

# Phenotyping COVID-19 Patients by Ventilation Therapy: Data Quality Challenges and Cohort Characterization

Patrick ESSAY<sup>a,1</sup>, Jarrod MOSIER<sup>a</sup> and Vignesh SUBBIAN<sup>a</sup>

<sup>a</sup> *The University of Arizona, Tucson, AZ, USA*

**Abstract.** The COVID-19 pandemic introduced unique challenges for treating acute respiratory failure patients and highlighted the need for reliable phenotyping of patients using retrospective electronic health record data. In this study, we applied a rule-based phenotyping algorithm to classify COVID-19 patients requiring ventilatory support. We analyzed patient outcomes of the different phenotypes based on type and sequence of ventilation therapy. Invasive mechanical ventilation, noninvasive positive pressure ventilation, and high flow nasal insufflation were three therapies used to phenotype patients leading to a total of seven subgroups; patients treated with a single therapy (3), patients treated with either form of noninvasive ventilation and subsequently requiring intubation (2), and patients initially intubated and then weaned onto a noninvasive therapy (2). In addition to summary statistics for each phenotype, we highlight data quality challenges and importance of mapping to standard terminologies. This work illustrates potential impact of accurate phenotyping on patient-level and system-level outcomes including appropriate resource allocation under resource constrained circumstances.

**Keywords.** computable phenotype, respiratory failure, severe COVID-19

## 1. Introduction

Evidence-based recommendations regarding treatment selection for COVID-19 patients are largely limited [1]. Ventilator shortages have further increased the importance of appropriate resource allocation [2]. Respiratory support during the pandemic has shifted from primarily invasive mechanical ventilation (IMV) during the early phases of the crisis (December 2019 to May 2020) to various modes of noninvasive ventilation (NIV) such as noninvasive positive pressure ventilation and high flow nasal insufflation (HFNI). This shift was due to preliminary data regarding safety and efficacy of NIV among other reasons that continue to be debated among the intensive care community [3]. Thus, retrospective analysis of patient subgroups stratified by ventilation therapy is critical for not only understanding clinical outcomes but also for evaluating which therapy is most appropriate for a critically ill patient under resource-constrained circumstances. To enable such cohort characterization, we applied a rule-based phenotyping algorithm, which was originally developed for acute respiratory failure [4]. In doing so, we aimed to (1) evaluate challenges associated with extraction and phenotyping of COVID-19

---

<sup>1</sup> Corresponding Author, Patrick Essay, 1127 E James E Rogers Way, Tucson, AZ. US; E-mail: pessay@email.arizona.edu.

patients using electronic health record (EHR) data, (2) analyze outcomes of patient cohorts stratified by ventilation therapy, and (3) explore factors affecting failure of a particular NIV therapy requiring endotracheal intubation.

2. Methods

De-identified data were obtained from a clinical data warehouse in a regional health system with hospitals across six states in the Western United States. All adult patients ( $\geq 18$  years) with COVID-19 diagnosis (ICD10 code, U07.1 or lab-confirmed positive) requiring IMV or NIV individually or any combination of ventilation therapy were included. Low flow oxygen therapies (i.e.,  $\leq 15$  liters per minute oxygen flow rate) were excluded. Data were extracted from intensive care unit (ICU) records of patients admitted between January 3<sup>rd</sup> and September 29<sup>th</sup> of 2020. This study was reviewed and approved by two local institutional review boards.

We used a rule-based phenotyping algorithm, adapted from [4], to classify COVID-19 patients requiring IMV and NIV. Two methods of NIV were considered: 1) noninvasive positive pressure ventilation (NIPPV) and 2) high flow nasal insufflation (HFNI). While standardized terminology exists for NIPPV, concepts for HFNI are lacking and inconsistent [5]. NIPPV records were located as *ventilation events* in the EHR in addition to other concepts such as medications. Record of an oxygen flow rate in excess of 15 liters per minute was used as a surrogate identifier of HFNI.

3. Results

The phenotyping algorithm identified 1334 patients with COVID-19 admitted to ICUs at 17 medical centers. A total of seven phenotypes were created (Table 1), including patients initially intubated and then weaned onto either of the two NIV methods. Patients were primarily male in all subgroups except the HFNI cohort. Patients treated with a single therapy totaled 54.5% (727). The failure rate of NIPPV (i.e., patient treated with NIPPV and subsequently required IMV) was 72.1% and the failure rate of HFNI was 91.5% (i.e., patient treated with HFNI and subsequently required IMV). NIPPV and NIPPV failure patients had a mortality rate of 24.2% and 36.8%, respectively, and HFNI and HFNI failure a mortality rate of 61.5% and 80.0%. Mortality rate of intubation-only patients was 43.3%. In addition, ICU lengths of stay were generally higher for patients that failed either mode of NIV therapy.

Table 1. Summary characteristics of critically ill COVID-19 phenotypes based on ventilation therapy.

	Invasive	NIPPV	NIPPV failure	HFNI	HFNI failure	IMV to NIPPV	IMV to HFNI
Patients, n	574	140	361	13	140	85	21
Age, years, median (IQR)	59 (20)	64 (24)	62 (18)	64 (22)	63 (18)	53 (19)	61 (11)
Gender, % male	61.3	56.9	57.3	25	63.6	60.0	57.1
APACHE score, median (IQR)	74 (43)	51 (24)	65 (31)	54 (24)	77 (34)	60 (35)	82 (39)
ICU length of stay, days, median (IQR)	12.9 (17)	9.66 (11)	19.3 (15)	17.7 (11)	20.1 (16)	21.1 (23)	30.4 (21)
Mortality, %	43.3	24.2	36.8	61.5	80.0	8.23	33.3
PaO <sub>2</sub> , median (IQR)	77 (63)	50 (24)	77 (60)	50 (16)	75 (50)	77 (79)	75 (45)

PaCO <sub>2</sub> , median (IQR)	37 (18)	20 (18)	39 (18)	20 (10)	38 (17)	40 (17)	34 (21)
FiO <sub>2</sub> , median (IQR)	80 (60)	38 (32)	100 (50)	38 (12)	90 (50)	100 (62)	80 (60)
Respiratory rate, median (IQR)	30 (22)	35 (12)	33 (12)	36 (12)	34 (11)	31 (13)	34 (24)

IQR: Interquartile Range; APACHE: Acute Physiology and Chronic Health Evaluation; PaO<sub>2</sub>: arterial oxygen partial pressure; PaCO<sub>2</sub>: arterial carbon dioxide partial pressure; FiO<sub>2</sub>: fractional inspired oxygen

4. Discussion

The rule-based phenotyping algorithm allows for scalable, secondary analysis of electronic health record (EHR) data and deeper understanding of various COVID-19 subgroups. There are, however, several challenges related to data completeness and consistency, making it difficult to extract accurate representations of clinical concepts from large-scale EHR data.

4.1 Data Quality Challenges

Data elements such as ventilation events for HFNI could not be directly mapped to standardized terminologies. This was alleviated by using surrogate ventilation events and oxygen flow rates. For example, *ventilation events* included more than 20 nonstandard terms for HFNI patients. In addition, only 8.7% of HFNI patients had *ventilation event* records while the remaining patients were identified by oxygen flow rate > 15 liters per minute. Without this surrogate identifier, most HFNI records would be overlooked. Mapping to Common Data Models (CDM) such as the Observational Medical Outcomes Partnership (OMOP) CDM [6] may offer a basic approach for data harmonization, provided there are standardized concepts for various ventilation events.

Accurately identifying critically ill patients was also a challenge. For this study, we included patients that were admitted to an ICU. However, ICU admission was less relevant during the pandemic with health systems near capacity and critically ill patients treated outside of ICUs [7]. This approach may have missed patients that were treated with NIV outside the ICU and were not included in the analysis. Inclusion of non-ICU patients will require definition of specific parameters for patients to be considered critically ill regardless of physical location. Parameters such as diagnosis and respiratory and cardiovascular lab measurements could be used to define physiological state of a patient as critically ill.

4.2 Potential Impact and Future Work

Guidelines for respiratory management of patients changed during the course of the pandemic and continue to change [8][9]. Presumably, patient outcomes would improve as more reliable evidence on ventilation treatment approaches are generated. Our approach allows for these results to be evaluated over time as evidence evolves. Improvements in patient stratification and the quality of data extracted will allow for deeper investigation of measurements of particular interest to clinicians and more direct relevance to the disease itself.

Indicators of NIV failure could have substantial patient-level impact through clinical decision-making support [10]. Clinical factors related to NIV failure could be used to guide clinical decision-making. This is particularly important during crucial times of resource management. Ultimately, these factors could be used as inputs for clinical

prediction models to predict diagnostic or prognostic outcomes of NIV failure. This work also has potential system-level impact as a result of improved resource allocation [11]. Patients likely to fail NIV would be treated with IMV rather than attempting NIV and then getting intubated later in their ICU stay. The overall phenotyping and analysis of ventilation classifications is applicable beyond COVID-19 patients as well and may be applied to any illness or patient population requiring ventilation.

#### 4.3 Recommendations

As much as is feasible, data warehousing should include CDM mapping and use of standard terminologies. In addition, standardized concepts for high flow nasal insufflation should be created where nonexistent and expanded such that surrogate measures are not necessary for accurate phenotyping. Rather, surrogate measures should be used for source record verification that patients are correctly classified.

As for COVID-19 patient outcomes, continuous evaluation of individual subgroups should be performed as more is learned about the disease and how best to treat critical illness [12]. The comparisons made in our analysis include data from January to September of 2020. Consequently, this includes multiple shifts in the way patients were treated. Respiratory management in particular has evolved in a short time period and will continue to evolve to best serve patients in light of potential resource constraints [13].

### 5. Conclusions

Standardized terminologies and phenotyping algorithms allow for efficient classification of patients using EHR data and facilitate retrospective analysis of patient outcomes. Data extraction and processing, however, require numerous validation and quality assurance steps and remains time consuming. Granular CDM mapping of ventilator data using standard concepts could alleviate extraction and preprocessing delays leading to improved, quicker, and deeper analysis.

Our preliminary results can be used to evaluate COVID-19 patients requiring IMV and NIV. Deeper analysis could further characterize NIV failure patients and factors which may be indicative of impending ventilation failure for COVID-19 and guide resource allocation during times of resource constraints (e.g., ventilator shortages).

### Acknowledgment

This work was funded in part by the Emergency Medicine Foundation and the State of Arizona Technology and Research Initiative Fund (TRIF). The authors acknowledge research data services and support from Don Saner and Mario Arteaga at Banner Health.

### References

- [1] Guan L, Zhou L, Le Grange JM, Zheng Z, Chen R. Non-invasive ventilation in the treatment of early hypoxemic respiratory failure caused by COVID-19: considering nasal CPAP as the first choice. *Crit Care*. 2020 Jun 11;24(1):333.
- [2] Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, Zhang C, Boyle C, Smith M, Phillips

- JP. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *N Engl J Med*. 2020 May 21;382(21):2049-2055.
- [3] Remy KE, Lin JC, Verhoef PA. High-flow nasal cannula may be no safer than non-invasive positive pressure ventilation for COVID-19 patients. *Crit Care*. 2020 Apr 23;24(1):169.
  - [4] Essay P, Mosier J, Subbian V. Rule-Based Cohort Definitions for Acute Respiratory Failure: Electronic Phenotyping Algorithm. *JMIR Med Inform*. 2020 Apr 15;8(4):e18402.
  - [5] Miller DC, Bime C, Parthasarathy S, Mosier JM. High-Flow Oxygen Therapy Concepts: Time to Standardize Nomenclature and Avoid Confusion. *J Intensive Care Med*. 2020 May;35(5):519-523.
  - [6] Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li YC, Stang PE, Madigan D, Ryan PB. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform*. 2015;216:574-8.
  - [7] Aziz S, Arabi YM, Alhazzani W, Evans L, Citerio G, Fischkoff K, Salluh J, Meyfroidt G, Alshamsi F, Oczkowski S, Azoulay E, Price A, Burry L, Dzierba A, Benintende A, Morgan J, Grasselli G, Rhodes A, Möller MH, Chu L, Schwedhelm S, Lowe JJ, Bin D, Christian MD. Managing ICU surge during the COVID-19 crisis: rapid guidelines. *Intensive Care Med*. 2020 Jul;46(7):1303-25.
  - [8] Alhazzani W, Möller MH, Arabi YM, Loeb M, Gong MN, Fan E, Oczkowski S, Levy MM, Derde L, Dzierba A, Du B, Aboodi M, Wunsch H, Cecconi M, Koh Y, Chertow DS, Maitland K, Alshamsi F, Belley-Cote E, Greco M, Laundry M, Morgan JS, Kesecioglu J, McGeer A, Mermel L, Mammen MJ, Alexander PE, Arrington A, Centofanti JE, Citerio G, Baw B, Memish ZA, Hammond N, Hayden FG, Evans L, Rhodes A. Surviving Sepsis Campaign: Guidelines on the Management of Critically Ill Adults with Coronavirus Disease 2019 (COVID-19). *Crit Care Med*. 2020 Jun;48(6):e440-e469.
  - [9] Corrêa TD, Matos GFJ, Bravim BA, Cordioli RL, Garrido ADPG, Assuncao MSC, Barbas CSV, Timenetsky KT, Rodrigues RDR, Guimarães HP, Rabello Filho R, Lomar FP, Scarin FCC, Batista CL, Pereira AJ, Guerra JCC, Carneiro BV, Nawa RK, Brandão RM, Pesaro AEP, Silva Júnior M, Carvalho FRT, Silva CSM, Almeida ACF, Franken M, Pesavento ML, Eid RAC, Ferraz LJR. Intensive support recommendations for critically-ill patients with suspected or confirmed COVID-19 infection. *Einstein (Sao Paulo)*. 2020 Jun 3;18:eAE5793.
  - [10] De Vita N, Scotti L, Cammarota G, Racca F, Pissai C, Mastrone C, Colombo D, Olivieri C, Della Corte F, Barone-Adesi F, Navalesi P, Vaschetto R; for COVID-19 Eastern Piedmont Network. Predictors of intubation in COVID-19 patients treated with out-of-ICU continuous positive airway pressure. *Pulmonology*. 2021 Jan 20:S2531-0437(21)00002-7.
  - [11] Kirkpatrick JN, Hull SC, Fedson S, Mullen B, Goodlin SJ. Scarce-Resource Allocation and Patient Triage During the COVID-19 Pandemic: JACC Review Topic of the Week. *J Am Coll Cardiol*. 2020 Jul 7;76(1):85-92.
  - [12] Subbian V, Solomonides A, Clarkson M, Rahimzadeh VN, Petersen C, Schreiber R, DeMuro PR, Dua P, Goodman KW, Kaplan B, Koppel R, Lehmann CU, Pan E, Senathirajah Y. Ethics and informatics in the age of COVID-19: challenges and recommendations for public health organization and public policy. *J Am Med Inform Assoc*. 2021 Jan 15;28(1):184-89.
  - [13] Robba C, Battaglini D, Ball L, Patroniti N, Loconte M, Brunetti I, Vena A, Giacobbe DR, Bassetti M, Rocco PRM, Pelosi P. Distinct phenotypes require distinct respiratory management strategies in severe COVID-19. *Respir Physiol Neurobiol*. 2020 Aug;279:103455.