

An Adjudication Protocol for Severe Pneumonia

Chiagozie I. Pickens,^{1,*} Catherine A. Gao,¹ Justin Bodner,¹ James M. Walter,¹ Jacqueline M. Kruser,^{1,2} Helen K. Donnelly,¹ Alvaro Donayre,¹ Katie Clepp,¹ Nicole Borkowski,¹ Richard G. Wunderink,¹ and Benjamin D. Singer,¹ The NU SCRIPT Study Investigators

¹Division of Pulmonary and Critical Care Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA, and ²Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, University of Wisconsin, Madison, Wisconsin, USA

Background. Clinical end points that constitute successful treatment in severe pneumonia are difficult to ascertain and vulnerable to bias. The utility of a protocolized adjudication procedure to determine meaningful end points in severe pneumonia has not been well described.

Methods. This was a single-center prospective cohort study of patients with severe pneumonia admitted to the medical intensive care unit. The objective was to develop an adjudication protocol for severe bacterial and/or viral pneumonia. Each episode of pneumonia was independently reviewed by 2 pulmonary and critical care physicians. If a discrepancy occurred between the 2 adjudicators, a third adjudicator reviewed the case. If a discrepancy remained after all 3 adjudications, consensus was achieved through committee review.

Results. Evaluation of 784 pneumonia episodes during 593 hospitalizations achieved only 48.1% interobserver agreement between the first 2 adjudicators and 78.8% when agreement was defined as concordance between 2 of 3 adjudicators. Multiple episodes of pneumonia and presence of bacterial/viral coinfection in the initial pneumonia episode were associated with lower interobserver agreement. For an initial episode of bacterial pneumonia, patients with an adjudicated day 7–8 clinical impression of cure (compared with alternative impressions) were more likely to be discharged alive (odds ratio, 6.3; 95% CI, 3.5–11.6).

Conclusions. A comprehensive adjudication protocol to identify clinical end points in severe pneumonia resulted in only moderate interobserver agreement. An adjudicated end point of clinical cure by day 7–8 was associated with more favorable hospital discharge dispositions, suggesting that clinical cure by day 7–8 may be a valid end point to use in adjudication protocols.

Keywords. adjudication; clinical; end point; pneumonia; severe.

Assessing response to therapy for severe pneumonia is critical for both clinical care and research studies but is complicated in the critically ill. For most antibiotic registration trials, the Food and Drug Administration (FDA) assesses efficacy using the standard of clinical cure *and* bacterial clearance from the site of infection as the primary end point [1]. However, clinical criteria for severe pneumonia are subjective and prone to errors, particularly in critically ill patients. Multiple alternative causes for radiographic infiltrates and systemic inflammation lead to misdiagnosis of pneumonia in up to 50% of cases [2–4]. A considerable proportion of these pneumonia mimics are noninfectious. These

alternative diagnoses can also confuse the assessment of clinical response to antibiotics [5], leading to antibiotic overuse and misuse. This diagnostic confusion led the FDA to exclude clinical response as a noninferiority end point for registration trials of new antibiotics for hospital-acquired pneumonia (HAP) and ventilator-associated pneumonia (VAP) [6, 7].

When conditions with a heterogenous clinical course are the subject of a clinical trial, study protocols use adjudication committees to provide consistent, unbiased determination of the presence or absence of a diagnosis, event, or end point [8–10]. Adjudication committees are established components of large cohort studies and randomized clinical trials in cardiovascular disease, oncology, interstitial lung disease, and sepsis [11–15]. The process of adjudicating clinical end points in a large cohort of patients with severe pneumonia has not been well described [16–19], although such processes have been incorporated into critical care trials. In severe pneumonia, end points such as death or extubation failure are easily ascertained, while other end points like clinical cure, microbiologic resolution, and safety are not standardized and subject to bias in interpretation [20]. Existing studies that leverage an adjudication committee for pneumonia do not describe the process or the reliability of the method in detail.

Assessment of response to therapy is an integral aspect of observational and interventional studies of pneumonia. We

Received 08 May 2023; editorial decision 26 June 2023; accepted 28 June 2023; published online 1 July 2023

Correspondence: Chiagozie Pickens, MD, Division of Pulmonary and Critical Care, Northwestern University Feinberg School of Medicine, 303 E. Superior Street, Simpson Querrey 5th Floor, Suite 5-406, Chicago, IL 60611 (chiagozie-ononye@northwestern.edu); or Richard Wunderink, MD, Division of Pulmonary and Critical Care, Northwestern University Feinberg School of Medicine, 303 E. Superior Street, Simpson Querrey 5th Floor, Suite 5-301, Chicago, IL 60611 (r-wunderink@northwestern.edu).

Open Forum Infectious Diseases®

© The Author(s) 2023. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

<https://doi.org/10.1093/ofid/ofad336>

therefore developed and validated a protocol for adjudication of clinical end points in a prospective cohort of critically ill, mechanically ventilated patients with suspected pneumonia as part of the Successful Clinical Response in Pneumonia Therapy (SCRIPT) study.

METHODS

SCRIPT is a single-center, prospective, observational cohort study of mechanically ventilated patients with suspected pneumonia (Northwestern University IRB #STU00204868). In the medical intensive care unit (ICU) at our institution, bronchoscopic bronchoalveolar lavage (BAL) or nonbronchoscopic BAL (NBBAL) is routinely performed to identify the presence and define the etiology of pneumonia in intubated patients [21–23]. Our adjudication protocol was developed before the start of SCRIPT patient enrollment in 2018 and subsequently modified during the COVID-19 pandemic in 2020 to include adjudication of viral pneumonia.

Clinical End Points

Our protocol was designed to ascertain interim clinical end points as well as a final clinical end point at the time of a patient's death or hospital discharge. Interim clinical end points were defined as cure, indeterminate, persistence, or superinfection. See the [Supplementary Data](#) for detailed definitions of each term. Of note, BAL fluid data incorporated into the assessment of clinical end points included culture results, multiplex polymerase chain reaction (PCR) results from the BioFire Pneumonia Panel, cell count, and differential and amylase levels. All other microbiology results, including blood cultures and urinary antigen testing for *Streptococcus pneumoniae* and *Legionella pneumophila*, were available to the adjudicators. For bacterial pneumonia or bacterial/viral coinfection pneumonia, interim clinical end points were assessed at day 7–8, day 10, and day 14, corresponding to common antibiotic discontinuation time points in clinical trials and clinical practice. These time points were not used to assess cases of viral-only pneumonia because standard durations of empirical antibacterial therapy in viral pneumonia are unknown. End points were defined based on literature review and expert consensus [24–26]. Because part of our objective was to evaluate patients at various time points during an episode of pneumonia, we did not use traditional measures like mortality, ventilator-free days, or microbiological cure as standalone end points.

Pneumonia Episode Definition

Pneumonia was defined as a positive BAL culture, positive BAL fluid multiplex PCR, or negative BAL fluid culture with a white blood cell differential containing >50% neutrophils and no alternative explanation in a patient with clinical signs and symptoms of pneumonia as identified by the treating physician. Note that culture-negative pneumonia includes cases that were

negative by PCR-based testing. We defined day 0 of the initial episode of pneumonia by the date of the clinical BAL procedure that prompted enrollment into SCRIPT. An episode of pneumonia ended when systemic antibiotics for pneumonia were discontinued without recurrent signs of pneumonia for at least 48 hours (defined in detail in the [Supplementary Data](#)). The onset of subsequent episodes of pneumonia within the same hospitalization was likewise defined by the timing of a diagnostic BAL procedure occurring after at least 48 hours off antibiotics directed at pneumonia and without persistent or recurrent signs of pneumonia during the interval. We applied standard definitions for pneumonia at the time of each BAL procedure [27, 28]. See the [Supplementary Data](#) for descriptions of clinical community-acquired pneumonia (CAP), HAP, and VAP and additional definitions used to standardize the adjudication process. Importantly, the definitions of CAP, HAP, and VAP were applied based on the duration of admission or ventilation preceding the time of BAL sampling irrespective of the timing or locale of symptom onset, as these durations typically dictate the spectrum of empirical antibiotics administered to patients with suspected pneumonia. We defined episodes on and subsequent to the date of the SCRIPT study enrollment BAL procedure up to day 99, censoring any episodes preceding SCRIPT enrollment or beginning after day 99. Patients who underwent lung transplantation during the enrollment hospitalization were adjudicated as having died.

Adjudication Committee

The adjudication committee was composed of 6 pulmonary and critical care physicians at Northwestern Medicine. Two adjudicators independently reviewed each patient's complete electronic health record and entered written responses to each question on the evaluation form (included in the [Supplementary Data](#)). Review was conducted after the patient's hospital discharge. If both adjudicators provided the same responses to all questions on the evaluation form, consensus was achieved. If discrepancy in at least 1 critical response occurred, the discrepant question was highlighted on a new, blank evaluation form and given to a third adjudicator, who answered the specific question based on chart review, blinded to the previous adjudicators' responses. Agreement between the third adjudicator and 1 of the original adjudicators was considered the final response. If a discrepancy remained between all 3 adjudicators, a committee review was performed with a minimum of 3 adjudicators present, and a consensus answer was provided by the committee. If a case was identified by the first adjudicator as a nonpneumonia episode, this classification was confirmed by a second adjudicator if requested by the first adjudicator. The first adjudicator could request direct committee review for unusual, complex cases.

Committee adjudication took place in person or by video conference on a weekly basis. The total committee group

membership at any time point was 5 physicians. The committee held several meetings for quality control to ensure that the manual data entered into the adjudication worksheet aligned with the data entered into the electronic database and the designated etiologies for pneumonia (ie, viral, bacterial, coinfection, culture-negative, nonpneumonia, and indeterminate) were accurate per written guidelines available to all committee members (included in the [Supplementary Data](#)).

Adjudication Worksheet

The full adjudication worksheet consisted of 20 questions and is included in the [Supplementary Data](#). Depending on the complexity of the case and the branching logic of the form, not all questions required a response. Additionally, some questions had objective, numerical answers that did not require multi-adjudicator concordance to resolve discrepancies. Because the answers to these questions could be identified by chart review, a research coordinator reviewed the case to resolve discrepant responses to these objectively verifiable questions rather than involving a third adjudicator: (1) number of days antibiotics were given for viral-only pneumonia; (2) if initial sample, has the patient been actively treated for pneumonia for >24 hours before sample collection?; (3) are serial procalcitonin values available?; and (4) level of evidence for extrapulmonary infection.

Statistical Analysis

Statistical analysis was performed using R Studio (RStudio, PBC, Boston, MA, USA). The Sankey diagram in [Figure 1](#) was created using Python 3.9. Comorbidities were mapped using ICD codes with Charlson comorbidity index definitions, with only comorbidities present before admission contributing to totals. Crude rates of interobserver agreement were calculated. To estimate a probability of chance agreement for Cohen's kappa statistic (observed agreement—chance agreement/1 – chance agreement), we calculated the probability of chance agreement on 3 mandatory questions: Appropriate Antibiotics (3 response options), Clinical Impression (5 response options), and End of Hospitalization Clinical Outcome (2 response options). This resulted in a 0.04 probability of chance agreement. In a highly cited review, McHugh et al. advise researchers to rely on the kappa statistic if frequent guessing among observers is likely, but if raters are well trained and likelihood of guessing is low, reliance on percent agreement as a reflection of interrater reliability is safe [29]. Reliability was assessed by using a test of proportions between the rates of interobserver agreement by year. The Pearson chi-square or Fisher exact test with Bonferroni correction for multiple comparisons was performed to associate adjudicated clinical end points with final discharge dispositions.

RESULTS

From June 2018 to June 2022, 784 pneumonia episodes from 593 patients were adjudicated. Some patients had >1 episode of pneumonia during hospitalization, and thus the total number of pneumonia episodes exceeded the number of patients in the study. Patient demographics and clinical characteristics are listed in [Table 1](#). A small number (6.2%) of hospitalizations were excluded from the analysis of interobserver agreement because (1) the case was not reviewed by multiple adjudicators because the first adjudicator identified the episode(s) as nonpneumonia or requested an expedited committee review for unusual, complex, or previously undefined conditions (n = 37) or (2) the case was reviewed using different versions of the protocol (n = 18); thus, the total number of hospitalizations reviewed by multiple adjudicators was 538 ([Figure 1](#)). Interobserver concordance between the first 2 adjudicators occurred in 259/538 (48.1%) of hospitalization-level examinations; hence, 279/538 (51.9%) cases went to a third adjudicator to evaluate at least 1 point of disagreement. Discrepancies between all 3 adjudicators remained in 114 of these 279 cases, requiring committee review to arrive at a final agreement on the appropriate end point(s) for the patient. Consensus (either agreement of 2/2 or 2/3 adjudicators) without the need for committee review was achieved in 424/538 (78.8%) cases.

To identify differences in the rates of consensus agreement over time, we calculated crude rates of interobserver agreement by year ([Figure 2](#)). At least 2 adjudicators reviewed 39 cases from 2018, 95 cases from 2019, 257 cases from 2020, 124 cases from 2021, and 23 cases from 2022 (538 total). In the years spanning 2018 to 2022, interobserver agreement, defined as agreement between 2 of 2 adjudicators, was 56%, 56%, 41%, 52%, and 57%, respectively. The Cohen's kappa coefficient per year for interobserver agreement based on agreement between 2 of 2 adjudicators was 0.52, 0.54, 0.39, 0.50, and 0.55, respectively. When interobserver agreement was defined as agreement between 2 of 3 adjudicators, the crude rate of agreement increased to 87%, 85%, 70%, 86%, and 100% agreement. A test for proportions demonstrated a significant difference in interobserver agreement by year ($\chi^2 = 27.3$; $df = 4$; $P \leq .01$). A test for proportions excluding the early pandemic year 2020 demonstrated no significant difference in interobserver agreement between 2018, 2019, 2021, and 2022 ($\chi^2 = 3.8$; $df = 3$; $P = .29$).

We sought to determine whether the etiology of the initial pneumonia episode and the number of pneumonia episodes in each hospitalization were associated with interobserver agreement, defined as consensus between 2 of 3 adjudicators without the need for committee review. The proportion of initial pneumonia episodes that were clinical CAP, HAP, or VAP did not differ between cases that had consensus without committee review and cases that required committee review for consensus. We examined the microbial etiology of the initial episode of pneumonia as a predictor of agreement between

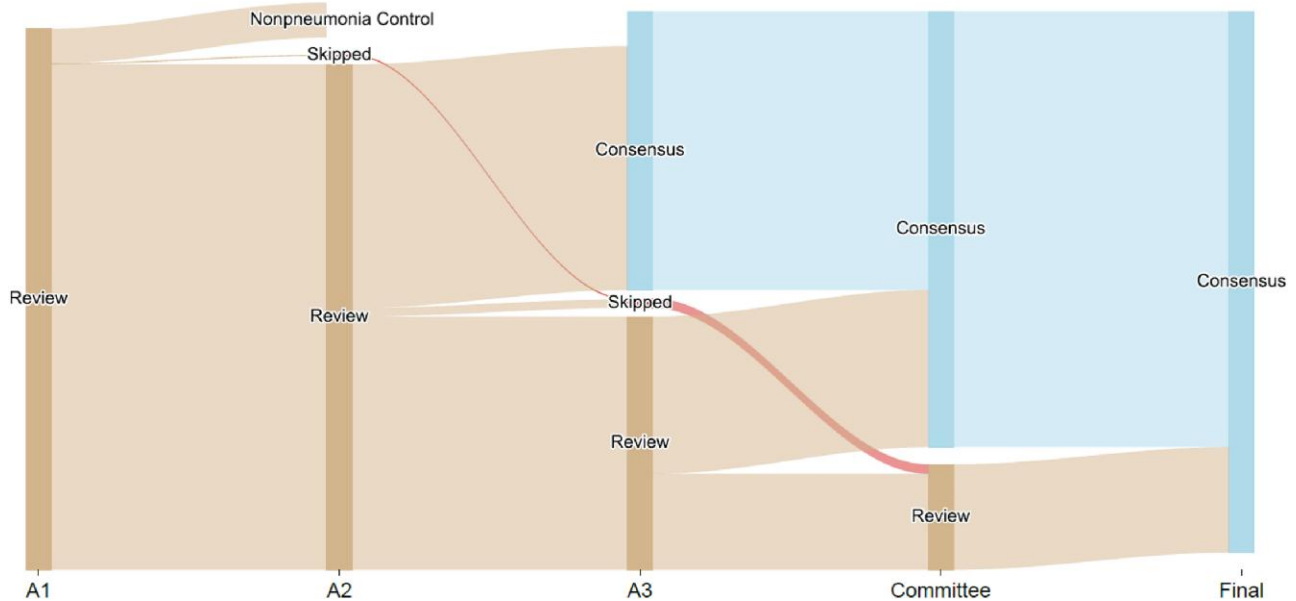


Figure 1. Diagram illustrating the adjudication process. Out of 593 hospitalizations, 55 had only 1 adjudicator (A1), 259 cases had 2 adjudicators (A2), and 279 cases required a third adjudicator (A3). A small subset went directly to committee review after A1 and A2.

adjudicators (Table 2). An initial episode of bacterial/viral coinfection resulted in low agreement between adjudicators (67.5%). This was significant when the proportion of bacterial coinfection cases requiring committee review was compared with the proportion of all other cases that required committee review ($P = .01$). Similarly, cases with multiple pneumonia episodes also resulted in low agreement between adjudicators (48.5%; $P = .01$). The specific bacteria identified by culture and viruses identified by PCR are listed in Supplementary Tables 2 and 3.

We identified the most common questions in the adjudication worksheet that led to discordant answers between the first 2 adjudicators. Cases of viral and bacterial/viral coinfection were excluded from this analysis because the viral adjudication worksheet underwent several iterations during the COVID-19 pandemic. The most common question discordant between adjudicators was “What was the overall global clinical outcome?”, which asks the adjudicator to assess the status of the patient with respect to pneumonia at the time of death or hospital discharge, followed by the question on clinical impression at day 7–8. The percentage of discordance for each question is outlined in Supplementary Figure 1.

To determine the construct validity of our clinical end points for bacterial pneumonia, we assessed the discharge disposition based on the day 7–8 adjudicated clinical end point of the initial pneumonia episode. A day 7–8 adjudicated clinical end point was available for 285 of the 593 hospitalizations; 308 did not have a day 7–8 clinical end point for the initial episode because 94 were nonpneumonia episodes, 153 had viral infection only, and 61 were discharged, transferred to another hospital, or died

before a day 7–8 assessment. Discharge dispositions were grouped into favorable outcomes (home, acute inpatient rehabilitation, skilled nursing facility, long-term acute care facility) vs unfavorable outcomes (death or hospice). As demonstrated in Figure 3, the day 7–8 adjudicated clinical end point of the initial episode of pneumonia was associated with distinct discharge dispositions. Eighty-five percent (136/160) of patients with a day 7–8 clinical end point of cure had a favorable discharge disposition; 47% (59/125) patients with a day 7–8 clinical end point other than cure (indeterminate, persistence, superinfection, or both persistence and superinfection) had a favorable discharge disposition. This difference in favorable discharge disposition based on a day 7–8 clinical end point of cure was significant (odds ratio, 6.3; 95% CI, 3.5–11.6). Significant differences remained when comparing the proportion of patients with favorable discharge dispositions between cure vs indeterminate (85% vs 40%; $P = .05$), cure vs superinfection (85% vs 45%; $P < .01$), and cure vs nonpneumonia (85% vs 55%; $P < .01$) but was not present when comparing cure vs persistence (85% vs 63%; $P = .07$). In patients with viral-only pneumonia, a higher proportion of patients with a clinical impression of cure were discharged to favorable dispositions compared with patients with an adjudicated clinical end point of indeterminate, superinfection, or persistence (Supplementary Table 4).

DISCUSSION

Pneumonia is a complex clinical syndrome with poorly defined end points for treatment success or failure. Because of this limitation, the FDA requires all-cause mortality at 14–28 days as a

Table 1. Demographics and Outcomes of Participants With Adjudicated Episodes of Pneumonia

Demographics and Outcomes of Study Participants	
Variable	No. (%)
Age, median [IQR], y	62 [51–72]
Gender	
Male	349 (59.0)
Female	244 (41.1)
Ethnicity	
Hispanic	124 (20.9)
Not Hispanic or Latino	446 (75.2)
Not answered	23 (3.9)
Race (self-identified)	
African American or Black	119 (20.1)
American Indian or Alaska Native or Asian Indian	6 (1.0)
Asian	17 (2.9)
Native Hawaiian or other Pacific Islander	2 (0.3)
White	351 (59.2)
None of the above or patient declined to answer	98 (16.5)
Pneumonia category for initial episode (n = 593)	
Clinical CAP	132 (22.3)
Clinical HAP	205 (34.6)
Clinical VAP	162 (27.3)
Nonpneumonia episode (37 single adjudicator, 57 adjudicated) ^a	94 (15.9)
Pneumonia etiology (n = 593)	
Viral only	153 (25.8)
Bacterial only	150 (25.3)
Viral/bacterial coinfection	87 (14.7)
Indeterminate	3 (0.1)
Culture-negative	106 (17.9)
Nonpneumonia	94 (15.9)
No. of episodes (n = 593)	
Multiple episodes	120 (20.2)
Single episode	473 (79.8)
Comorbid conditions	
Cerebrovascular disease	124 (20.9)
Congestive heart failure	173 (29.2)
Malignancy	200 (33.7)
Diabetes	213 (35.9)
Chronic pulmonary disease	212 (35.7)
Liver disease	155 (25.2)
Renal disease	159 (26.8)
Hospitalization	
Hospital length of stay, median [IQR]	23 [13–37]
ICU length of stay, median [IQR]	14 [6–25]
Patients with day 7–8 adjudication ^a	257 (60.0)
Hospital mortality	189 (37.5)

Abbreviations: CAP, community-acquired pneumonia; HAP, hospital-acquired pneumonia; ICU, intensive care unit; IQR, interquartile range; VAP, ventilator-associated pneumonia.

^aPatients with viral-only pneumonia, nonpneumonia, and those who were deceased, discharged, or transferred to another facility before day 7–8 did not have a day 7–8 adjudication. The etiologies of pneumonia and of nonpneumonia respiratory failure are listed in [Supplementary Table 1](#).

primary end point for noninferiority registration clinical trials of new therapy for bacterial HAP/VAP [30]. However, mortality fails to capture important patient-centered outcomes [31]. Using the Delphi method, a panel of international experts

attempted to establish a consensus composite end point for clinical trials related to HAP/VAP [24]. The panel discussed a wide range of possible end points, including clinical cure, ventilator-free days, decrease in clinical pulmonary infection score, microbiological cure, safety, change in procalcitonin, and acquisition of antimicrobial resistance. Clinical cure was ranked as the most important primary outcome, yet a universally accepted definition for clinical cure in pneumonia does not exist. Weiss et al. used an iterative process to ultimately define clinical cure as resolution of signs and symptoms of infection, improvement in oxygenation, and no appearance of new signs of sepsis [24].

Adjudication of clinical end points in patients with severe pneumonia has been incorporated into trials, but existing studies that leverage an adjudication committee for pneumonia do not describe the process or the reliability of the method in detail. A large randomized controlled CAP trial used an adjudication committee to determine the primary outcome—clinically indicated treatment with antibiotics—but did not detail the internal or external validity of the adjudication process [16]. Another used a clinical evaluation committee to define the subpopulation of enrolled subjects for an optimal assessment of the mortality benefit of an intervention [32]. Other studies focus on diagnostic adjudication of pneumonia presence and do not address end point adjudication [17, 19]. Because no standardized protocol for adjudication of pneumonia currently exists, clinical trials of pneumonia may have different outcomes when adjudicated using different adjudication procedures.

In this study, we developed a protocol for adjudicating the diagnosis and clinical end points of severe bacterial and viral pneumonia in mechanically ventilated patients, including incorporation of the criteria of Weiss et al. into our definition of cure at days 7–8, 10, and 14. Despite using consensus definitions, our protocol for clinical classification and adjudication of episodes of pneumonia resulted in 48.1% agreement between 2 adjudicators and a 78.8% overall crude rate of consensus agreement when a third adjudicator attempted to resolve discrepancies. For FDA registration trials, usually only a single investigator assesses clinical response. Our study demonstrates that these single-evaluator assessments of clinical response in pneumonia are prone to error.

There are multiple potential reasons for discordance between adjudicators. The adjudication guide was 4 pages long and may be interpreted differently by different adjudicators. Future versions of the adjudication guide may need to include clarifying language. The adjudication worksheet questions that resulted in the most discordance between 2 reviewers were “What was the overall global clinical outcome?” ([Supplementary Figure 1](#)) and “What was the clinical impression at day 7–8 of the pneumonia episode?” The discordance may be due to the fact that some clinical interpretation was required when choosing the best response to these questions. For example, a day 7–8 end point of

Interobserver Agreement by Year

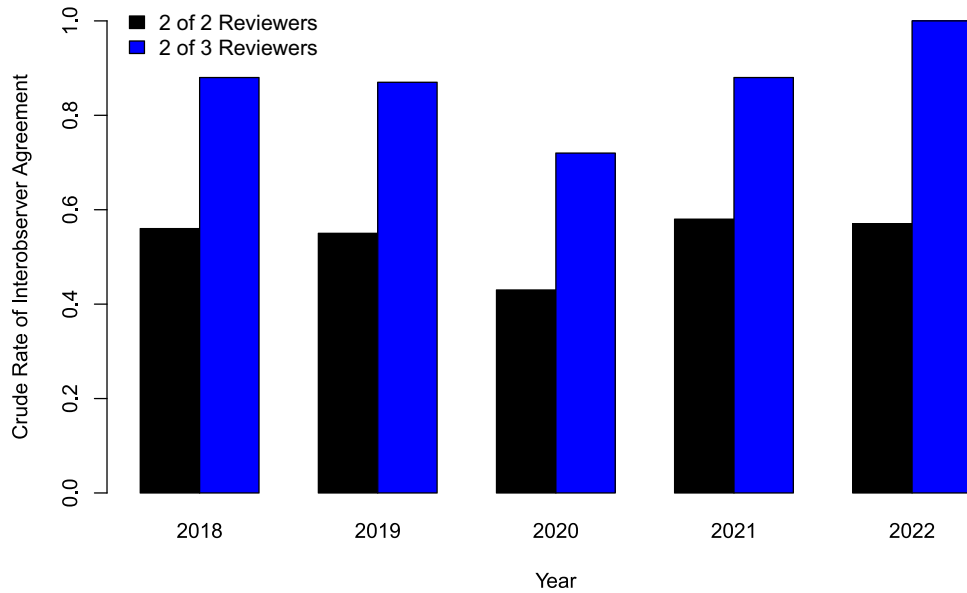


Figure 2. Crude rates of interobserver agreement by year. When interobserver agreement was defined as consensus between 2 of 2 adjudicators, the crude rates of interobserver agreement did not exceed 0.6. Interobserver agreement, defined as agreement between 2 of 3 adjudicators, increased the rate of agreement.

Table 2. Rates of Interobserver Agreement Based on Characteristics of the First Pneumonia Episode or the Presence of Multiple Episodes of Pneumonia

Etiology of Pneumonia Episode	Method to Achieve Consensus			Percentage of Cases That Achieved Agreement Without Need for Committee Review
	No. of Cases With Agreement Between 2 of 2 Adjudicators (%)	No. of Cases With Agreement Between 2 of 3 Adjudicators (%)	No. of Cases Needing Committee Review (%)	
Bacterial etiology defined (n = 150)	83 (55.3)	43 (28.7)	24 (16.0)	84.0
Bacterial and viral coinfection (n = 80)	32 (40.0)	22 (27.5)	26 (32.5)	67.5
Culture-negative (n = 106)	57 (53.8)	36 (34.0)	13 (12.3)	87.8
Nonpneumonia (n = 57)	28 (49.1)	16 (28.1)	13 (22.8)	77.2
Viral etiology defined (n = 142)	57 (40.1)	47 (33.1)	38 (26.8)	73.2
Indeterminate (n = 3)	2 (66.7)	1 (33.3)	0 (0)	100
Multiple pneumonia episodes (n = 120)	19 (15.8)	39 (32.5)	62 (51.7)	48.3

The "multiple pneumonia episodes" category consists of patients with 2 or more pneumonia episodes during a single hospitalization. Note that the "pneumonia etiology" section in [Table 1](#) lists all cases included in this study (n = 593). [Table 2](#) only lists the cases that were adjudicated by at least 2 adjudicators.

cure requires the signs and symptoms of pneumonia to be improving. If a patient's oxygen requirement changed from 60% fraction of inspired oxygen (FiO₂) to 50% FiO₂, some clinicians may consider this an improvement while others may not consider this a meaningful improvement. Similarly, if an initial respiratory culture grew >100 000 cfu/mL of a bacterium, and a

repeat culture on day 7 grew 100 cfu/mL of the same bacterium, some clinicians may call this persistent infection while others may consider this to be a cure.

Other factors were associated with relatively lower rates of consensus. The presence of multiple episodes of pneumonia during a hospitalization and an initial episode of bacterial/viral

Hospital Discharge Disposition Grouped By Day 7-8 Clinical End Point

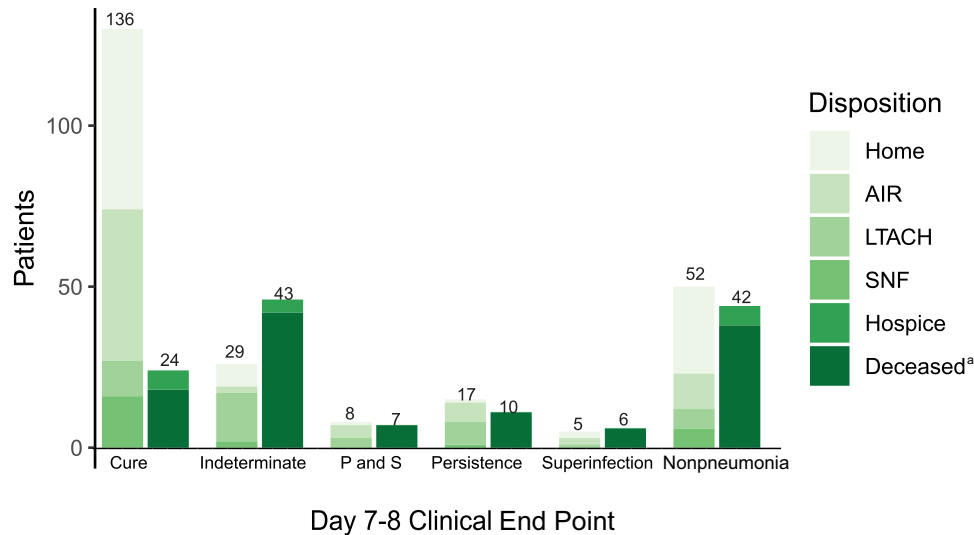


Figure 3. The association between day 7–8 clinical end points (cure, indeterminate, persistent, or superinfection) for the initial episode of bacterial or bacterial–viral coinfection pneumonia and discharge disposition at the end of the hospitalization. Favorable discharge dispositions are home, AIR, LTACH, and SNF. Unfavorable discharge dispositions are hospice or deceased. ^aDeceased includes patients who underwent lung transplantation for refractory respiratory failure during their hospitalization. Abbreviations: AIR, acute inpatient rehabilitation; LTACH, long-term acute care hospital; P and S, persistence and superinfection; SNF, skilled nursing facility.

coinfection pneumonia decreased the rates of interobserver agreement. The latter finding likely reflects the increased complexity of assessing coinfection, particularly when microbiologic resolution for 1 pathogen and persistence of the other exists. Conversely, interobserver agreement was high for culture-negative cases. One explanation may be that a culture-negative case may already be responding to antibiotics and more consensus among adjudicators was found for cure. Alternatively, the reason may simply be that fewer variables need adjudication in culture-negative cases, for example, appropriateness of antibiotics or a decrease in colony-forming units of an organism on serial BALs.

There are multiple strengths to our adjudication protocol. First, the extent of our protocol minimizes the probability that 2 observers could provide the same responses to all of the questions by chance. The adjudicators were all pulmonary and critical care physicians, with extensive clinical experience in pneumonia treatment. The defined end points correlated with objective and clinically meaningful discharge dispositions and were consistent with consensus guidelines. Our starting point was pneumonia diagnosed by BAL criteria, minimizing contamination of the assessment by cases that were not pneumonia. Despite this level of rigor, we found poor interobserver agreement between 2 adjudicators and only moderate improvement with an additional independent review.

Our study also has limitations. We adjudicated clinical outcomes independent of the physicians treating the patient, although we did review all clinical documentation from the

bedside teams during our case reviews. We did not have access to additional pertinent clinical information not readily available from the electronic health record. Another limitation is the external validity of our findings, which may depend on the experience of the adjudicators comprising the panel. In our study, key data incorporated into adjudications came from BAL fluid results, which may limit the generalizability of our protocol in centers where BAL sampling is used less often in pneumonia diagnosis and management. Despite these limitations, for patients with both bacterial and viral-only pneumonia, a day 7–8 adjudicated clinical impression of cure was associated with discharge from the hospital alive. This important association between adjudicated end points and objective outcomes supports the construct validity of our procedures for the initial bacterial episode of pneumonia. Use of an adjudication committee similar to ours could be considered to optimize use of clinical response as an end point for registration trials and cohort studies of pneumonia.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

The authors would like to acknowledge Nikolay Markov for programming assistance, Elizabeth Gray with the Biostatistics Collaboration

Center for input on the analysis of interobserver agreement, and the NU SCRIPT Study Investigators.

Author contributions. C.I.P., C.A.G., J.M.W., R.G.W., and B.D.S. wrote and edited the manuscript and performed data analysis. C.I.P., C.A.G., J.M.W., J.M.K., R.G.W., and B.D.S. adjudicated the episodes of pneumonia. J.B. reviewed the cohort and identified the questions leading to interobserver disagreement. R.G.W. led the conceptualization and methodology of the adjudication process. H.K.D., A.D., K.C., and N.B. prepared the forms for adjudication, performed data entry, and organized the committee reviews. All authors confirm that they had full access to all the data in the study and accept responsibility for submission for publication. All authors read and approved the final draft of the manuscript.

Patient consent. All participants, or legally authorized representatives for the participant, in this study provided written consent for data to be published. Approval for this study was obtained from the Northwestern University Institutional Review Board (#STU00204868).

Financial support. This work was supported by the National Institutes of Health, National Institute of Allergy and Infectious Diseases (AI135964). C.A.G. is supported by NIH awards T32HL076139 and F32HL162377. J.M.K. is supported by NIH/NHLBI K23HL146890. R.G.W. is supported by NIH awards U19AI135964, U01TR003528, P01HL154998, R01HL14988, R01LM013337. B.D.S. is supported by NIH awards R01HL149883, R01HL153122, P01HL154998, P01AG049665, and U19AI135964.

Potential conflicts of interest. B.D.S. holds US patent 10,905,706, “Compositions and methods to accelerate resolution of acute lung inflammation,” and serves on the Scientific Advisory Board of Zoe Biosciences, for which he holds stock options. C.I.P., C.A.G., J.B., J.M.W., J.M.K., H.K.D., A.D., K.C., N.B., and R.G.W. declare no conflicts of interest. All other authors report no potential conflicts.

References

1. US Food and Drug Administration. *Guidance for Industry: Community Acquired Bacterial Pneumonia: Developing Drugs for Treatment*. US Food and Drug Administration; 2014. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM123686.pdf>. Accessed April 26, 2023.
2. Meduri GU, Mauldin GL, Wunderink RG, et al. Causes of fever and pulmonary densities in patients with clinical manifestations of ventilator-associated pneumonia. *Chest* 1994; 106:221–35.
3. Fagon JY, Chastre J, Wolff M, et al. Invasive and noninvasive strategies for management of suspected ventilator-associated pneumonia. A randomized trial. *Ann Intern Med* 2000; 132:621–30.
4. Pickens CO, Gao CA, Cuttica MJ, et al. Bacterial superinfection pneumonia in patients mechanically ventilated for COVID-19 pneumonia. *Am J Respir Crit Care Med* 2021; 204:921–32.
5. Ioanas M, Ferrer M, Cavalcanti M, et al. Causes and predictors of nonresponse to treatment of intensive care unit-acquired pneumonia. *Crit Care Med* 2004; 32: 938–45.
6. Talbot GH, Das A, Cush S, et al. Evidence-based study design for hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia. *J Infect Dis* 2019; 219:1536–44.
7. Hospital-Acquired Bacterial Pneumonia and Ventilator-Associated Bacterial Pneumonia: Developing Drugs for Treatment. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/hospital-acquired-bacterial-pneumonia-and-ventilator-associated-bacterial-pneumonia-developing-drugs>. Accessed April 25, 2023.
8. Held C. When do we need clinical endpoint adjudication in clinical trials? *Ups J Med Sci* 2019; 124:42–5.
9. Kahan BC, Feagan B, Jairath V. A comparison of approaches for adjudicating outcomes in clinical trials. *Trials* 2017; 18:266.
10. Meah MN, Denvir MA, Mills NL, Norrie J, Newby DE. Clinical endpoint adjudication. *Lancet* 2020; 395:1878–82.

11. Agnelli G, Becattini C, Meyer G, et al. Apixaban for the treatment of venous thromboembolism associated with cancer. *N Engl J Med* 2020; 382:1599–607.
12. Seltzer JH, Turner JR, Geiger MJ, et al. Centralized adjudication of cardiovascular end points in cardiovascular and noncardiovascular pharmacologic trials: a report from the Cardiac Safety Research Consortium. *Am Heart J* 2015; 169:197–204.
13. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018; 198:e44–68.
14. Gainor JF, Curigliano G, Kim DW, et al. Pralsetinib for RET fusion-positive non-small-cell lung cancer (ARROW): a multi-cohort, open-label, phase 1/2 study. *Lancet Oncol* 2021; 22:959–69.
15. Del Prato S, Kahn SE, Pavo I, et al. Tirzepatide versus insulin glargine in type 2 diabetes and increased cardiovascular risk (SURPASS-4): a randomised, open-label, parallel-group, multicentre, phase 3 trial. *Lancet* 2021; 398:1811–24.
16. Bielicki JA, Stöhr W, Barratt S, et al. Effect of amoxicillin dose and treatment duration on the need for antibiotic re-treatment in children with community-acquired pneumonia: the CAP-IT randomized clinical trial. *JAMA* 2021; 326:1713–24.
17. Cook D, Walter S, Freitag A, et al. Adjudicating ventilator-associated pneumonia in a randomized trial of critically ill patients. *J Crit Care* 1998; 13:159–63.
18. Klein Klouwenberg PM, Ong DS, Bos LD, et al. Interobserver agreement of Centers for Disease Control and Prevention criteria for classifying infections in critically ill patients. *Crit Care Med* 2013; 41:2373–8.
19. Walker M, Hamer D, Musso M, O’Neal C, Thomas C, O’Neal H Jr. Accuracy of physician adjudication of infection in patients with systemic inflammatory response syndrome (SIRS). *Open Forum Infect Dis* 2018; 5:S633–4.
20. Dela Cruz CS, Evans SE, Restrepo MI, et al. Understanding the host in the management of pneumonia. An official American Thoracic Society workshop report. *Ann Am Thorac Soc* 2021; 18:1087–97.
21. Gao CA, Bailey JI, Walter JM, et al. Bronchoscopy on intubated patients with COVID-19 is associated with low infectious risk to operators. *Ann Am Thorac Soc* 2021; 18:1243–6.
22. Walter JM, Helmin KA, Abdala-Valencia H, Wunderink RG, Singer BD. Multidimensional assessment of alveolar T cells in critically ill patients. *JCI Insight* 2018; 3:e123287.
23. Grant RA, Morales-Nebreda L, Markov NS, et al. Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. *Nature* 2021; 590:635–41.
24. Weiss E, Zahar JR, Alder J, et al. Elaboration of consensus clinical endpoints to evaluate antimicrobial treatment efficacy in future hospital-acquired/ventilator-associated bacterial pneumonia clinical trials. *Clin Infect Dis* 2019; 69:1912–8.
25. Timsit JF, de Kraker MEA, Sommer H, et al. Appropriate endpoints for evaluation of new antibiotic therapies for severe infections: a perspective from COMBACTE’s STAT-Net. *Intensive Care Med* 2017; 43:1002–12.
26. Weiss E, Essaïed W, Adrie C, Zahar JR, Timsit JF. Treatment of severe hospital-acquired and ventilator-associated pneumonia: a systematic review of inclusion and judgment criteria used in randomized controlled trials. *Crit Care* 2017; 21: 162.
27. Kalil AC, Metersky ML, Klompas M, et al. Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin Infect Dis* 2016; 63:e61–111.
28. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med* 2019; 200:e45–67.
29. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012; 22:276–82.
30. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. 2001.
31. Veldhoen RA, Howes D, Maslove DM. Is mortality a useful primary end point for critical care trials? *Chest* 2020; 158:206–11.
32. Wunderink RG, Laterre PF, Francois B, et al. Recombinant tissue factor pathway inhibitor in severe community-acquired pneumonia: a randomized trial. *Am J Respir Crit Care Med* 2011; 183:1561–8.