Emergency department patient and physician survey accuracy compared to chart abstraction in patients with acute respiratory illness

Katherine M. Hunold, MD, MPH[1], Andrew L. Schwaderer, MD[2], Matthew Exline, MD, MPH[3], Courtney Hebert, MD, MS[4,5], Brent C. Lampert, DO[6], Lauren T. Southerland, MD[1], Julie A. Stephens, MS[7], Edward W. Boyer, MD, PhD[1], Tanya R. Gure, MD[8], Lorraine C. Mion, PhD[9], Michael Hill, RN[1], Ching-Min B. Chu, BS[1], Edriane Ernie, BA[1], Jeffrey M. Caterino, MD, MPH[1]

[1]Department of Emergency Medicine, The Ohio State University, Columbus, Ohio, USA

[2]Section of Pediatric Nephrology, Indiana University, Indianapolis, Indiana, USA

[3]Department of Internal Medicine, The Ohio State University, Columbus, Ohio, USA

[4]Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio, USA

[5]Division of Infectious Disease, The Ohio State University, Columbus, Ohio, USA

[6]Division of Cardiovascular Medicine, The Ohio State University, Columbus, Ohio, USA

[7]Center for Biostatistics, Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio, USA

[8]Division of General Internal Medicine & Geriatrics, The Ohio State University, Columbus, Ohio, USA

[9]College of Nursing, The Ohio State University, Columbus, Ohio, USA

Abstract

Background: High-quality research studies in older adults are needed. Unfortunately, the accuracy of chart review data in older adult patients has been called into question by previous

Correspondence Katherine M. Hunold, MD, The Ohio State University Prior Hall, 376 W 10th Avenue, Columbus, OH 43210, USA. katherine.buck@osumc.edu.

studies. Little is known on this topic in patients with suspected pneumonia, a disease with 500,000 annual older adult U.S. emergency department (ED) visits that presents a diagnostic challenge to ED physicians. The study objective was to compare direct interview and chart abstraction as data sources.

Methods: We present a preplanned secondary analysis of a prospective, observational cohort of ED patients 65 years of age with suspected pneumonia in two Midwest EDs. We describe the agreement between chart review and a criterion standard of prospective direct patient survey (symptoms) or direct physician survey (examination findings). Data were collected by chart review and from the patient and treating physician by survey.

Results: The larger study enrolled 135 older adults; 134 with complete symptom data and 129 with complete examination data were included in this analysis. Pneumonia symptoms (confusion, malaise, rapid breathing, any cough, new/worse cough, any sputum production, change to sputum) had agreement between patient/legally authorized representative survey and chart review ranging from 47.8% (malaise) to 80.6% (confusion). All examination findings (rales, rhonchi, wheeze) had percent agreement between physician survey and chart review of 80%. However, all kappas except wheezing were less than 0.60, indicating weak agreement.

Conclusions: Both patient symptoms and examination findings demonstrated discrepancies between chart review and direct survey with larger discrepancies in symptoms reported. Researchers should consider these potential discrepancies during study design and data interpretation.

INTRODUCTION

The importance of "diagnostic excellence" in older adult patients has been described. Diagnostic excellence has several components including that both over- and underdiagnosis should be avoided but recognizes that older adults have increased medical complexity and vague or atypical symptoms that can complicate accurate diagnosis.[1] Critical to achieving this goal of diagnostic excellence in older adults is the availability of high-quality research to inform clinical practice. Retrospective chart review of patient data is commonly employed in studies of this population and in the emergency department (ED), but chart review may underreport symptoms of infection, as demonstrated in prior reports of difficulty previously examining urinary tract infections (UTIs) among older adult ED patients.[2] Little is known about the accuracy of chart review in identifying symptoms and examination findings among ED patients with respiratory illness.

There are over 500,000 older adult ED visits for pneumonia annually in the United States.[3] The diagnosis of pneumonia in older adults is often delayed due to diagnostic uncertainty,[4,5] which results in inadequate treatment, worsening infection,[6,7] longer lengths of stay, and increased mortality.[8] Thus, improving the accuracy of pneumonia diagnosis in older adult ED patients has the potential to markedly improve clinical care and outcomes. As previous studies rely heavily on chart review, determining the accuracy of chart review in this population is critical to the appropriate interpretation of research results.[4,9,10]

This study's objective was to describe the agreement of chart abstraction versus a criterion standard of direct interview for the presence of symptoms (patient interview) and

examination findings (ED physician interview) of pneumonia in older adult patients with suspected pneumonia.

METHODS

This was a preplanned secondary analysis of patients in a prospective, observational cohort study of older adult ED patients 65 years of age who presented to two EDs with suspected pneumonia. The goal of the primary study was to prospectively evaluate the utility of existing pneumonia diagnostic criteria in older adult ED patients and the potential utility of using antimicrobial peptide (AMP) levels as a biomarker for this diagnosis.

Patients were eligible for inclusion if they were aged 65 years or older and the treating physician ordered chest radiography and had initial suspicion for pneumonia (score of 2 [unlikely] to 5 [very likely] on Likert scale) when asked by research assistants. Patients with a Likert scale of 1 (very unlikely) were not eligible for inclusion. Physician suspicion was used as inclusion criteria to ensure the enrolled population met the goal of the larger study; this evaluated the utility of existing diagnostic criteria and therefore required patients both with and without pneumonia. Exclusion criteria were the inability to obtain consent from the patient or legally authorized representative (LAR), active cancer, organ transplant, immunosuppression, trauma activation, or incarceration. Enrollment via proxy was completed for patients without capacity via patient assent as in previous work.[11] This study was approved by the institutional review board at our institution.

Patients were enrolled by research assistants in two U.S. EDs in the same hospital system, an academic ED with 76,000 annual visits (26% older adults) and a community ED with 44,000 annual visits (17% older adults). Patients were enrolled when research staff were available; this included weekdays and weekends beginning at 7:00 a.m. with coverage as late as 11:00 p.m. Data were collected from the patient (or proxy) directly during the ED visit via survey completed by research assistants, from the treating physician at ED disposition via a paper survey, and via chart review by trained abstractors. Relevant portions of the patient and physician surveys are included in the Appendix. Chart abstractors were trained by a study coordinator or the principal investigator and had a standard codebook, and 10% of charts were abstracted twice to ensure consistency between abstractors.

The presence, absence, or inability to determine presence of pneumonia was determined by expert chart adjudication as done in ours and others previous work.[5,12,13] Each chart was reviewed by two randomly assigned attending physicians ("adjudicators") with a third reviewer if there was disagreement on pneumonia diagnosis between the two initial reviewers. Any remaining disagreements were resolved via in- person meeting. The adjudicators were five attending physicians with training in (emergency medicine, infectious disease, pulmonary/critical care, cardiology, or geriatrics). Adjudicators were aware of existing diagnostic criteria but were not required to follow them; they had access to all patient surveys as well as ED, inpatient, and outpatient records.

We report the presence of symptoms by patient self- report and chart review, agreement between patient self-report and chart review ("percent agreement," kappa, and sensitivity/

specificity) with patient self-report treated as the criterion standard. Similarly, we report presence of examination findings by physician survey and chart review, agreement between physician survey and chart review ("percent agreement"), kappa, and sensitivity/specificity with physician survey treated as the criterion standard. Patient self-report via survey and physician report via survey were treated as the criterion standard based on the previous UTI literature.[2] Patients with missing data were not included in the analyses.

RESULTS

Patients were enrolled from October 3, 2019, to March 23, 2022, with a pause from March 10, 2020, to August 4, 2020, due to the onset of the COVID-19 pandemic. A total of 256 patients were eligible; 135 patients were enrolled in the parent study, with 26.7% via LAR consent. Twenty-seven (20%) had pneumonia by criterion standard review; criterion standard review for pneumonia diagnosis was by a multidisciplinary physician panel review of the chart. The primary reason for exclusion from the study was due to patient refusal (n = 85). Most enrolled patients were female (50.4%) with a mean ± SD age of 75.7 ± 8.4 years. The study population was primarily White/Caucasian (74.8%) or Black/African American (23.7%), a finding consistent with the demography of the region. Most older adults lived at home (88.2%); six (4.5%) resided in a subacute or long- care facility prior to ED presentation. Most patients (109, 80.7%) were admitted to the hospital with 28 (20.7%) going to the intensive care unit or intermediate care unit; none died in the ED. Patients presented with high Emergency Severity Index (ESI) scores with 10 (7.4%) were ESI Level 1 and 95 (70.4%) were ESI Level 2 (Table 1).

Several were excluded from this analysis sample after enrollment in the parent study. One patient was enrolled by a proxy who did not have enough information to answer symptom questionnaire for the patient. Six other patients did not have a physician survey completed. Therefore, subjects with complete included in the data analysis was 134 with completed patient symptom questionnaires and 129 with complete provider evaluations.

We found that pneumonia symptoms had varying levels of percent agreement between patient/LAR survey and chart review ranging from 47.8% (malaise) to 80.6% (confusion). There was consistent percent agreement in examination findings between physician survey and chart review of 80%. When there was disagreement, symptoms or examination findings were more often present by patient or physician survey and absent on chart review. For example, for malaise (47.8% agreement), 43.3% documented malaise by direct patient survey but it was absent on chart review while the remaining 9.0% were present on chart review but absent on patient survey. New/worse cough and wheeze were the only symptoms/examination findings with more present on chart review compared to direct survey (Table 2).

Kappas for all symptoms (confusion, malaise, rapid breathing, any cough, new/worse cough, any sputum production, change to sputum) and examination findings of rales and rhonchi were less than 0.60, indicating weak agreement. The only examination finding with kappa > 0.6 was wheeze (0.62).

Sensitivity and specificity of chart review were calculated using direct interview/survey as the criterion standard. The sensitivity of chart review for both symptoms and examination findings was low (<65%) with the worst being rapidly breathing (25.0 [95% CI 15.0–37.4]) and best wheeze (62.6 [40.6–81.2]). Specificity had a wider range with the worst being new or worse cough (68.2 [57.4–77.7]) and best any sputum production (97.7 [92.0–99.7]).

DISCUSSION

We found consistently weak agreement (kappa < 0.60) with conventional interpretation[14] as well as consistently poor sensitivity for all symptoms and examination findings except wheezing in a prospective study of older adult ED patients with suspected pneumonia. Both symptoms and examination findings are systematically underrepresented in chart documentation. When symptoms are not present, specificity is generally high. However, chart documentation likely underreports symptoms and abnormal examination findings, in some cases by a substantial degree.

The discrepancy between direct survey and chart review could be due to many reasons including but not limited to the fact that (1) patients may alter their story between interviewers either by accident or due to remembering with additional prompting; (2) treating physicians may not specifically ask for each symptom and patients may not volunteer unless specifically asked; (3) treating physicians do not document every symptom they elicit from the patient in the chart; (4) there may be delay between physician patient interview and chart completion that could cause unintentional chart omissions or inclusions; and (5) charting details may be changed by clarity of diagnostic tests, for example, an imaging study that shows clear pneumonia.

It is important to consider what the observed kappa levels mean in terms of data reliability. The majority of the kappas in this study were less than 0.60. In the best- case scenario, a kappa of 0.40–0.69 indicates that 15%–35%[14] of the data are reliable. Kappa accounts for the fact that some agreement may be due to chance whereas percent agreement does not.[15]

In comparison, the best percent agreement for patient- supplied data and chart review was 80.6%; for physician- supplied data and chart reviewed was 85.8%. This translates to 20.4% of symptom data and 14.2% of examination findings data being incorrect by chart review. It is unlikely that these rates of potentially incorrect data are acceptable for research data.

Data collected by chart review/data abstraction by human reviewers and/or direct extraction from electronic medical records (EMRs) are commonly used in research, quality improvement, and national metrics. A previous study demonstrated similar results in older adult ED patients with UTIs with agreement percentages between patient self-report and chart review ranging from 60% to 95% for genitourinary symptoms.[2] Taken together, these studies suggest that, when possible, symptoms and examination findings in older adult ED patients should be collected prospectively in a standardized manner. Further, there are many considerations to ensure that chart reviewed data are accurately abstracted[16]; these were considered in this study and should be considered in any using retrospective chart reviewed data.

There are several practical reasons that chart reviewed data may be preferred by researchers. First, prospective data collection is more expensive than chart review; it requires personnel to administer and record surveys in addition to time by those being surveyed. Second, retrospective chart reviews often allow for larger sample sizes as is often the case when large data sets are used for "data mining." While this methodology may be appropriate for initial investigations, findings must be examined in a prospective manner to confirm or refute results. Third, chart data are "more practical" in that it is what would be available to machine learning algorithms that are currently in use in EMRs. Both chart review studies and prospective data collection have a role in research, but addressing the potential limitations based on the chosen methodology are critical.

Chart review data are often used for quality-of-care analysis or recommendations for public policy. Unfortunately, based on the results presented here, this must be reconsidered and/or the results must be viewed with caution. Chart review or data mining methods frequently demonstrate the minimum presence of a findings rather than the maximum. For example, if a hypothetical study using EMR data found that 20% of patients given antibiotics for pneumonia had no examination findings, the conclusions should be tempered because our study demonstrates the proportions of symptoms and examination findings obtained via chart review may be inaccurate. Instead, it is likely that those truly without symptoms are much less than 20%. This would have direct impacts on potential recommendations from these data as the perceived quality of care may be lower than actual; for example, overprescribing of antibiotics may be lower than the data suggest.

Although not assessed in this study, there may be clinical implications of inaccurate charting. Accurate and complete charting was associated with improved measures of mortality in trauma patients.[17] One proposed solution to chart data inaccuracies is direct patient- entered data,[18] which is the clinical equivalent of direct survey in research.

STRENGTHS AND LIMITATIONS

Strengths of our study that improve generalizability include the inclusion of patients who required LAR for consent as well as ill patients. Our study has several imitations. First, we do not know who is "correct"—direct survey or chart review. Thus, we have speculated potential causes for the observed discrepancy but cannot be sure of the cause. Second, we treat data collected from patients and LARs the same in our analysis. While LARs provided consent for over one- quarter of participants, often the patient helped provide the answers to the surveys. Thus, this cannot be fully separated. Third, only patients with suspected pneumonia were included, but there is no reason these patients are systematically different from older adult patients with other presenting complaints and similar results have been demonstrated in older adult UTI patients.[2] It is unknown how this generalizes to younger patients. Fourth, patients were enrolled at times when research assistants were available, and this did not include overnight hours. If patients who present during overnight hours are different from the hours of enrollment, results may not be generalizable, but we believe this is unlikely. Finally, physician surveys were completed by attending physicians; it is unknown if different results would be obtained with other care team members such as resident physicians, nurse practitioners, or physicians assistants.

CONCLUSIONS

In a cohort of older adults presenting to the ED with suspected pneumonia, chart review data may not be reliable compared to prospective data collection directly from patient and the treating attending physician. Chart review greatly underreported symptoms compared to direct interview; physical examination findings were also underreported via chart review but to a lesser extent. It is imperative that researchers carefully consider how their data collection method could impact the reliability of data and how this may impact interpretation of the results.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding information

REFERENCES

1. Cassel C, Fulmer T. Achieving diagnostic excellence for older patients. JAMA. 2022;327:919–920. [PubMed: 35175301]

2. Caterino JM, Stephens JA, Camargo CA Jr, et al. Asymptomatic bacteriuria versus symptom underreporting in older emergency department patients with suspected urinary tract infection. J Am Geriatr Soc. 2020;68:2696–2699. [PubMed: 33460062]

3. National Hospital Ambulatory Medical Care Survey. Emergency department summary tables. Ambulatory and Hospital Care Statistics Branch of the National Center for Health Statistics. 2018. Accessed February 20, 2023. https://www.cdc.gov/nchs/data/nhamcs/web_tables/2018-ed-web-tables-508.pdf

4. Chandra A, Nicks B, Maniago E, Nouh A, Limkakeng A. A multi-center analysis of the ED diagnosis of pneumonia. Am J Emerg Med. 2010;28:862–865. [PubMed: 20887906]

5. Hunold KM, Schwaderer AL, Exline M, et al. Diagnosing dyspneic older adult emergency department patients: a pilot study. Acad Emerg Med. 2020;28:675–678. [PubMed: 33249675]

6. Kollef MH, Sherman G, Ward S, Fraser VJ. Inadequate antimicrobial treatment of infections: a risk factor for hospital mortality among critically ill patients. Chest. 1999;115:462–474. [PubMed: 10027448]

7. Davey PG, Marwick C. Appropriate vs. inappropriate antimicrobial therapy. Clin Microbiol Infect. 2008;14(Suppl 3):15–21. [PubMed: 18318875]

8. Green SM, Martinez-Rumayor A, Gregory SA, et al. Clinical uncertainty, diagnostic accuracy, and outcomes in emergency department patients presenting with dyspnea. Arch Intern Med. 2008;168:741–748. [PubMed: 18413557]

9. Nouvenne A, Ticinesi A, Folesani G, et al. The association of serum procalcitonin and high-sensitivity C-reactive protein with pneumonia in elderly multimorbid patients with respiratory symptoms: retrospective cohort study. BMC Geriatr. 2016;16:16. [PubMed: 26772604]

10. Basi SK, Marrie TJ, Huang JQ, Majumdar SR. Patients admitted to hospital with suspected pneumonia and normal chest radiographs: epidemiology, microbiology, and outcomes. Am J Med. 2004;117:305–311. [PubMed: 15336579]

11. Caterino JM, Kline DM, Leininger R, et al. Nonspecific symptoms lack diagnostic accuracy for infection in older patients in the emergency department. J Am Geriatr Soc. 2019;67:484–492. [PubMed: 30467825]

12. Caterino JMKD, Leininger R, Southerland LT, et al. . Nonspecific symptoms lack diagnostic accuracy for infection in older emergency department patients. J Am Geriatr Soc. 2019;67:484–492. [PubMed: 30467825]

13. Caterino JM, Leininger R, Kline DM, et al. Accuracy of current diagnostic criteria for acute bacterial infection in older adults in the emergency department. J Am Geriatr Soc. 2017;65:1802–1809. [PubMed: 28440855]

14. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22:276–282. [PubMed: 23092060]

15. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: measures of agreement. Perspect Clin Res. 2017;8:187–191. [PubMed: 29109937]

16. Vassar M, Holzmann M. The retrospective chart review: important methodological considerations. J Educ Eval Health Prof. 2013;10:12. [PubMed: 24324853]

17. Elkbuli A, Godelman S, Miller A, et al. Improved clinical documentation leads to superior reportable outcomes: an accurate representation of patient's clinical status. Int J Surg. 2018;53:288–291. [PubMed: 29653245]

18. Wuerdeman L, Volk L, Pizziferri L, et al. How accurate is information that patients contribute to their electronic health record? AMIA Annu Symp Proc. 2005;2005:834–838. [PubMed: 16779157]

TABLE 1

Demographic characteristics of older ED patients with suspected pneumonia ($n = 135$).

Patient characteristic	
Age (years), mean (±SD)	75.7 (±8.4)
Female	68 (50.4)
Race	
White/Caucasian	101 (74.8)
African American	32 (23.7)
Other/unknown	2 (1.5)
Hispanic	1 (0.8)
Marital status	
Married	58 (43.0)
Never married	13 (9.6)
Divorced	32 (23.7)
Separated	1 (0.7)
Widowed	31 (23.0)
Education level	
Less than high school	13 (9.6)
High school graduate	36 (26.7)
Some college	44 (32.6)
College graduate	21 (15.6)
Graduate degree	19 (14.1)
Unknown	2 (1.5)
Place of residence	
Home/apartment	119 (88.2)
Assisted living	8 (5.9)
Rehabilitation	1 (0.7)
SNF/nursing home	6 (4.5)
Unknown	1 (0.7)
Emergency Severity Index (ESI) score	
1 (most urgent)	10 (7.4)
2 (emergent)	95 (70.4)
3 (urgent)	29 (21.5)
4 (less urgent)	1 (0.7)
5 (nonurgent)	0 (0)
ED disposition	
Admit to floor	81 (60)
Admit to intermediate or intensive care	28 (20.7)
ED or hospital observation unit	3 (2.2)
Discharge home	21 (15.6)
Transfer to rehab/SNF directly	2 (1.5)

Note: All data are presented as *n* (%) unless otherwise noted. Abbreviation: SNF, skilled nursing facility.

TABLE 2

Agreement and test characteristics of pneumonia symptoms (134 subjects) and examination findings (129 subjects) present on chart review compared to a criterion standard.

	Patient or physician survey positive[a]	ED chart review positive[a]	Agreement[a]	Present by patient or physician survey, absent on chart review[a]	Present on chart review, absent by patient or physician survey	k (95% CI)	Sensitivity of chart review[b]	Specificity of chart review[b]
Symptoms								
Confusion (n = 133[c])	42 (31.6)	23 (17.3)	108 (80.6)	22 (16.5)	3 (2.3)	0.51 (0.35–0.66)	47.6 (32.0–63.6)	96.7 (90.7–99.3)
Malaise	94 (70.2)	48 (35.8)	64 (47.8)	58 (43.3)	12 (9.0)	0.06 (−0.07–0.19)	38.3 (28.5–48.9)	70.0 (53.5–83.4)
Rapid breathing	64 (47.8)	26 (19.4)	76 (56.7)	48 (35.8)	10 (7.5)	0.11 (−0.03–0.25)	25.0 (15.0–37.4)	85.7 (75.3–92.9)
Any cough	65 (48.5)	54 (40.3)	85 (63.4)	30 (22.4)	19 (14.2)	0.26 (0.10–0.43)	53.8 (41.0–66.3)	72.5 (60.4–82.5)
New/worse cough	46 (34.3)	54 (40.3)	86 (64.2)	20 (14.9)	28 (20.9)	0.24 (0.07–0.40)	56.5 (41.1–71.1)	68.2 (57.4–77.7)
Any sputum production	46 (34.3)	16 (11.9)	100 (74.6)	32 (23.9)	2 (1.5)	0.33 (0.18–0.49)	30.4 (17.7–45.8)	97.7 (92.0–99.7)
Change to sputum	29 (21.6)	16 (11.9)	105 (78.4)	21 (15.7)	8 (6.0)	0.24 (0.04–0.43)	27.6 (12.7–47.2)	92.4 (85.5–96.7)
Exam								
Rales	20 (15.5)	15 (11.6)	112 (83.5)	11 (8.5)	6 (4.7)	0.44 (0.22–0.66)	60.0 (32.3–83.7)	90.4 (83.4–95.1)
Rhonchi	20 (15.5)	7 (5.4)	110 (82.1)	16 (12.4)	3 (2.3)	0.24 (0.01–0.46)	57.1 (18.4–90.1)	86.9 (79.6–92.3)
Wheeze	20 (15.5)	24 (18.6)	115 (85.8)	5 (3.9)	9 (7.0)	0.62 (0.44–0.80)	62.5 (40.6–81.2)	95.2 (89.2–98.4)

[a] Data are reported as n (%).

[b] Data are reported as % (95% CI).

[c] One legally authorized representative did not know if the patient had been experiencing confusion.