Esophageal rupture	3 (1.7)	2 (9.5)	0
Trauma	11 (6.0)	0	1 (4.8)
Diagnostic criteria			
Autopsy diagnosis of empyema	4 (2.2)	–	–
Pus in the pleural space			
Yes	171 (94.0)	–	–
No	5 (2.8)	17 (81.0)	18 (85.7)
No information	6 (3.3)	4 (19.1)	3 (14.3)
Gram stain or culture of pleural fluid			
Positive	118 (64.8)	–	–
Negative	57 (31.3)	16 (76.2)	13 (61.9)
No microbiological evaluation	7 (3.9)	5 (23.8)	8 (38.1)
Radiographic features of empyema			
Yes	174 (95.6)	21 (100.0)	4 (19.1)
No	5 (2.8)	0	15 (71.4)
No information	3 (1.7)	0	2 (9.5)

(95% CI: 81.3–86.1) when focusing only on definite cases. PPVs of empyema diagnosed in 1995–1999, 2000–2004, and 2005–2009 were 90.7% (95% CI: 81.7–96.2), 94.6% (95% CI: 86.7–98.5), and 86.7% (95% CI: 76.8–93.4), respectively, indicating that no major changes in empyema coding validity occurred over the 15-year study period (P -value for trend = 0.40). The PPVs in different strata of calendar period, hospital type, department type, and cause of empyema showed only slight differences from the overall PPV. However, the PPV appeared to decrease with increasing age; that is, from 95.7% (95% CI: 78.1–99.9) in patients aged 15–39 years to 87.5% (95% CI: 67.6–97.3) in patients aged 80 years and over (P -value for trend < 0.001) (Table 2).

Discussion

Our finding of a high overall PPV indicated good agreement between ICD-10 codes for pleural empyema and medical records. In addition, we detected little variation in the PPV by calendar period, hospital type, department type, or cause. However, PPV was slightly higher in younger age groups than in elderly patients.

Although we examined only a segment of the Danish national population, the results can be assumed to be generalizable, thanks to the uniform structure of record keeping and the universal nature of the health care system across Denmark. However, a number of limitations need to be considered in the interpretation of the results. First, we only examined one dimension of the validity: the predictive value of a positive registration of empyema. The optimal validation study would also examine sensitivity and specificity,¹⁰ but since no other data source with empyema exists, only the predictive value could be estimated. Second, most of the medical records were evaluated by one reviewer. However, any cases with an uncertain diagnosis were evaluated by both reviewers independently. Thereafter, final diagnoses were made according to consensus. The reviewers were not blinded to the registered discharge diagnosis codes, and blinding is usually recommended in diagnostic studies.²¹ However, since the result of the “test under study” (the register diagnosis) was by definition given for all the examined patients (as only patients who had a diagnosis were included), it is unlikely that lack of blinding of the reviewers would have influenced the

PPV in general. Third, pleural effusions represent a spectrum of inflammatory fluid collections that ranges from transudative effusion to empyema, and the differentiation of this spectrum into patients with empyema and those without is somewhat arbitrary.^{1,6} Studies have shown that neither radiography, ultrasound, nor computed tomography can reliably identify the stage of pleural disease in adult patients with effusions and empyema.^{22,23} Although we carefully evaluated the reported clinical and radiographic findings, we may therefore have overestimated the PPV among patients with probable empyema. Finally, we do not know to what extent our findings can be extrapolated to other places, where the prevalence of empyema and the diagnostic strategy may be different.

Whether the data quality documented in our study is sufficient for registry-based studies depends on the proposed research questions and the study design used.¹⁰ If databases such as the DNRP are used to assess changes in incidence of empyema over time, the PPV must remain stable over time to obtain valid estimates.²⁴ We found a slight variation in PPV by study period but there was not a significant linear trend over the 15-year period of our study. Thus, the increasing availability of advanced diagnostic tools, including computed tomography and magnetic resonance imaging, between 1995 and 2009 apparently did not improve the quality of diagnostic coding. We do not know whether there might have been any substantial time trends in the data quality before 1995. In analytical studies, inaccurate or missing data tend to bias the association toward the null hypothesis rather than to cause spurious associations,

Table 2 Validity of pleural empyema diagnosis

	Total No. of cases	No. definite cases	No. probable cases	PPV (%)	95% CI
All	224	182	21	90.6	86.0–94.1
Period					
1995–1999	75	63	5	90.7	81.7–96.2
2000–2004	74	65	5	94.6	86.7–98.5
2005–2009	75	54	11	86.7	76.8–93.4
Hospital type					
Referral hospital	139	117	10	91.4	85.4–95.5
District hospital	85	65	11	89.4	80.8–95.0
Department type					
Pulmonary medicine or thoracic surgery	101	85	7	91.1	83.8–95.8
Other	123	97	14	90.2	83.6–94.9
Cause of empyema					
Medical	179	144	19	91.1	85.9–94.8
Surgical incl. esophageal rupture	33	27	2	87.9	77.8–96.6
Trauma	12	11	0	91.7	61.5–99.8
Age group (years)					
15–39	23	19	3	95.7	78.1–99.9
40–64	95	76	11	91.6	84.1–96.3
65–79	82	69	4	89.2	80.2–94.9
80 and older	24	18	3	87.5	67.6–97.3

Abbreviations: PPV, positive predictive value; CI, confidence interval.

as long as the sensitivity and specificity are similar in the compared groups.¹⁰ At the same time, if empyema is used as the outcome of interest, imperfect sensitivity, no matter how low, is not expected to bias relative measures of association as long as the specificity is 100%.²⁵ We have shown that empyema cases that are registered largely represent true instances of this event. If the misclassification of empyema is associated with the examined exposure of interest, however, the risk estimates may be biased, both toward and away from the null, depending on the direction of the association.

Denmark and a few other places in the world are privileged by having a broad network of population-based registries,^{9,14,18,26} which enables use of routinely collected data to rapidly assemble large cohorts for epidemiological studies, including that of pleural empyema. The high PPV estimated in this study indicates a low false-positive number of empyema cases in the DNRP, suggesting that the DNRP is a valid source for identifying cases of empyema for epidemiological research.

Acknowledgments

We are very grateful to Elisabeth Kristoffersen who helped to retrieve and systemize the medical records. We also thank statistician Rikke Beck Nielsen for sampling the DNRP.

Disclosure

The authors report no conflicts of interest in this work.

References

- Light RW. A new classification of parapneumonic effusions and empyema. *Chest*. 1995;108(2):299–301.
- Sahn SA. Management of complicated parapneumonic effusions. *Am Rev Respir Dis*. 1993;148(3):813–817.
- Davies CW, Kearney SE, Gleeson FV, Davies RJ. Predictors of outcome and long-term survival in patients with pleural infection. *Am J Respir Crit Care Med*. 1999;160(5 Pt 1):1682–1687.
- Ferguson AD, Prescott RJ, Selkon JB, Watson D, Swinburn CR. The clinical course and management of thoracic empyema. *QJM*. 1996;89(4):285–289.
- Cham CW, Haq SM, Rahamim J. Empyema thoracis: a problem with late referral? *Thorax*. 1993;48(9):925–927.
- Septimus E. Pleural effusion and empyema. In: Mandell GL, Bennett JE, Dolin R, editors. *Principles and Practice of Infectious Diseases*. 6th ed. Churchill, Livingstone: Elsevier Inc.; 2005:845–853.
- Finley C, Clifton J, Fitzgerald JM, Yee J. Empyema: an increasing concern in Canada. *Can Respir J*. 2008;15(2):85–89.
- Li ST, Tancredi DJ. Empyema hospitalizations increased in US children despite pneumococcal conjugate vaccine. *Pediatrics*. 2010;125(1):26–33.
- Grijalva CG, Nuorti JP, Zhu Y, Griffin MR. Increasing incidence of empyema complicating childhood community-acquired pneumonia in the United States. *Clin Infect Dis*. 2010;50(6):805–813.
- Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol*. 1996;25(2):435–442.
- Schneeeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58(4):323–337.
- Baron JA, Weiderpass E. An introduction to epidemiological research with medical databases. *Ann Epidemiol*. 2000;10(4):200–204.
- Nickelsen TN. Data validity and coverage in the Danish National Health Register. *Ugeskr Laeger*. 2002;164(01):33–37.
- Frank L. Epidemiology. When an entire country is a cohort. *Science*. 2000;287(5462):2398–2399.
- Statistics Denmark, StatBank Denmark, Population and Electronics. Available from: <http://www.statistikbanken.dk/statbank5a/default.asp?w=1280>. Accessed September 15, 2010.
- Andersen TF, Madsen M, Jørgensen J, Møller-Jensen L, Olsen JH. The Danish National Hospital Register. A valuable source of data for modern health sciences. *Dan Med Bull*. 1999;46(3):263–268.
- Malhotra P, Aggarwal AN, Agarwal R, Ray P, Gupta D, Jindal SK. Clinical characteristics and outcomes of empyema thoracis in 117 patients: a comparative analysis of tuberculous vs. non-tuberculous aetiologies. *Respir Med*. 2007;101(3):423–430.
- Ahmed RA, Marrie TJ, Huang JQ. Thoracic empyema in patients with community-acquired pneumonia. *Am J Med*. 2006;119(10):877–883.
- Chalmers JD, Singanayagam A, Murray MP, Scally C, Fawzi A, Hill AT. Risk factors for complicated parapneumonic effusion and empyema on presentation to hospital with community-acquired pneumonia. *Thorax*. 2009;64(7):592–597.
- Evans AL, Gleeson FV. Radiology in pleural disease: state of the art. *Respirology*. 2004;9(3):300–312.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ*. 2003;326(7379):41–44.
- Donnelly LF, Klosterman LA. CT appearance of parapneumonic effusions in children: findings are not specific for empyema. *AJR Am J Roentgenol*. 1997;169(1):179–182.
- Kearney SE, Davies CW, Davies RJ, Gleeson FV. Computed tomography and ultrasound in parapneumonic effusions and empyema. *Clin Radiol*. 2000;55(7):542–547.
- Schneeeweiss S. Understanding secondary databases: a commentary on “Sources of bias for health state characteristics in secondary databases”. *J Clin Epidemiol*. 2007;60(7):648–650.
- Brenner H, Savitz DA. The effects of sensitivity and specificity of case selection on validity, sample size, precision, and power in hospital-based case-control studies. *Am J Epidemiol*. 1990;132(1):181–192.
- Koshy E, Murray J, Bottle A, Sharland M, Saxena S. Impact of the seven-valent pneumococcal conjugate vaccination (PCV7) programme on childhood hospital admissions for bacterial pneumonia and empyema in England: national time-trends study, 1997–2008. *Thorax*. 2010;5(9):770–774.

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic

Submit your manuscript here: <http://www.dovepress.com/clinical-epidemiology-journal>

Dovepress

reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.