ICD-10 codes are a valid tool for identification of pneumonia in hospitalized patients aged ⩾65 years

S. A. SKULL[1,2*], R. M. ANDREWS[2], G. B. BYRNES[3], D. A. CAMPBELL[3,4],
T. M. NOLAN[3], G. V. BROWN[5] AND H. A. KELLY[3,6]

[1] *Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Parkville, Victoria, Australia*
[2] *Menzies School of Health Research, Darwin, Northern Territory, Australia*
[3] *School of Population Health, University of Melbourne, Victoria, Australia*
[4] *Monash Institute of Health Services Research, Monash Medical Centre, Clayton, Victoria, Australia*
[5] *Department of Medicine, University of Melbourne, Royal Melbourne Hospital, Parkville, Victoria, Australia*
[6] *Victorian Infectious Diseases Reference Laboratory, North Melbourne, Victoria, Australia*

SUMMARY

This study examines the validity of using ICD-10 codes to identify hospitalized pneumonia cases. Using a case-cohort design, subjects were randomly selected from monthly cohorts of patients aged ⩾65 years discharged from April 2000 to March 2002 from two large tertiary Australian hospitals. Cases had ICD-10-AM codes J10–J18 (pneumonia); the cohort sample was randomly selected from all discharges, frequency matched to cases by month. Codes were validated against three comparators: medical record notation of pneumonia, chest radiograph (CXR) report and both. Notation of pneumonia was determined for 5098/5101 eligible patients, and CXR reports reviewed for 3349/3464 (97%) patients with a CXR. Coding performed best against notation of pneumonia: kappa 0·95, sensitivity 97·8% (95% CI 97·1–98·3), specificity 96·9% (95% CI 96·2–97·5), positive predictive value (PPV) 96·2% (95% CI 95·4–97·0) and negative predictive value (NPV) 98·2% (95% CI 97·6–98·6). When medical record notation of pneumonia is used as the standard, ICD-10 codes are a valid method for retrospective ascertainment of hospitalized pneumonia cases and appear superior to use of complexes of symptoms and signs, or radiology reports.

INTRODUCTION

All-cause pneumonia is an important clinical endpoint for determining vaccine effectiveness (VE) for 23-valent pneumococcal polysaccharide vaccine (23vPPV) and influenza vaccine. It represents a greater part of the burden of disease due to these organisms, and a trade-off between a highly specific but insensitive outcome measure such as pneumonia associated with pneumococcal bacteraemia or microbiologically proven influenza and a sensitive, but non-specific surrogate outcome such as all-cause mortality. However, accurate identification of pneumonia cases is not straightforward. Clinical criteria alone are imprecise [1]. There remains no internationally agreed definition for pneumonia based on clinical symptoms and signs and no one sign or symptom, nor combination of these has ever been shown to clearly differentiate pneumonia from other respiratory illnesses [2–4]. There is also no ideal diagnostic test for microbiological diagnosis [1], and while chest radiograph (CXR) is often useful to confirm the diagnosis of pneumonia and its severity [1], it also has limitations [5, 6]. Because of the difficulties

* Author for correspondence: Dr S. A. Skull, Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Flemington Rd, Parkville, Victoria, Australia 3052.
(Email: saskull@unimelb.edu.au)

in defining and identifying cases of pneumonia using clinical, radiological or microbiological criteria, codes from the International Classification of Diseases (ICD), overseen by the World Health Organization (WHO), are frequently used as surrogate measures to identify hospitalized patients with pneumonia in studies of VE for 23vPPV and influenza vaccine [7–11]. These codes have become the international standard for disease classification [12].

Use of standardized codes to retrospectively identify cases of pneumonia among hospitalized patients is appealing to researchers primarily because of time efficiencies (compared with the alternative of reviewing hospital records for clinical, radiological and/or laboratory evidence consistent with pneumonia). Despite the practical advantages and continued use of ICD codes by researchers to identify cases of pneumonia, at the time of this research only two small studies (<150 subjects) have examined the validity of this approach for all-cause pneumonia [13, 14], while a third has examined codes for pneumococcal pneumonia [15]. These studies, all from North America, used ICD-9-CM codes and suggested ICD codes may be a valid tool for case ascertainment of pneumonia. However, further examination is prudent given the paucity of available data and the potential for differences in other settings. As part of a case-cohort study [16, 17] examining VE for 23vPPV and influenza vaccine against pneumonia in the elderly in Australia where these vaccines are provided free of charge, we examined the validity of ICD-10-AM codes to identify cases of pneumonia among hospitalized patients.

METHODS

Study subjects

Cases of pneumonia were identified from monthly separation lists of completed admissions for patients aged ⩾65 years from two major teaching hospitals (Royal Melbourne Hospital and Western Hospital Footscray) for the period 1 April 2000 to 31 March 2002, using ICD-10-AM codes J10–J18 (pneumonia including those cases due to influenza) [18]. These two hospitals represent about 11% of the total hospitalized population and 13% of all hospitalizations for pneumonia for those aged ⩾65 years in Victoria [19]. Eight coders at each hospital (each with a three-year university degree in Health Information Management) assigned codes as per Australian standards [20]. Pneumonia was identified if one or more of these codes

appeared in any of the 14 diagnostic code positions for each hospital separation. Cases were also eligible for selection in the cohort as a case-cohort design. If a subject appeared on the hospital separation list more than once in any given month, one episode was selected at random, and the rest excluded from analysis. For month-to-month repeat separations for pneumonia for an individual, the first selected admission was retained and subsequent episodes excluded to minimize any Hawthorne effect from study participation affecting vaccination status [21]. Patients were excluded from monthly separation lists if not resident in Victoria or if admitted for short-stay procedures such as dialysis and chemotherapy (ICD-10-AM codes Z49.1, Z49.2 and Z51.1). Cohort subjects were randomly selected from monthly separation lists, frequency-matched to the cases. Over-sampling was conducted to allow for subsequent exclusion of repeat admissions in the cohort, as well as those subjects also selected as cases. After exclusions, a total of 1·2 times the number of cases was selected using a random number generator. A cohort subject could be selected only once each month and admissions selected for the same subject in subsequent months were excluded.

Examination of the validity of codes for subgroups of microbiologically proven pneumococcal pneumonia (J13) or pneumonia associated with proven influenza (J10 and J11) was not conducted due to small numbers in these subgroups (*S. pneumoniae* pneumonia was coded in only 11 first-presentation pneumonia cases, and there were none coded as influenza pneumonia).

Development of comparators for pneumonia ICD-10 codes

Given the difficulty of defining a reference standard for the diagnosis of pneumonia, three comparators were developed for the purpose of examining the validity of ICD-10 coded cases using retrospective chart review: (1) medical record notation of 'pneumonia', (2) CXR report and (3) both, since interpretation of both clinical and radiological findings is generally used in clinical practice to make a definitive diagnosis of pneumonia. As ICD-10 codes were not integrated with the database until after completion of the study, record review occurred blinded to coding status.

Hospital records for the selected admission were reviewed for notation of 'pneumonia' as a diagnosis

considered probable by the clinical team under whose care the patient was admitted. This notation was considered most likely to be consistent with a diagnosis of pneumonia given it would be based on all information available to the clinical team at the time of discharge (thus having both face and content validity). To examine relevance of signs and symptoms in retrospective identification of pneumonia, for patients with notation of pneumonia, the documentation of cough, sputum production, pleuritic chest pain, fever ⩾37·5 °C, shortness of breath, crackles (crepitations), and aspiration was also sought (definitely present, definitely absent or not recorded); these being the most common symptoms and signs of pneumonia suggested by descriptive studies of pneumonia [3, 22].

Two trained research assistants used pre-specified criteria to interpret radiologists' reports for all study subjects with CXRs undertaken as part of routine management. Pneumonia was defined as 'lobar' (any opacity confined to a lobar anatomical distribution), 'bronchopneumonia' (opacity distributed beyond a single lobe in conjunction with terms that are similar to or include the words 'patchy' and/or 'airspace'), 'other' (opacity consistent with pneumonia not previously classified) and 'not pneumonia' (none of the above).

For patients with more than one CXR during their selected admission, the first abnormal report was reviewed blinded to other reports. When reports could not be confidently interpreted, two of the investigators (S.S., a paediatrician, and D.C., an adult respiratory physician) made the final assessment. High inter-operator agreement was first established on a sample of pilot subjects (the study was piloted in full for 1 month prior to study commencement). We independently reviewed CXR reports for consecutive groups of 20 pilot subjects and compared interpretation for agreement. No further groups were examined and review of study reports did not commence until a kappa statistic [23] indicating >95% concurrence was achieved. Consensus was obtained on subsequent 'in dispute' reports after independent review and before data entry. For the main analysis, outcomes were categorized as 'consistent with pneumonia' (lobar pneumonia, bronchopneumonia or other pneumonia) or not.

Statistical methods

The validity of ICD-10 codes for identification of cases of pneumonia in hospitalized patients was examined by comparing codes J10–J18 as a group *vs.* the three comparators. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated using STATA version 9.1 [24]. In addition, because of the absence of a true reference standard, raw data are presented with percentage agreement between ICD-10 codes and the comparators and kappa statistics (agreement adjusted for chance agreement) were calculated. The effect of hospital of separation and season on coding validity was examined using stratification. The influenza season, defined by influenza surveillance independent of this study [25] was used as a proxy for a period of increased pneumonia activity.

To determine the extent of any effect of repeat admission for pneumonia on coding practices, analyses were repeated with inclusion of all selected subjects.

The study was approved by the Human Research Ethics Committee, Melbourne Health (ref. 2000.022). Free and informed consent was obtained from subjects or their legal guardians.

RESULTS

Study subjects

There were 2319 first presentations coded as pneumonia and 2912 first-presentation cohort subjects, including 130 who were also selected as cases, giving a total of 5101 eligible study subjects (Figures 1 and 2). The mean age of eligible subjects was 77 years and 2740 (54%) were male. CXRs were conducted for 3464/5101 (68%) subjects (96% of cases and 47% of cohort subjects), and of these 3349 (97%) (97% of cases and 97% of cohort subjects) had radiology reports available for review.

Validity using medical record notation of pneumonia as the comparator

Clinical notation of pneumonia (yes/no) was able to be determined for 5098/5101 subjects (99·9%). Of these, 2281 (45%) had pneumonia documented as a probable diagnosis, representing 2230/2318 (96%) ICD-10-coded cases and 51/2780 (2%) ICD-10-coded non-cases. Among cohort subjects, 128/179 (72%) pneumonia notations had pneumonia defined by ICD-10 codes. There was a very high level of agreement between ICD-10-coded pneumonia or non-pneumonia and clinical notation, with a kappa statistic of 0·95 and high sensitivity, specificity, PPV
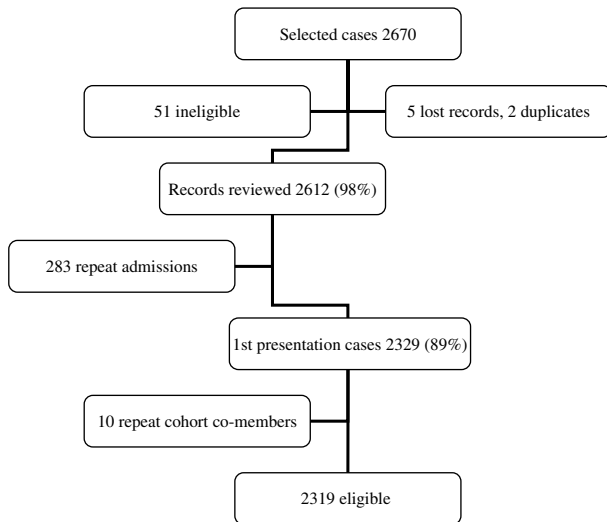
Fig. 1. Flow chart of eligible first-presentation cases, based on ICD-10 codes J10–J18.
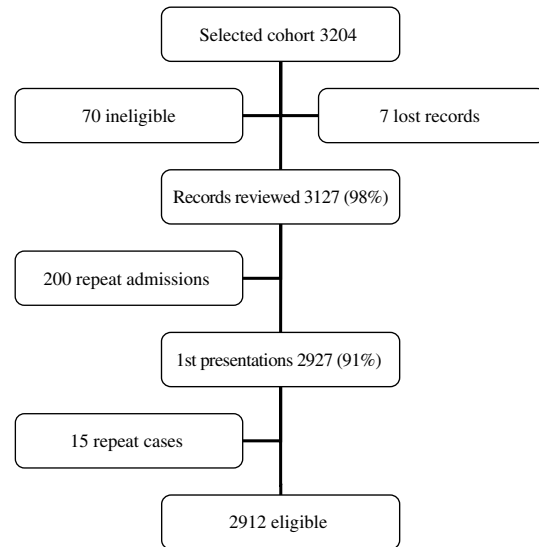

Fig. 2. Flow chart of eligible first-presentation cohort subjects.

and NPV (Table 1). Stratification by season and hospital of selection indicated these factors did not play an important role, with small differences between strata in real terms (range 0·1–5·5%).

Among the 2281 subjects with notation of pneumonia, a median of four (range 0–6) of the seven symptoms and signs of interest were present. Three or more were present in 1911/2281 (84%). The symptoms and signs that were most frequently recorded as present or absent and that were present most often were crackles (92%), shortness of breath (74%), cough (71%), fever $\geqslant 37\cdot5\ ^{\circ}\mathrm{C}$ (66%) and sputum production (54%) (Table 2).

Validity using CXR as the comparator

Of 5101 eligible subjects, 3464 (68%) had had a CXR conducted, representing 2239/2329 (96%) subjects with ICD codes for pneumonia and 1374/2927 (47%) subjects not coded as having pneumonia. Eighty-seven of 5101 (1·7%) subjects had no CXR performed and an ICD-coded diagnosis of pneumonia. In total, 3345/3464 (97%) subjects with a CXR had radiology reports available for review and 1724/3345 (51%) with CXR reports had some form of pneumonia based on review (bronchopneumonia 24·8%, lobar pneumonia 24·8%, other pneumonia 0·2%, not pneumonia 46·8%, investigator unsure 0·1%). This represented 1538/2154 (71%) ICD-10-coded cases and 186/1191 (16%) ICD-10-coded non-cases with a report. A good level of agreement was present between pneumonia status according to ICD-10 codes

and CXR report (kappa 0·52) (Table 1). No difference in estimates was found when stratifying by season, and only one difference in strata-specific estimates for NPV when stratifying by hospital suggesting a true difference ($-5\cdot5$, 95% CI $-9\cdot8$ to $-1\cdot3$).

Validity using CXR plus medical record notation of pneumonia as the comparator

The level of agreement was similar to that of CXR report alone (kappa 0·60) (Table 1). Indicators of validity were within the range provided by the previous two comparators, except for PPV which was lower (Table 1). Stratification indicated no effect of season or hospital of separation on estimates (data not shown). Estimates for validity also changed very little when all cases of pneumonia were included rather than just first presentations (data not shown).

ICD-10 codes and diagnostic positions used for pneumonia

The most common ICD-10 codes used and the diagnostic positions (1–14) in which they occurred during the study period for the 2319 eligible first-presentation cases of pneumonia are shown in Table 3. Eight subjects (0·3%) had two codes for pneumonia assigned. By far the most common ICD-10 code used for cases of pneumonia was J18.9 (pneumonia, unspecified) which comprised 91·5% (2122/2319) of all first cases of pneumonia. The next most common codes were J18.0

Table 1. *Validity of ICD-10 coding* vs. *three comparators*

Comparator	Cases (J10–J18)	Non-cases (other codes)	Kappa	Sensitivity (95 % CI)	Specificity (95 % CI)	PPV (95 % CI)	NPV (95 % CI)
CXR report (*n* = 3345)	1538/2154 71 %	1005/1191 84 %	0·52	89·2 (87·7–90·6)	62·0 (59·6–64·4)	71·4 (69·4–73·3)	84·4 (82·2–86·4)
Notation of pneumonia (*n* = 5098)	2230/2318 96 %	2729/2780 98 %	0·95	97·8 (97·1–98·3)	96·9 (96·2–97·5)	96·2 (95·4–97·0)	98·2 (97·6–98·6)
Both (*n* = 3343)	1509/2153 70 %	34/1190 2·9 %	0·60	97·8 (96·9–98·5)	64·2 (62·0–66·4)	68·1 (68·1–72·0)	97·1 (96·1–98·1)

CI, Confidence interval; PPV, positive predictive value; NPV, negative predictive value.

(bronchopneumonia, unspecified): 1·6 % (37/2319) and J15.1 (pneumonia due to *Pseudomonas*): 1·4 % (32/2319). Of first-presentation cases of pneumonia with codes J10–J18 listed, codes for pneumonia occurred most frequently in diagnostic position 1 (50·8 %). A total of 82 % of cases were documented in the first four positions and 95 % in the first eight positions.

DISCUSSION

These data confirm the validity of ICD-10 codes for the retrospective identification of persons discharged from hospital with a diagnosis of pneumonia. Using medical record notation of pneumonia as the comparator, we were able to exclude estimates for sensitivity, specificity, PPV and NPV of less than 95 %. Given that coding staff are trained to translate hospital record notations into codes in a way that captures as much information as possible [20], rather than by searching for individual symptoms and signs, these data confirm that the coding process is being performed at a high standard in the two hospitals studied.

This study found somewhat higher levels of internal validity for ICD-10 coding as a tool for identifying persons discharged from hospital with pneumonia than previous studies conducted outside Australia using ICD-9 codes [13–15]. The differences in results may be explained by differences in design (see below) or setting (e.g., coding practices or training). In Victoria, for instance, a high level of training is required for clinical coders, there is linking of hospital funding to codes, and annual audits of coding accuracy are conducted by independent and/or government agencies. In general, results from earlier studies were nonetheless favourable towards use of ICD codes as a diagnostic tool.

Marrie and colleagues examined ICD-9-CM codes (011.6, 021.2, 136.3, 480–487, 506–507) in a prospective study of 105 adult patients hospitalized with pneumonia [13, 26]. Codes 480–487 correspond to ICD-10-AM codes J10–J18 [18]. The comparator utilized was clinical pneumonia diagnosed within 48 h of admission by medical staff, plus a new opacity on CXR consistent with pneumonia confirmed by the researchers. The study estimated sensitivity of 69 % and PPV of 57 % for these ICD codes as a group. Another small study of agreement compared 144 ICD-9-CM classified cases of community-acquired pneumonia (CAP) (codes 480–487 plus 13 other codes that might capture pneumonia [27]) with a reference standard for pneumonia using retrospective review of clinical records and CXR reports [14]. Confirmation of CAP by clinical review required symptoms compatible with pneumonia within 24 h of admission and a report consistent with pneumonia from a CXR within 48 h of admission. Where the diagnostic code for CAP was in the principal diagnosis position, compared with review of clinical records, codes had a sensitivity of 84 %, specificity of 86 %, PPV of 92 % and kappa of 0·68. A further study by Guevara and colleagues is not directly comparable as the investigators examined the validity of ICD-9-CM codes for the subcategory of pneumococcal pneumonia against various clinical definitions [15]. Inclusion criteria for the analysis of CAP requiring hospitalization included age ⩾18 years, CXR within 48 h of admission consistent with pneumonia in a patient with any one of fever, abnormal white blood cell count, hypothermia or productive cough. With removal of the narrowest of the six diagnostic coding groups (code for pneumococcal septicaemia only: 38·20), ranges for a combination of codes indicative of pneumococcal pneumonia were sensitivity (55–85 %) and NPV (93–95 %) [15]. With removal of the

Table 2. *Frequency of symptoms and signs extracted from medical records for subjects with notation of pneumonia as a probable diagnosis (n = 2281)*

Symptom or sign	Present n (%)	Absent n (%)	Not recorded n (%)
Crackles	2110 (92·5)	138 (6·0)	33 (1·5)
Shortness of breath	1680 (73·6)	496 (21·7)	105 (4·6)
Cough	1623 (71·1)	437 (19·2)	221 (9·7)
Fever ⩾37·5 °C	1508 (66·1)	764 (33·5)	9 (0·4)
Sputum production	1229 (53·9)	757 (33·2)	295 (12·9)
Pleuritic chest pain	508 (22·3)	915 (40·1)	858 (37·6)
Evidence of aspiration	154 (6·7)	469 (20·6)	1658 (72·7)

broadest of the six diagnostic coding groups (all six evaluated codes: 38.20, 481.00, 38.00, 482.30, 518.81, 486.00), the range for specificity and PPV was 96–100% and 72–95% respectively. A recent study conducted since completion of our study confirms estimates for validity in the same range as the studies by Marrie *et al.* and Guevara *et al.* [28]. Aronsky and colleagues compared ICD-9 codes 480–483 plus 485–487 with a reference standard requiring: a CXR report compatible with pneumonia, an ICD-9 code for or discharge diagnosis of pneumonia, at least a 1% probability of pneumonia calculated by a decision support system [29], notation of 'pneumonia' in the medical notes and a consensus vote of pneumonia as the diagnosis by three independent physicians. Estimates for validity were: sensitivity 55% (95% CI 48–61), specificity 99% (95% CI 99–99), PPV 84% (95% CI 77–90) and NPV 96% (95% CI 95–97).

Our choice of ICD-10-AM codes J10–J18 to identify cases of hospitalization of pneumonia is consistent with previous studies examining VE of influenza vaccine and 23vPPV against pneumonia [7, 8, 10, 11, 30]. Most researchers have utilized ICD-9 codes 480–487, equivalent to ICD-10-AM codes J10–J18 [7, 10, 11, 30]. Two of the previous studies examining validity of ICD-9 codes used a more inclusive set of codes [13, 14] which may also partially explain their lower levels of estimated validity.

Although this study did not examine individual signs and symptoms consistent with pneumonia for all participants, previous studies suggest that symptom complexes are likely to be inferior to ICD-10 codes as a tool for researchers to retrospectively identify cases of pneumonia [2–4]. In our study, review of hospital records for subjects with notation of pneumonia found 84% had at least three of the seven symptoms and signs of interest.

It may not be surprising that using radiology reports as a reference standard to define pneumonia retrospectively did not result in close agreement with ICD-10 codes. First, non-specific language was often used. Words such as 'opacity' were frequently used to describe the appearance of a CXR rather than reporting a definitive diagnosis, and may indicate pathology other than pneumonia. We did not attempt to review chest radiographs themselves. While it is possible that radiologist review of CXRs (rather than of their associated reports) may be of greater diagnostic value, limited data suggest this is also imperfect. One study of 282 patients with pneumonia confirmed by a radiologist found that the agreement rate by a further two radiologists was only 79% [5]. A standardized approach to the interpretation of adult CXRs is not yet available, however, future developments may improve the usefulness of radiology reports in reference standards for pneumonia for future studies. Although a standardized approach to interpretation of paediatric CXRs has been developed, this has not yet been correlated with clinical disease and is only valid for prospective studies following specific training of reviewers [31].

Code J18.9 for 'unspecified pneumonia' comprised over 91% of all hospital separations for pneumonia. Therefore while ICD-10 codes are both sensitive and specific for the identification of all-cause pneumonia, they are unlikely to be helpful, at least in this setting, for the identification of subcategories of pneumonia.

A key limitation in this area of research is the lack of a reference standard for diagnosis of pneumonia against which to compare ICD-10 codes. However, analyses were conducted using three comparators suggested by review of the literature and this study was large enough to exclude a sensitivity, specificity, NPV and PPV for ICD-10 codes for pneumonia of

Table 3. *ICD-10 codes used to describe pneumonia by diagnostic code position (1–14)*

Code	Description	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total	%
J10.0	Influenza with pneumonia, influenza virus identified	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0·0
J11.0	Influenza with pneumonia, virus not identified	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0·0
J11.1	Influenza with other respiratory manifestations, virus not identified	13	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2	0·0
J12.9	Viral pneumonia, unspecified	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0·0
J13	Pneumonia due to *Streptococcus pneumoniae*	7	8	2	0	0	0	0	1	0	0	0	0	0	0	0	11	0·5
J14	Pneumonia due to *Haemophilus influenzae*	5	10	4	1	0	2	0	0	0	0	0	0	0	0	0	17	0·7
J15.0	Pneumonia due to *Klebsiella pneumoniae*	10	0	1	2	1	0	0	0	1	2	1	0	0	0	0	8	0·3
J15.1	Pneumonia due to *Pseudomonas*	3	9	12	2	2	2	2	1	0	0	2	0	0	0	0	32	1·4
J15.2	Pneumonia due to staphylococcus	4	8	6	7	1	0	1	1	0	1	0	3	1	0	0	29	1·3
J15.3	Pneumonia due to streptococcus, group B	8	10	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0·4
J15.4	Pneumonia due to other streptococci	10	3	2	1	0	1	0	0	0	0	0	0	1	0	0	8	0·3
J15.5	Pneumonia due to *Escherichia coli*	13	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2	0·1
J15.6	Pneumonia due to other aerobic Gram-negative bacteria	9	5	1	0	0	0	1	0	0	0	1	0	1	0	0	9	0·4
J15.7	Pneumonia due to *Mycoplasma pneumoniae*	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0·0
J15.8	Other bacterial pneumonia	12	1	0	0	0	0	1	0	0	0	0	0	1	0	0	3	0·1
J15.9	Bacterial pneumonia, unspecified	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0·0
J16.8	Pneumonia due to other specified infectious organisms	12	0	0	0	0	1	1	0	0	1	0	0	0	0	0	3	0·1
J17.0	Pneumonia in bacterial diseases classified elsewhere	9	0	5	1	0	2	0	0	0	0	0	1	0	0	0	9	0·4
J17.2	Pneumonia in mycoses	11	0	1	0	2	0	1	0	0	0	0	0	0	0	0	4	0·2
J17.3	Pneumonia in parasitic diseases	14	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0·0
J17.8	Pneumonia in diseases classified elsewhere	13	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0·1
J18.0	Bronchopneumonia, unspecified	2	20	5	3	1	2	0	2	0	0	1	0	1	1	1	37	1·6
J18.1	Lobar pneumonia, unspecified	6	8	2	1	0	0	0	0	1	0	0	0	0	0	0	12	0·5
J18.8	Other pneumonia, organisms unspecified	11	2	1	0	0	0	0	0	1	0	0	0	0	0	0	4	0·2
J18.9	Pneumonia, unspecified	1	1096	387	165	105	83	74	60	43	29	18	27	15	11	9	2122	91·5
Total		—	1182	432	183	114	95	81	65	46	33	23	31	20	12	10	2327	—
%			50·8	18·6	7·9	4·9	4·1	3·5	2·8	2·0	1·4	1·0	1·3	0·9	0·5	0·4		

less than 95% when compared with medical record notation of pneumonia. Kappa statistics for agreement were very high. There were few missing data for any comparator, with 97% of radiology reports available for eligible subjects, and notation of pneumonia able to be determined for all but three subjects. Selection bias was minimized by random selection of the cohort, frequency sampling by month, and exclusion of non-Victorian residents. Measurement bias was reduced by blinding data collectors to ICD-coded case status, rigorous training and monitoring, and piloting of the study. Estimates made using all episodes of pneumonia were virtually identical to those made using only first presentations, suggesting that repeat presentations were not coded differently and their exclusion from the primary analyses was unlikely to have biased the estimates of validity. Generalizability to the wider population of hospitalized elderly persons in Victoria may be limited, however, the two participating hospitals were very large central tertiary centres and likely to be representative of this setting.

In conclusion, when medical record notation of pneumonia is used as the standard, we found ICD-10 codes are a valid method for retrospective ascertainment of hospitalized cases of pneumonia and are likely to be superior to use of complexes of symptoms and signs, or interpretation of radiology reports.

ACKNOWLEDGMENTS

DECLARATION OF INTEREST

None.

REFERENCES

1. **Smith PR.** What diagnostic tests are needed for community-acquired pneumonia? *Medical Clinics of North America* 2001; **85**: 1381–1396.

2. **Fine MJ, Chowdhry T, Ketema A.** Outpatient management of community-acquired pneumonia. *Hospital Practice (Office Edition)* 1998; **33**: 123–133.

3. **Metlay JP, et al.** Influence of age on symptoms at presentation in patients with community-acquired pneumonia. *Archives of Internal Medicine* 1997; **157**: 1453–1459.

4. **Marrie TJ, et al.** Community-acquired pneumonia requiring hospitalization. Is it different in the elderly? *Journal of the American Geriatrics Society* 1985; **33**: 671–680.

5. **Albaum MN, et al.** Interobserver reliability of the chest radiograph in community-acquired pneumonia. PORT Investigators. *Chest* 1996; **110**: 343–350.

6. **Syrjala H, et al.** High-resolution computed tomography for the diagnosis of community-acquired pneumonia. *Clinical Infectious Diseases* 1998; **27**: 358–363.

7. **Nichol KL, et al.** The health and economic benefits associated with pneumococcal vaccination of elderly persons with chronic lung disease. *Archives of Internal Medicine* 1999; **159**: 2437–2442.

8. **Christenson B, et al.** Effects of a large-scale intervention with influenza and 23-valent pneumococcal vaccines in adults aged 65 years or older: a prospective study. *Lancet* 2001; **357**: 1008–1011.

9. **Dear K, et al.** Vaccines for preventing pneumococcal infection in adults [Cochrane Review]. *The Cochrane Database of Systematic Reviews* (http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD000422/frame.html). Accessed 24 February 2004.

10. **Vila-Córcoles A, et al.** Protective effect of pneumococcal vaccine against death by pneumonia in elderly subjects. *European Respiratory Journal* 2005; **26**: 1086–1091.

11. **Fisman DN, et al.** Prior pneumococcal vaccination is associated with reduced death, complications, and length of stay among hospitalized adults with community-acquired pneumonia. *Clinical Infectious Diseases* 2006; **42**: 1093–1101.

12. **World Health Organization.** *International Classification of Diseases* (*ICD*). WHO (http://www.who.int/classifications/icd/en/). Accessed 13 August 2006.

13. **Marrie TJ, Durant H, Sealy E.** Pneumonia – the quality of medical records data. *Medical Care* 1987; **25**: 20–24.

14. **Whittle J, et al.** Community-acquired pneumonia: can it be defined with claims data? *American Journal of Medical Quality* 1997; **12**: 187–193.

15. **Guevara RE, et al.** Accuracy of ICD-9-CM codes in detecting community-acquired pneumococcal pneumonia for incidence and vaccine efficacy studies. *American Journal of Epidemiology* 1999; **149**: 282–289.

16. **Comstock GW.** Evaluating vaccination effectiveness and vaccine efficacy by means of case-control studies. *Epidemiologic Reviews* 1994; **16**: 77–89.

17. **Wacholder S.** Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* 1991; **2**: 155–158.

18. **National Centre for Classification in Health.** *International Statistical Classification of Diseases and Related Health Problems, 10th Edition, Australian*

Modification, 1st edn. Sydney: University of Sydney, 1998.

19. **Department of Human Services Victoria.** Victorian Inpatient Medical Database. Performance Reporting & Analysis, Department of Human Services Victoria. Accessed 15 May 2004.

20. **The National Centre for Classification in Health.** *ICD-10-AM Australian Coding Standards*, 2nd edn. Sydney: The National Centre for Classification in Health, 2000.

21. **Last JM.** *A Dictionary of Epidemiology*, 3rd edn. New York: Oxford University Press, 1995.

22. **Neill AM, *et al.*** Community acquired pneumonia: aetiology and usefulness of severity criteria on admission. *Thorax* 1996; **51**: 1010–1016.

23. **Szklo M, Nieto FJ.** Quality assurance and control: the kappa statistic. In: Szklo M, Nieto FJ, eds. *Epidemiology. Beyond the Basics.* Gaithersburg: Aspen Publishers; 2000, pp. 343–406.

24. **Intercooled Stata 9.1 for Windows.** Computer program, version 9.1. College Station, TX: StataCorp LP, 2005.

25. **Watts CG, *et al.*** Establishing thresholds for influenza surveillance in Victoria. *Australia and New Zealand Journal of Public Health* 2003; **27**: 409–412.

26. **Commission on Professional and Hospital Activities.** *International Classification of Diseases, 9th Revision, Clinical Modification, Vol 1.* Ann Arbor, Michigan, 1980.

27. **Fine MJ, *et al.*** Validation of a pneumonia prognostic index using the MedisGroups comparative hospital database. *American Journal of Medicine* 1993; **94**: 153–159.

28. **Aronsky D, *et al.*** Accuracy of administrative data for identifying patients with pneumonia. *American Journal of Medical Quality* 2005; **20**: 319–328.

29. **Aronsky D, Chan KJ, Haug PJ.** Evaluation of a computerized diagnostic decision support system for patients with pneumonia: study design considerations. *Journal of the American Medical Informatics Association* 2001; **8**: 473–485.

30. **Ansaldi F, *et al.*** Effectiveness of a 23-valent polysaccharide vaccine in preventing pneumonia and non-invasive pneumococcal infection in elderly people: a large-scale retrospective cohort study. *Journal of International Medical Research* 2005; **33**: 490–500.

31. **Cherian T, *et al.*** Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. *Bulletin of the World Health Organization* 2005; **83**: 353–359.