PAPER ID: PMC7060632

TITLE:
Predicting the microbial cause of community-acquired pneumonia: can physicians or a data-driven method differentiate viral from bacterial pneumonia at patient presentation?

ABSTRACT:
BACKGROUND: Community-acquired pneumonia (CAP) requires urgent and specific antimicrobial therapy. However, the causal pathogen is typically unknown at the point when anti-infective therapeutics must be initiated. Physicians synthesize information from diverse data streams to make appropriate decisions. Artificial intelligence (AI) excels at finding complex relationships in large volumes of data. We aimed to evaluate the abilities of experienced physicians and AI to answer this question at patient admission: is it a viral or a bacterial pneumonia? METHODS: We included patients hospitalized for CAP and recorded all data available in the first 3-h period of care (clinical, biological and radiological information). For this proof-of-concept investigation, we decided to study only CAP caused by a singular and identified pathogen. We built a machine learning model prediction using all collected data. Finally, an independent validation set of samples was used to test the pathogen prediction performance of: (i) a panel of three experts and (ii) the AI algorithm. Both were blinded regarding the final microbial diagnosis. Positive likelihood ratio (LR) values > 10 and negative LR values < 0.1 were considered clinically relevant. RESULTS: We included 153 patients with CAP (70.6% men; 62 [51–73] years old; mean SAPSII, 37 [27–47]), 37% had viral pneumonia, 24% had bacterial pneumonia, 20% had a co-infection and 19% had no identified respiratory pathogen. We performed the analysis on 93 patients as co-pathogen and no-pathogen cases were excluded. The discriminant abilities of the AI approach were low to moderate (LR+ = 2.12 for viral and 6.29 for bacterial pneumonia), and the discriminant abilities of the experts were very low to low (LR+ = 3.81 for viral and 1.89 for bacterial pneumonia). CONCLUSION: Neither experts nor an AI algorithm can predict the microbial etiology of CAP within the first hours of hospitalization when there is an urgent need to define the anti-infective therapeutic strategy.

Background:
The World Health Organization (WHO) estimates that due to antimicrobial resistance, bacterial infections will outcompete any cause of death by 2050 [1], meaning that there is an urgent need for new strategies to improve antibiotic treatments. The Agency for Healthcare Research and Quality (AHRQ) Safety Program for Improving Antibiotic Use recently proposed a structured approach to improve antibiotic decision making by clinicians, which emphasizes the 4 critical time points in antibiotic prescribing [2, 3]. The first time point of this organized approach requires the physician to ask: "Does this patient have an infection that requires antibiotics?". This question aims to remind the clinician to synthesize all relevant patient information to determine the likelihood of an infection that requires antibiotic therapy. The questionable ability of physicians to answer this first question properly in the context of pneumonia was the impetus for this study. Community-acquired pneumonia (CAP) is a major global healthcare burden associated with significant morbidity, mortality and costs [4–9]. Identifying the etiology of CAP is an utmost priority for its management and treatment decisions [10]. Although the range of pathogens that may be involved in these cases is broad, physicians must at least determine whether a bacterial or a viral pathogen (or both) is causing the pneumonia to determine if antibiotic treatment is appropriate. Whether the etiology of CAP is viral or bacterial should be determined based on the patient interview, clinical symptoms and signs, biological findings and radiological data from the very first hours of the patient's presentation (a time when microbiological findings are typically not yet available). Physicians must use the knowledge obtained from their routine practice and medical education to make sense of these diverse data input streams, triage the resulting complex dataset, and make appropriate decisions. A growing body of research has recently suggested that difficulties in accessing, organizing, and using a substantial amount of data could be significantly ameliorated by use of emerging artificial intelligence (AI)-derived methods, which are nowadays applied in diverse fields including biology, computer science and sociology [11]. AI excels at finding complex relationships in large volumes of data and can rapidly analyze many variables to predict outcomes of interest. In the context of CAP in intensive care units (ICUs), where information are particularly diverse, we wondered if an AI data-driven approach to reducing the

medical complexity of a patient could allow us to make a better hypothesis regarding the microbial etiology at the patient's presentation.

The aim of our study was to evaluate and compare the abilities of experienced physicians and a data-driven approach to answer this simple question within the first hours of a patient's admission to the ICU for CAP: is it a viral or a bacterial pneumonia?

## Step 1: patient data collection ::: Methods:

Prospective data collection was conducted in a single center over an 18-month period. The study complied with French law for observational studies, was approved by the ethics committee of the French Intensive Care Society (CE SRLF 13–28), was approved by the Commission Nationale de l'Informatique et des Libertés (CNIL) for the treatment of personal health data. We gave written and oral information to patients or next-of-kin. Patients or next-of-kin gave verbal informed consent, as approved by the ethic committee. Eligible patients were adults hospitalized in ICU for CAP. Pneumonia was defined as the presence of an infiltrate on a chest radiograph and one or more of the following symptoms: fever (temperature ≥ 38.0 °C) or hypothermia (temperature < 35.0 °C), cough with or without sputum production, or dyspnea or altered breath sounds on auscultation. Community-acquired infection was defined as infection occurring within 48 h of admission. Cases of pneumonia due to inhalation or infection with pneumocystis, pregnant women and patients under guardianship were not included. Cases with $PaO_2$ ≥ 60 mmHg in ambient air or with the need for oxygen therapy ≤4 L/min or without mechanical ventilation (invasive or non-invasive) were not included.

Baseline patient information was collected at case presentation through in-person semi-structured interviews with patients or surrogates (see Supplementary Table 1). Observations from the physical examination at presentation, including vital signs and auscultation of the lungs, were recorded. Findings of biological tests done at presentation (within the first three-hour period) were also recorded (hematology and chemistry tests), as were findings from chest radiography. Two physicians interpreted chest x-rays; a third physician reviewed the images in cases of disagreements in interpretation.

Microbiological investigations included blood cultures, pneumococcal and legionella urinary antigen tests, bacterial cultures and multiplex PCR RespiFinder SMART 22® (PathoFinder B.V., Oxfordlaan, Netherlands) analyses on respiratory fluids (sputum and/or nasal wash and/or endotracheal aspirate and/or bronchoalveolar lavage [BAL]).

## Clinician predictions ::: Step 2: clinician and data-driven predictions of microbial etiology ::: Methods:

An external three member expert panel reviewed the work dataset to familiarize themselves with the dataset containing the patient characteristics. Then, the experts were asked to predict the microbial etiologies in the external validation dataset (Fig. 1a). The clinicians had to answer the question: is it a viral or a bacterial pneumonia? They were also asked to give a confidence index regarding the accuracy of their answer: 1 (very low), 2 (low), 3 (moderate), 4 (high). Agreement of at least two of the three experts was required for the final predicted etiology.

## Data-driven approach predictions ::: Step 2: clinician and data-driven predictions of microbial etiology ::: Methods:

The data were analyzed using an AI method (Fig. 1b) involving a logistic regression analysis using forward stepwise inclusion. This method was employed to optimize the ability of the algorithm to distinguish viral and bacterial pneumonia based on the combination of parameters available in the work dataset. All available data were thus included in the model, regardless of the data type. Qualitative data were processed as binary information (i.e. influenza immunization: present "1", absent "0"). Raw data were provided for quantitative values (no cut-offs defined). We built the predictive mathematical model from the work dataset using the Random Forest method and Leave-One-Out Cross-Validation. We started by determining the most relevant item to use through a variable selection procedure using the Random Forest method and the Mean Decrease in Gini criterion (value 0.75). Then, the population in the work dataset was randomly separated into two independent datasets: 80% of cases were assigned to the training set and 20% were assigned to the test set. N models with bootstrap resampling (with N = 25) were performed on the training set and validated on the test set. The model providing the best prediction criteria was selected, and the final model was built from the entire work dataset. Finally, an independent validation set of samples was used to test the pathogen prediction performance of the AI algorithm. To decipher the relative importance of clinical versus biological/radiological variables in

the predictions, we generated three algorithms built from different parameters of the work dataset: (i) clinical variables only, (ii) biological and radiological variables only, and (iii) all variables. For each parameter tested, the area under the ROC curve (AUC) was calculated, and the best cutoff value that yielded the highest accuracy was determined along with the sensitivity and specificity.

Statistical analysis ::: Methods:
We compared the concordance between the predictions and the final microbial etiologies for the experts and for the algorithm and calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and likelihood ratios (LRs) for the predictions [12]. Given the importance of this diagnostic prediction in the patient's therapeutic management, we determined that the discriminant properties should be "high" (LR + > 10 and/or LR- < 0.1) for the prediction to be considered useful for clinical practice [13, 14]. Table 1 summarizes the LR cutoff values defining the discriminant properties of the predictions [13]. Quantitative data are reported as the median value and interquartile range (IQR). Statistical analyses were done with JMP software (SAS, version 7.2).

Data-driven approach predictions ::: Clinician predictions ::: Results:
Predictions by the data-driven algorithms generated from clinical data alone resulted in an ROC curve with a corresponding AUC of 0.72. Predictions by the data-driven algorithms generated from biological and radiological variables data alone resulted in an ROC curve with an AUC of 0.81. Finally, predictions generated from the dataset that included all data sources outperformed the other algorithms and resulted in an ROC curve with an AUC of 0.84 (Table 3, Fig. 3). This model based on the more inclusive dataset was considered the final model for comparison with the expert panel. The final algorithm made predictions with a sensitivity of 0.57, specificity of 0.91, PPV of 0.80 and NPV of 0.77 for the diagnosis of bacterial pneumonia. The LR+ for diagnosing a viral pneumonia was 2.12, and the corresponding LR- was 0.16. The LR+ for diagnosing a bacterial pneumonia was 6.29, and the corresponding LR- was 0.47. Consequently, the discriminant abilities of the data-driven algorithm to distinguish viral and bacterial etiologies for pneumonia were categorized as low to moderate (according to defined cutoff values for the interpretation of likelihood ratios, see Table 1).

Discussion:
Addressing antimicrobial resistance requires investment in several critical areas, the most pressing of which is the ability to make rapid diagnoses to promote appropriate anti-infective therapeutics and limit unnecessary antibiotic use. Here, we set up a pilot study and demonstrated that neither experts nor a mathematical algorithm could accurately predict the microbial etiology of severe CAP within the first 3 hours of hospitalization when there is an urgent need to define the appropriate anti-infective therapeutic strategy.
We encoded all information available in the first 3 hours after admission for a large cohort comparable with other published cohorts in terms of the distribution of causal microbial pathogens, patient characteristics and severity of disease [15–17]. We demonstrated that experienced clinicians synthesizing all this information failed to adequately answer the question: "is it a viral or a bacterial pneumonia?", as the discriminant ability between the two diagnoses was considered low. We interpreted our results mainly based on the calculation of likelihood ratios, as recommended for reports of a diagnostic test for an infectious disease [12]. Likelihood ratios incorporate both sensitivity and specificity and, unlike predictive values, do not vary with prevalence, making them good statistical tools to facilitate translation of knowledge from research to clinical practice [12]. In parallel, we designed a data-driven approach. Different AI methods were available; we selected the Random Forest method because it is one of the most efficient strategies for providing a predictive algorithm in this context [18–21]. Importantly, the final algorithm was tasked with providing predictions for a novel population independent of the dataset used for the algorithm construction. The discriminant abilities of the AI approach restricted to the binary choice "viral" or "bacterial" were superior to those of experts but still considered low or moderate and were ultimately insufficient to provide an indisputable therapeutic decision. It is important to emphasize that we chose a high cutoff value for determining the discriminant ability of the AI approach (LR+ > 10, LR- < 0.1); this choice was made for two reasons. First, in this proof-of-concept study, we did not analyze co-infections and restricted the possible choices to a binary prediction. Because we reduced the complexity of the cases, we expected high predictive

performances. Second, the goal of this study was not a prediction of outcomes (e.g., ICU length of stay, mortality), which are informative but do not determine patient management; it was to provide a clear and immediate medical decision: whether or not to prescribe antibiotics. The immediate clinical consequences in this situation demand a high predictive performance. Still, it is important to highlight that the machine learning method we developed achieved an AUC of 0.84, which is superior or at least equal to AUC values usually observed for predictive mathematical models developed for the ICU environment. For instance, the Systemic Inflammatory Response Syndrome (SIRS) criteria, the Simplified Acute Physiology Score II (SAPS II) and the Sequential Organ Failure Assessment (SOFA) have AUC values of 0.61, 0.70 and 0.73, respectively, for identifying sepsis [22]. An AI Sepsis Expert algorithm for early prediction of sepsis has been engineered and achieved AUC values ranging from 0.83–0.85 according to the time of the prediction. An AI method for predicting prolonged mechanical ventilation achieved an AUC of 0.82 [23].

How can it be that AI or machine-learning predictive algorithms that can already automatically drive cars or successfully understand human speech failed to predict the microbial cause of pneumonia accurately? First, having data of excellent quality is critical for the success of AI predictions. The ICU environment is data-rich, providing fertile soil for the development of accurate predictive models [24], but it is also a challenging environment with heterogeneous and complex data. In our study, the data that fueled the AI method were from a real-world data source. It is probably more difficult to create a consistent data format when merging data from interviews, patient examinations, biological and radiological information than when using datasets from the insurance or finance industries. Additionally, data arising from patient examinations and interviews are still strictly dependent on the physician's skill and experience. Finally, although we hypothesized that the AI capabilities would exceed human skills and make accurate predictions when physicians cannot, we must also consider the null hypothesis: viral and bacterial pneumonias share the same characteristics and cannot be distinguished based on initial clinical, biological or radiological parameters. The dividing lines between the signs and symptoms of a viral versus a bacterial infection could be too blurry to permit the two diagnoses to be discerned without microbial analyses.

Our results emphasize the need to use a rapid turnaround time system for the accurate identification of respiratory pathogens from patient specimens. Utilizing rapid molecular respiratory panel assays may increase the likelihood of optimal treatment of acute respiratory infections [25–29]. However, antibiotic consumption was not reduced by the use of a molecular point-of-care strategy in adults presenting with acute respiratory illness in a large randomized controlled trial [28]. It seems that we are experiencing a switch in perspectives regarding microbial diagnoses of respiratory infections: physicians are used to dealing with an absence of information, but they will likely be overloaded with information in the near future [30]. The positive detection of respiratory viruses may or may not be useful for the immediate management of a patient [31]. Thus, the development of molecular point-of-care analysis techniques will not lessen the usefulness of our AI strategy. On the contrary, we believe that AI could be a great help in dealing with information overload, which could soon be a common problem. AI methods should not be viewed as ways to replace human expertise but rather as catalysts that accelerate human expertise–based analyses of data. AI methods can assist–rather than replace–in clinical decision-making by transforming complex data into more actionable information. Further studies are needed to assess if AI system integrated with point-of-care rapid molecular respiratory panel assays could be a useful addition for the clinician. Ultimately, randomized controlled trial should determine the effect of this strategy on the decision making regarding antibiotic use.

Our study should be interpreted in the context of several limitations. First, this was a proof-of-concept study, and we excluded cases of CAP with mixed etiology or without microbiological documentation. Consequently, the results were obtained from artificially dichotomized situations (viral or bacterial pneumonia, 93 patients in total) and cannot be directly extrapolated to real-life practice. Moreover, we did not include cases of acute pneumonia with non-infectious origins. Second, the experts were asked to make their predictions based on case reports exhaustively described in Excel files. They did not have the opportunity to interview or directly examine the patients themselves. Furthermore, the experts' predictions were not performed in "real-life" situation. This could have affected the experts' predictive performance. Third, we cannot rule out the possibility that some bacterial or viral pneumonia cases were misdiagnosed. We relied on state-of-the-art methods for microbial discovery, but it is possible that our current technology is sometimes suboptimal for detecting respiratory microbial pathogens.

Conclusion:
Neither a panel of experts nor a data-driven approach could accurately distinguish viral from bacterial pneumonia within the first hours of patient admission in ICU for CAP. The heterogeneous and complex data generated in the ICU environment are likely difficult to use to generate an AI algorithm with a high predictive quality. The results of our pilot study at least highlight that we should not treat machine learning and data science as crystal balls for making predictions and automating decision-making; we should rather use these techniques to more critically examine all available information and enhance existing human expertise.