

Generating Evidence for Chronic Obstructive Pulmonary Disease (COPD) Clinical Guidelines Using EHR Data

Amber M. Johnson, MS^{1,2}, Marvi Bikak, MD³, Paul M. Griffin, PhD¹, and Mohammad Adibuzzaman, PhD¹

¹Regenstrief Center for Healthcare Engineering

²Department of Computer Science, Purdue University, West Lafayette, Indiana, USA

³Indiana University School of Medicine, Indianapolis, Indiana, USA

ABSTRACT

Objectives: The aim of this research was to develop data-driven models using electronic health records (EHRs) to conduct clinical studies for predicting clinical outcomes through probabilistic analysis that considers temporal aspects of clinical data. We assess the efficacy of antibiotics treatment and the optimal time of initiation for in-hospitalized diagnosed with acute exacerbation of COPD (AECOPD) as an application to probabilistic modeling.

Materials and Methods: We developed a semi-automatic Markov Chain Monte Carlo (MCMC) modeling and simulation approach that encodes clinical conditions as computable definitions of health states and exact time duration as input for parameter estimations using raw EHR data. We applied the MCMC approach to the MIMIC-III clinical database, where ICD-9 diagnosis codes (491.21, 491.22, and 494.1) were used to identify data for 697 AECOPD patients of which 25.9% were administered antibiotics.

Results: The average time to antibiotic administration was 27 hours, and 32% of patients were administered *vancomycin* as the initial antibiotic. The model simulations showed a 50% decrease in mortality rate as the number of patients administered antibiotics increased. There was an estimated 5.5% mortality rate when antibiotics were initially administered after 48 hours vs 1.8% when antibiotics were initially administered between 24 and 48 hours.

Discussion: Our findings suggest that there may be a mortality benefit in initiation of antibiotics early in patient with severe respiratory failure in settings of COPD exacerbations warranting an ICU admission.

Conclusion: Probabilistic modeling and simulation methods that considers temporal aspects of raw clinical patient data can be used to adequately generate evidence for clinical guidelines.

Keywords: Markov chain, COPD, clinical guidelines, Monte Carlo, EHR

BACKGROUND AND SIGNIFICANCE

Clinical guidelines are systematically developed documents that support patient care for specific clinical circumstances.¹ They are typically developed through either evaluation of the “best” available evidence generated via meta-analyses, randomized controlled trials, or expert opinion.² Such evidence-based sources supplement physician experience, which is typically based on individual clinical experience and discussion with other physicians.^{3,4} In practice, physicians may not follow the clinical guidelines due to a lack of awareness, conflicting recommendations from the presence of different clinical conditions in a patient, or limited recommendations due to lack of evidence.⁵ For example, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) national clinical practice guidelines⁶ have been used worldwide by healthcare professionals for treating and managing chronic obstructive pulmonary disease (COPD). COPD refers to a group of diseases including chronic bronchitis and emphysema, that affects over 16 million persons in the U.S., and is the third leading cause of death.⁷ GOLD recommends antibiotic therapy for patients with severe acute exacerbation of COPD (AECOPD), a sudden worsening of symptoms,⁸ as they can shorten hospital length-of-stay (LOS) and decrease mortality. Though several studies⁹ have been conducted to assess the short-term

and long-term efficacy of antibiotics for AECOPD none have explored the timing of when antibiotics are administered to AECOPD patients in the intensive care unit (ICU). Hence, there are no guidelines or recommendations for the initial timing for administering antibiotics.

AECOPD often requires multiple hospitalizations over a patient's lifetime. During these hospitalizations physician care decisions are collected in patients' electronic health records (EHRs) and clinical notes. Each visit is a source of patient information (e.g., diagnosis, treatment, outcome) about clinical events that effect their health state during their hospital encounters (i.e., care pathway). While this information is crucial for treating patients, there is often not a standardized method of processing and documenting the information.¹⁰ Moreover, data can be challenging to retrieve and properly interpret due to its high volume and unstructured nature.¹¹ Existing software systems that analyze EHR data^{12,13} do not allow for longitudinal processes like chronic disease progression to be observed directly, nor do they support analysis of how patients transition from one health state to another as a direct cause of an intervention (e.g., drug administration, oxygen therapy, surgery).

Despite the limitations of current tools, there are examples of retrospective data in EHRs being used for evidence generation through longitudinal analysis.¹³⁻¹⁵ This process helps identify medical trends for a patient population and improve the overall quality of care, which is particularly important for chronic conditions. Markov models are one approach that have been used to estimate healthcare costs, utilization, and disease progression over time.¹⁶⁻¹⁸ They provide support for evaluating uncertainties in decision making by approximating patient transitions through a set of "health states", each of which corresponds to a clinical event.¹⁹ While using Markov chains has been useful for chronic conditions such as COPD, directly using EHR data in Markov models presents certain challenges. In particular, Markov chains assume a one-step time-invariant transition. EHRs, however, contain highly-dimensional, time-variant, clinical data observed at irregular time intervals. Hence, the nature of EHR data limits the ability to represent granular timing of the transitions, which can lead to process misrepresentation in the model. Thus, EHR data points must be transformed and common modeling parameter estimation techniques refined to develop Markov models that properly capture temporal characteristics.

This research presents an approach that systematically captures the effect of interventions during medical encounters, and hence supports the development of clinical guidelines. A semi-automatic Markov Chain Monte Carlo (MCMC) modeling and simulation approach that encodes clinical conditions as computable definitions of health states using raw EHR data is described. We further demonstrate how data-driven models can be used to conduct clinical studies to predict clinical outcomes through probabilistic analysis. While our approach is general, we present an illustration by considering the efficacy of antibiotics treatment for in-hospitalized AECOPD patients. A model of the pathway of AECOPD is generated for hospitalized patients over time by curating retrospective EHR data collected during critical care encounters. Recommendations from the national GOLD guidelines⁶ and physicians were incorporated to define model components. We estimated outcomes for two research questions related to antibiotic treatment for AECOPD patients, namely: i) the impact of antibiotic administration on in-hospital death and ii) the impact on in-hospital death based on the initial timing of antibiotic administration. We performed Monte Carlo simulations, which we validated by comparing the estimated model outcomes to actual outcomes from the EHR data, and conducted sensitivity analyses.

MATERIALS AND METHODS

Data

Data was used from MIMIC-III (Medical Information Mart for Intensive Care),^{20,21} a curated database from Beth Israel Deaconess Medical Center, comprised of nearly 40,000 distinct patients and nearly 50,000 hospital admissions for adult patients in critical care units from 2001 to 2012. Each patient visit was associated with multiple diagnoses, which were labeled by the International Classification of Diseases, Ninth Revision (ICD-9) diagnosis codes and ordered by priority. We defined a primary diagnosis of AECOPD as a hospital admission having at least one of the following ICD-9 codes as either primary, secondary or tertiary diagnosis:^{22,23} 491.21, 491.22, and 494.1. We limited our data extraction to these three codes since we wanted to analyze patients with acute respiratory failure (ARF) secondary to AECOPD primarily, without other confounding etiologies that may be present. Based on this selection criteria (Figure 1), we identified 697 unique AECOPD patients with at least one ICU admission.

GOLD guidelines suggest that severe exacerbations requiring hospitalization may be associated with ARF,⁶ a build up

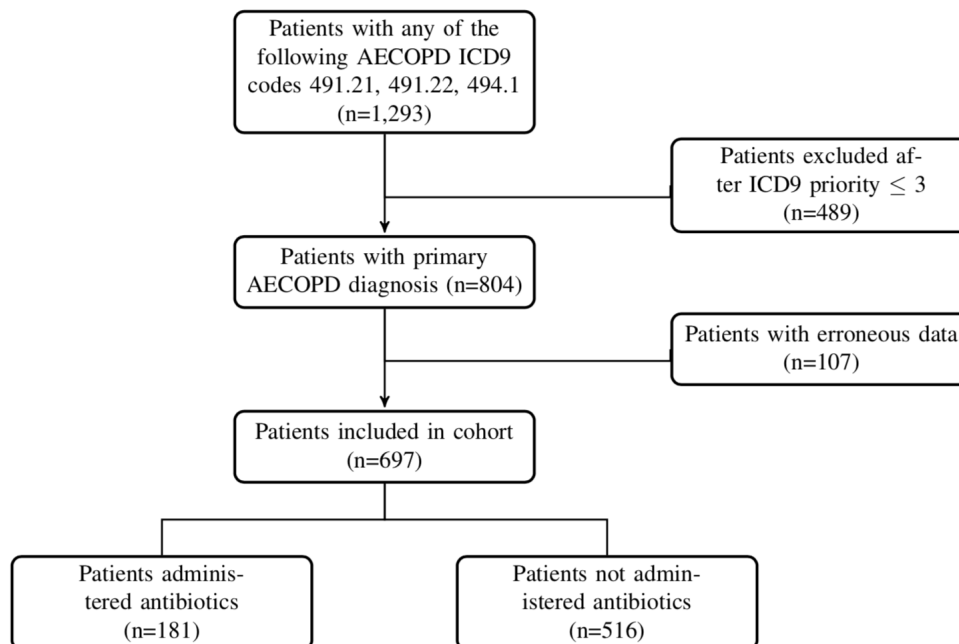


Figure 1: Cohort selection criteria for AECOPD patients administered and not administered antibiotics in the MIMIC-III database.

in the air sac of the lungs that inhibits the release of oxygen into the blood.⁸ There are two types of ARF, hypoxemic (type 1) and hypercapnic (type 2), which can be identified by changes in clinical variables that GOLD recommends for classifying exacerbations⁶ are: partial pressure of oxygen (PaO_2), partial pressure of carbon dioxide (PaCO_2), and pH. We used recommendations from GOLD and consulted with COPD experts to identify and define health states on the basis of established clinical criteria⁶ in terms of these variables (Table 1). We extracted the time that a clinical measurement or event was recorded and determined the health state for the patient at that time to create temporal paths, represented by pairs of ordered time-stamps and state sequences.

We performed an exhaustive search within the group of AECOPD patients extracted from MIMIC-III for antibiotics used to treat COPD. We then extracted information regarding administration times of those antibiotics for the subset of patients by preparing a comprehensive list of antibiotics names and codes from the National Drug Codes (NDC) directory²⁴ and the National Library of Medicines RXNorm database.²⁵ The antibiotics used to treat these patients were: *doxycycline*, *azithromycin*, *levofloxacin*, *oseltamivir*, *vancomycin*, *trimethoprim*, *fluoroquinolones*, *vibramycin*, *ofloxacin*, *clarithromycin*, *telithromycin*, *amoxicillin*, *cefuroxime*, *piperacillin-tazobactam*, *cefepime*. Data extraction and preprocessing was done via Structured Query Language (SQL) queries and Python scripts. This process included the removal of duplicate and implausible data entries. For example, if a time-stamp for a clinical measurement occurred after a patient was discharged, that patient record was excluded.

Markov Chain Model

The Markov chain model consisted of three mutually exclusive (transient) health states: no acute respiratory failure (NARF), acute respiratory failure type 1 (ARF_1), and acute respiratory failure type 2 (ARF_2), and two (absorbing) outcome states: discharge and death (Figure 2). The patient histories, $H = \{h_0, h_1, \dots, h_n\}$, were represented by a set of state transition events, $\{(s_0, t_0), \dots, (s_n, t_n)\}$. Here a time-stamp, t denoted the exact time a patient entered a particular state, s . We used this information to compute transitions as a function of time. For every sequential pair of transition events, we calculated the actual time between them in hours as $(s_{i+1} - s_i)$. We aggregated all patient

Table 1: Description of health and outcome states.

State Identifier		Definition	State Type
0	NARF	No acute respiratory failure	Health
1	ARF ₁	PaO ₂ ≤ 60	Health
		PaCO ₂ ≤ 50	
2	ARF ₂	PaCO ₂ > 50	Health
		pH ≤ 7.25	
		PaO ₂ ≤ 60	
3	Discharge	Discharge from hospital	Outcome
4	Death	In-hospital death	Outcome

The first column represents the numerical identifier for each state, following the state name identifier.

histories to construct the state transition matrix T , where x_{ij} denotes the total number of hours spent in state i before moving to state j .

Transition probabilities matrices (TPMs) for the Markov model were estimated using the ratio of the number of hours spent in a specific state before transitioning over the total number of hours spent in that state. Each transition probability (or model parameter) represents the estimated likelihood of a patient changing from one state to another during their hospital stay. Specifically, we calculated transition probabilities by:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=0}^n x_i} \quad (1)$$

where p_{ij} is the probability of transitioning from state i to state j at any given time. We set p_{ij} equal to the percentage of hours spent by individuals in state i and ending in state j relative to the total amount of time spent in state i . Note that $p_{ii} = 1$ for an outcome state. The model has a corresponding initial probability vector that is calculated from the initial percentage of patients starting in each of the health states.

Monte Carlo Simulation

Monte Carlo simulations were performed to estimate the outcomes from antibiotic administration on AECOPD patients in the ICU. We simulated the transitions of individual patients over time until they were either discharged or they died for the following two questions of interest:

- i. What is the impact of antibiotics on death for hospitalized AECOPD patients?
- ii. What is the initial timing of antibiotics for hospitalized AECOPD patients and the impact on death?

We ran the simulation to compare the in-hospital death rate for two populations, those that received antibiotics and those that did not. Patients were classified into four groups, ALL (the full population), ANTIBIOTICS (those for whom antibiotics were administered), and NO ANTIBIOTICS (those who did not receive antibiotics). For each of the two questions, the simulation was run for 100 replications (or cohorts), each of size 1000 patients. Each patient was probabilistically assigned to an initial health state based on the initial state probabilities. The occurrence of transitions

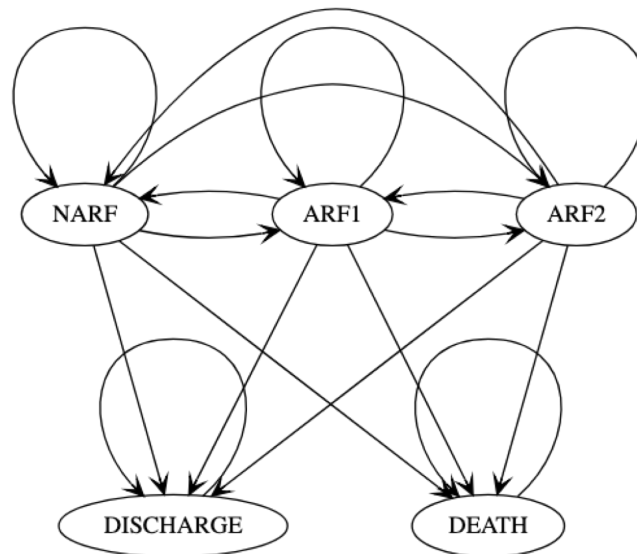


Figure 2: A 5-state first-order Markov model for AECOPD.

between transitioning between two states was based on the transition probabilities from the possible range of values, specified in the P matrix. Because the probability of a patient leaving an outcome state is zero, once a patient entered an outcome state, the process ended and a new patient was simulated. The mean, minimum, and maximum of the patient outcomes for the 100 replications was calculated and used for simulation statistics and model validation. In order to validate the simulation, we obtained a baseline from the raw EHR data and compared the observed outcomes (mean and variance) to estimated results from simulation.

Sensitivity analysis was conducted to quantify the effects of changes in model parameters (i.e., transition probabilities). A small change randomly chosen between plus or minus 10% of each transition probability value was applied to the model parameters. That is:

$$p'_{ij} = (p_{ij} - (.1 * p_{ij})) + (r * (.1 * p_{ij})) \quad (2)$$

and normalized using the row sum for each state.

A new TPM was obtained by normalizing the generated values so that the row sums equalled 1. The outcomes were then re-computed from the simulation using the new TPM.

RESULTS

Table 2 describes the characteristics of the patient groups, including the percentage of patients who received antibiotics. Notably, the NO ANTIBIOTICS group had the highest death rate (9.88%) and 0% of the patients were administered antibiotics. Figure 3 shows the transition probabilities and initial state probability vectors for each group, which derived from the Markov chain models created from the raw data. Additionally, figure 3 includes an additional TPM for a group, HALF, which was computed using the TPMs from the ANTIBIOTICS and NO ANTIBIOTICS group. The HALF TPM was used to conduct an analysis for administering 50% of the population antibiotics. We used the initial state probabilities from the ALL group to simulate each of the four models.

Table 2: Descriptive characteristics of patients included.

	ALL (n = 697)	ANTIBIOTICS (n = 181)	NO ANTIBIOTICS (n = 516)
Age in years ^a	70.81 ± 11.18	70.6 ± 10.76	70.88 ± 11.34
Women ^b	357 (51.22%)	91 (50.28%)	266 (51.55%)
Death ^b	59 (8.46%%)	8 (4.41%%)	51 (9.88%%)
LOS, days ^c	8.63	6.92	9.26
Administered antibiotics	25.9%	100%	0%
Initial antibiotics			
Drug name			-
<i>vancomycin</i>	58 (8.32%)	58 (32.04%)	-
<i>azithromycin</i>	47 (6.59%)	47 (25.97%)	-
<i>levofloxacin</i>	38 (5.45%)	38 (20.99%)	-
Administration, hours ^c	27.37	27.37	-

^a mean ± standard deviation

^b count (%)

^c mean

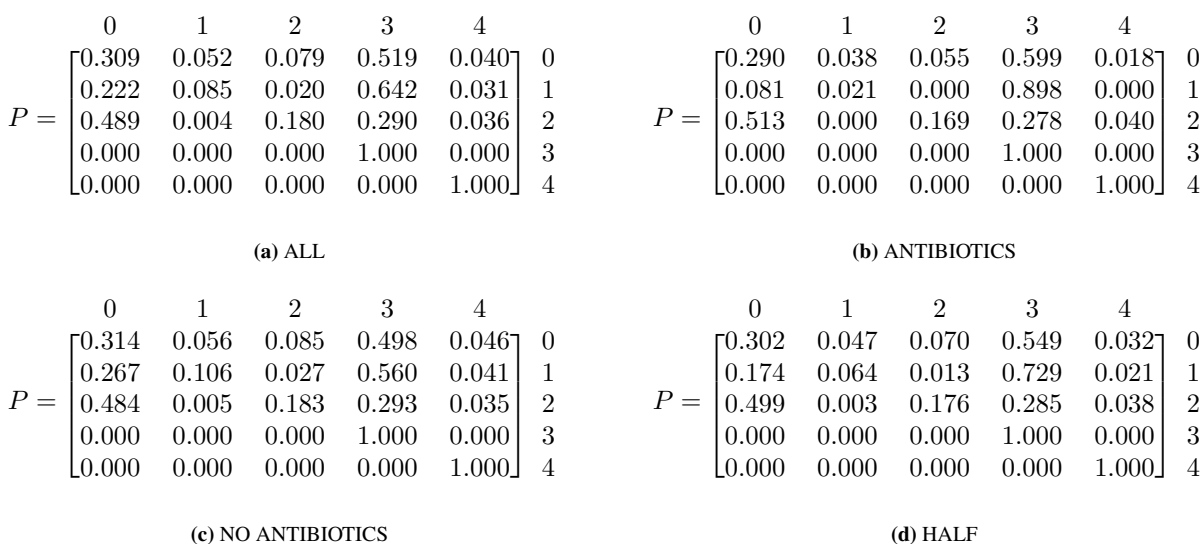


Figure 3: Model parameter estimations for all AECOPD patients administered antibiotics and not administered antibiotics are presented by group. Row and column labels correspond to health and outcome states in Table 1.

Figure 4a shows the estimated percentage of ICU patients with AECOPD that died as a function of the percentage that received antibiotics. Note that the actual number of patients that received antibiotics in the EHR was 25.9% with a death rate of 8.46% for the full cohort. This falls within the death rate interval [5.3%, 9.6%] estimated by the simulation. Also note that as the number of patients administered antibiotics increased, the percentage of deaths significantly decreased. Specifically, when all patients were given antibiotics, the death rate decreased by 50% vs not

administering antibiotics.

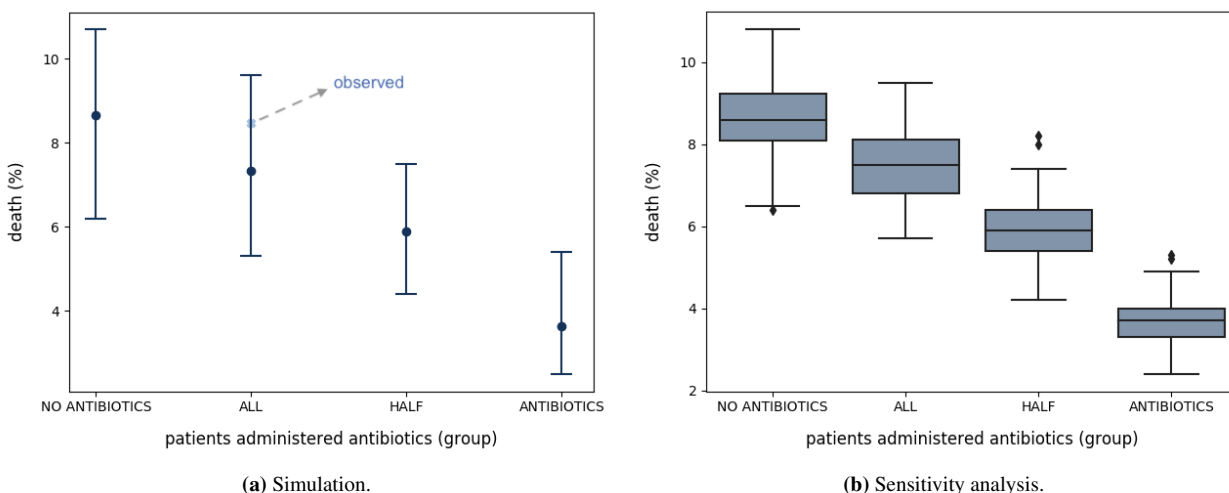


Figure 4: Results for (a) The estimated percentage of death for AECOPD based on antibiotics or no antibiotics administration. The x-axis is ordered by the percentage patients administered antibiotics and labeled by group. The y-axis represents the percentage deaths for that particular percentage of patients. The top and bottom horizontal lines are the maximum and minimum percentage of patient deaths, respectively. The black data point represents the average percentage of patient deaths, and the data point labeled, *observed*, represents the actual percentage of AECOPD patients that died when administered antibiotics. (b) The death percentage estimated for sensitivity analysis. Outliers are represented by data points. The median death percentage is represented by the horizontal line inside of the box. Top and bottom horizontal lines represent values higher than and the median.

The average time to antibiotic administration was 27 hours, and 32% of ANTIBIOTICS patients were administered *vancomycin* as the initial antibiotic (Table 2). Models were created using data from patients that were given antibiotics within the following specified time-frames: i) within 6 hours of hospital admission, ii) between 6 and 24 hours of admission, iii) between 24 and 48 hours of admission, and iv) after 48 hours of admission. We also used the initial state probabilities from the ALL group to simulate each of the four models.

Interestingly, Figure 6a shows a 5.5% mortality in the group of cohorts that received antibiotics after 48 hours vs 1.8% in the group that received antibiotics between 24 and 48 hours. The sensitivity analyses was performed by the perturbation of each probability in the transition probability matrix, and then observing how that perturbation changed the outcome (death). Figure 6b shows the results of the sensitivity analysis for different groups.

DISCUSSION

Physicians frequently make critical decisions in the face of uncertainty,³ weighing the risks and benefits of patient treatments and outcomes. Uncertainty is inevitable in medical reasoning and decision making, especially in critical care settings where the complete patient history may not be available. We introduced an approach to systematically capture, analyze, and estimate the effect of treatment interventions on outcomes through modeling and estimating uncertainties using EHR data. A major contribution of our work is the ability to derive transition probabilities from empirical time-series data by scaling the probabilities over time. This approach captures the duration of time that a patient spent in a particular state and allows extraction of temporal inferences from mathematical models using real-world data. Additionally, we demonstrated our approach to, for the first time, assess the administration and timing of antibiotics for AECOPD and the corresponding impact on in-hospital mortality. Our findings support a general approach of estimating patient outcomes from raw clinical data.

Chronically ill patients are susceptible to frequent changes in health status,²⁶ especially when under care in an ICU. For AECOPD patients in the ICU, clinical indicators can change suddenly, causing a patient to move between the health

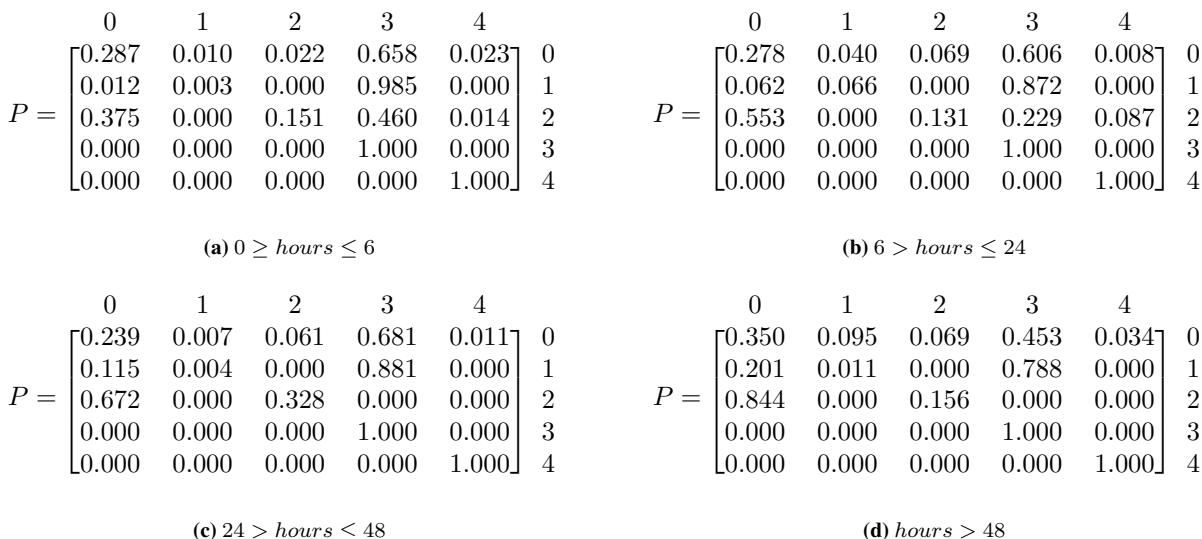


Figure 5: Model parameter estimations for AECOPD patients given antibiotics according to initial timing administered in hours are presented by group. Row and column labels correspond to health and outcome states in Table 1.

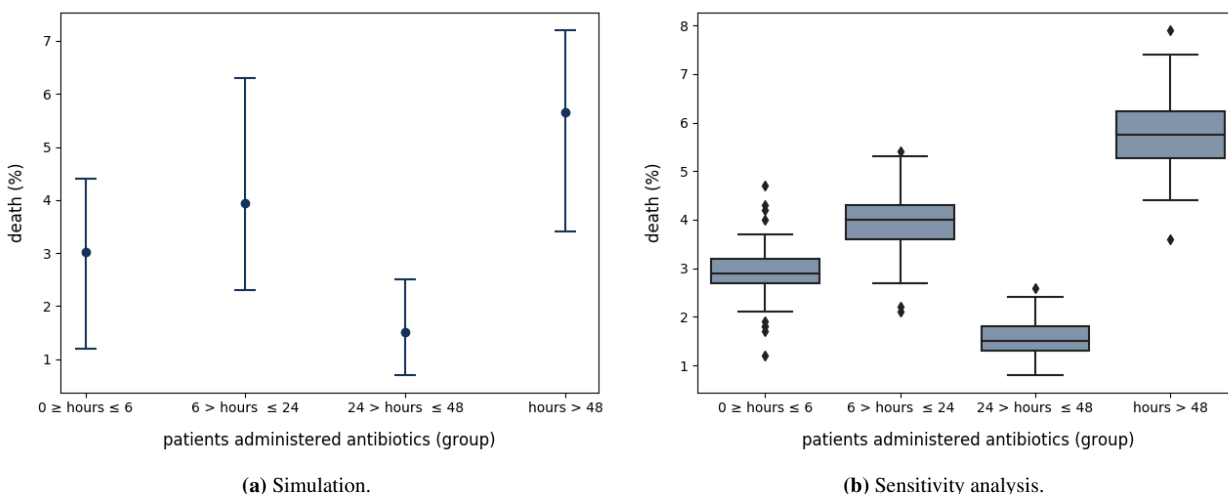


Figure 6: Results for (a) The estimated percentage of death for ICU patients with AECOPD based on the initial timing antibiotics were administered. The x-axis represents the time from admission in which patients were administered antibiotics, labeled by group. The y-axis represents the percentage deaths for that particular percentage of patients. The top and bottom horizontal lines are the maximum and minimum percentage of patient deaths, respectively. The black data point represents the average percentage of patient deaths, and the data point labeled, *observed*, represents the actual percentage of AECOPD patients that died when administered antibiotics. (b) Outliers are represented by data points. The median death percentage is represented by the horizontal line inside of the box. Top and bottom horizontal lines represent values higher than and the median.

states repeatedly, as illustrated in Figure 2. In our study we have seen an overall reduced mortality rate (Figure 4a) in patients with AECOPD admitted to the ICU who received antibiotics in comparison to those who did not. These findings are consistent with previous evidence that supports the use of antibiotics in critically ill patients given they

frequently have community acquired pathogens.²⁷ Antibiotics were shown to reduce all-cause mortality in critically ill patients and also reduce treatment failure after 4 weeks of discharge.²⁸ Our findings suggest that there may be a reduction in mortality rate when initiation of antibiotics occurs earlier in settings of AECOPD patients with severe respiratory failure warranting an ICU admission.

Limitations

One limitation of our study is that all of our data was limited to that in the ICU. Additional patient history might have helped to refine the findings. Another limitation is the process of selecting patient cohorts. The retrospective EHR data used required modeling patients who were assigned an AECOPD ICD-9 code. Though ICD-9 codes have been used for identifying patients when conducting retrospective clinical studies, accurately identifying COPD patients can be difficult in the absence of patient medical history.²³ COPD is diagnosed by lung function tests and staged according to the progression of the disease. Thus, simply using ICD-9 codes may not be enough information to identify an individual with COPD but used identify an individuals with COPD-like symptoms (e.g., wheezing, or noisy breathing, chest pain) who was given a COPD ICD-9 diagnosis. Such symptoms may be documented as free-text, unstructured data in EHR clinical notes. We also acknowledge limitations with modeling EHR data as Markov chains in that state transitions in first-order Markov chains only depend on the current state. Thus, such modeling does not account for confounding factors that previously impacted a patient prior to entering their current state. While Markov models are valuable when history is not important or available, they potentially lack accuracy when history matters.²⁹ Work has been done²⁹ to develop more complex Markov chain modeling methods that incorporate history.

Future Work

The development of clinical data networks such as Patient-Centered Clinical Research Network (PCORnet)³⁰ and Informatics for Integrating Biology and the Bedside (i2b2),¹³ allow for the integration of various clinical and research data sources into a single repository, increasing the amount and availability of clinical data. These platforms offer tools for cohort discovery and analytics that are currently being used for clinical research. Specifically, in addition to using ICD-9 codes, i2b2's natural language processing (NLP) feature for cohort discovery can be used for identifying AECOPD patient by extracting medications, smoking status, and diagnosis from clinical notes. While these tools allow simple data analysis, they do not offer support for granular temporal data analysis. Our methods can be integrated into such existing systems to allow more complete analysis.

Our methodology can also supplement real-time analysis of physiological data streams³¹ in hospital settings and be applied to both common and more complex clinical conditions. That is, we defined health states for our model that encompassed several clinical variables. However, given the vast space of clinical conditions and corresponding clinical criteria that defines the conditions, some health states may not be well represented. For example, in our state space NARF is defined as no acute respiratory failure, which means that the only criteria that determines if a patient is in NARF is that the constraints used to define ARF₁ and ARF₂ are not met. In this case, NARF does not reveal much information about the condition of the patient. To address this, information from various data sources such as past research studies,⁸ clinical trials,³² and medical devices can be combined to define additional clinical states. Clinical data warehouses^{12,13,30} have proven to be valuable for overcoming the barriers of data access and availability as they provide access and structure of medical data from multiple sources as well as tools for clinical discoveries through data analysis. Our work can be integrated as an API for platforms such as i2b2¹³ for complex data analytics and generalized to fit clinical data standards such as Fast Healthcare Interoperability Resources (FHIR)³³ for integration with healthcare systems.

CONCLUSION

Early administration of antibiotics for AECOPD patients in ICU settings can potentially reduce mortality rate. Such evidence can be generated for clinical guidelines and more generally, clinical discoveries, by developing comprehensive probabilistic modeling and simulation techniques that integrate temporal aspects of EHR data. As clinical researchers continue to strive to achieve data standardization, interoperability, and reproducibility, the methods presented in this paper can help leverage multiple EHR data sources and technological advancements.

References

1. Joan Vlayen, Bert Aertgeerts, Karin Hannes, Walter Sermeus, and Dirk Ramaekers. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *International Journal for Quality in Health Care*, 17(3):235–242, 2005.
2. George Weisz, Alberto Cambrosio, Peter Keating, Loes Knaapen, Thomas Schlich, and Virginie J Tournay. The emergence of clinical practice guidelines. *The Milbank Quarterly*, 85(4):691–727, 2007.
3. Andrew M Fine, Ben Y Reis, Lise E Nigrovic, Donald A Goldmann, Tracy N LaPorte, Karen L Olson, and Kenneth D Mandl. Use of population health data to refine diagnostic decision-making for pertussis. *Journal of the American Medical Informatics Association*, 17(1):85–90, 01 2010.
4. RI Horwitz, A Hayes-Conroy, R Caricchio, and BH Singer. From evidence based medicine to medicine based evidence. *The American journal of medicine*, 130(11):1246, 2017.
5. Michael D Cabana, Cynthia S Rand, Neil R Powe, Albert W Wu, Modena H Wilson, Paul-Andre C Abboud, and Haya R Rubin. Why don't physicians follow clinical practice guidelines?: A framework for improvement. *Jama*, 282(15):1458–1465, 1999.
6. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease 2018 report, 2018. https://goldcopd.org/wp-content/uploads/2017/11/GOLD-2018-v6.0-FINAL-revised-20-Nov_WMS.pdf, (accessed 27 March 2019).
7. Earl S. Ford. Trends in mortality from copd among adults in the united states. *Chest*, 148(4):962 – 970, 2015.
8. Stephan Budweiser, Rudolf A Jörres, and Michael Pfeifer. Treatment of respiratory failure in copd. *International journal of chronic obstructive pulmonary disease*, 3(4):605, 2008.
9. Robert Wilson, Sanjay Sethi, Antonio Anzueto, and Marc Miravittles. Antibiotics for treatment and prevention of exacerbations of chronic obstructive pulmonary disease. *Journal of Infection*, 67(6):497 – 515, 2013.
10. Lise Poissant, Jennifer Pereira, Robyn Tamblyn, and Yuko Kawasumi. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. *Journal of the American Medical Informatics Association*, 12(5):505–516, 2005.
11. Jesus J Caban and David Gotz. Visual analytics in healthcare opportunities and research challenges. *Journal of the American Medical Informatics Association*, 22(2):260–262, 2015.
12. George Hripcsak, Jon D. Duke, Nigam H. Shah, Christian G. Reich, Vojtech Huser, Martijn J. Schuemie, Marc A. Suchard, Rae Woong Park, Ian Chi Kei Wong, Peter R. Rijnbeek, Johan Van Der Lei, Nicole Pratt, G. Niklas Norén, Yu Chuan Li, Paul E. Stang, David Madigan, and Patrick B. Ryan. Observational health data sciences and informatics (ohdsi): Opportunities for observational researchers. In *Studies in Health Technology and Informatics*, volume 216 of *Studies in Health Technology and Informatics*, pages 574–578. IOS Press, 2015.
13. Shawn N Murphy, Griffin Weber, Michael Mendis, Vivian Gainer, Henry C Chueh, Susanne Churchill, and Isaac Kohane. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *Journal of the American Medical Informatics Association*, 17(2):124–130, 2010.
14. Sunil Nair, Douglas Hsu, and Leo Anthony Celi. Challenges and opportunities in secondary analyses of electronic health record data. In *Secondary Analysis of Electronic Health Records*, pages 17–26. Springer International Publishing, 2016.
15. Eric J Topol and Dick Hill. *The creative destruction of medicine: How the digital revolution will create better health care*. Basic Books New York, 2012.

16. Christopher H Jackson, Linda D Sharples, Simon G Thompson, Stephen W Duffy, and Elisabeth Couto. Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(2):193–209, 2003.
17. A Rodina-Theocharaki, K Bliznakova, and N Pallikarakis. Markov chain monte carlo simulation for projection of end stage renal disease patients in greece. *Computer Methods and Programs in Biomedicine*, 107(1):90–96, 2012.
18. Emma Tan, Ruud Boessen, David Fishwick, Rinke Klein Entink, Tim Meijster, Anjoeka Pronk, Birgit van Duuren-Stuurman, and Nick Warren. A microsimulation model for the development and progression of chronic obstructive pulmonary disease. *Respiratory medicine*, 109(12):1521–1531, 2015.
19. Daniel Abler, Vassiliki Kanellopoulos, Jim Davies, Manjit Dosanjh, Raj Jena, Norman Kirkby, and Ken Peach. Data-driven markov models and their application in the evaluation of adverse events in radiotherapy. *Journal of radiation research*, 54(suppl_1):i49–i55, 2013.
20. Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3, 2016.
21. Mohammad Adibuzzaman, Ken Musselman, Alistair Johnson, Paul Brown, Zachary Pitluk, and Ananth Grama. Closing the data loop: An integrated open access analysis platform for the MIMIC database. *Computing in cardiology*, 43:137, 2016.
22. Centers for Disease Control and Prevention. ICD - Classification of Diseases, Functioning, and Disability. <https://www.cdc.gov/nchs/icd/index.htm>, (accessed 28 May 2019).
23. Brian D Stein, Adriana Bautista, Glen T Schumock, Todd A Lee, Jeffery T Charbeneau, Diane S Lauderdale, Edward T Naureckas, David O Meltzer, and Jerry A Krishnan. The validity of international classification of diseases, ninth revision, clinical modification diagnosis codes for identifying patients hospitalized for COPD exacerbations. *Chest*, 141(1):87–93, 2012.
24. Center for Drug Evaluation and Research. National drug code directory. <https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory>, (accessed 12 Jan 2019).
25. National Library of Medicine. rxnorm. <https://www.nlm.nih.gov/research/umls/rxnorm/>, (accessed 12 Jan 2019).
26. Judith E Nelson, Christopher E Cox, Aluko A Hope, and Shannon S Carson. Chronic critical illness. *American journal of respiratory and critical care medicine*, 182(4):446–454, 2010.
27. Marc Miravittles and Antonio Anzueto. Role of infection in exacerbations of chronic obstructive pulmonary disease. *Current opinion in pulmonary medicine*, 21(3):278–283, 2015.
28. Daniela J Vollenweider, Harish Jarrett, Claudia A Steurer-Stey, Judith Garcia-Aymerich, and Milo A Puhan. Antibiotics for exacerbations of chronic obstructive pulmonary disease. *Cochrane Database of Systematic Reviews*, 12, 2012.
29. Tao Wu and David F. Gleich. Retrospective higher-order markov processes for user trails. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17*, pages 1185–1194, New York, NY, USA, 2017. ACM.
30. Rachael L Fleurence, Lesley H Curtis, Robert M Califf, Richard Platt, Joe V Selby, and Jeffrey S Brown. Launching PCORnet, a national patient-centered clinical research network. *Journal of the American Medical Informatics Association*, 21(4):578–582, 2014.
31. Allan F Simpao, Luis M Ahumada, Jorge A Gálvez, and Mohamed A Rehman. A review of analytics and clinical informatics in health care. *Journal of medical systems*, 38(4):45, 2014.

32. Shawn D Aaron, Katherine L Vandemheen, François Maltais, Stephen K Field, Don D Sin, Jean Bourbeau, Darcy D Marciniuk, J Mark FitzGerald, Parameswaran Nair, and Ranjeeta Mallick. Tnf antagonists for acute exacerbations of copd: a randomised double-blind controlled trial. *Thorax*, 68(2):142–148, 2013.
33. Barbara Franz, Andreas Schuler, and Oliver Krauss. Applying fhir in an integrated health monitoring system. *EJBI*, 11(2):51–56, 2015.