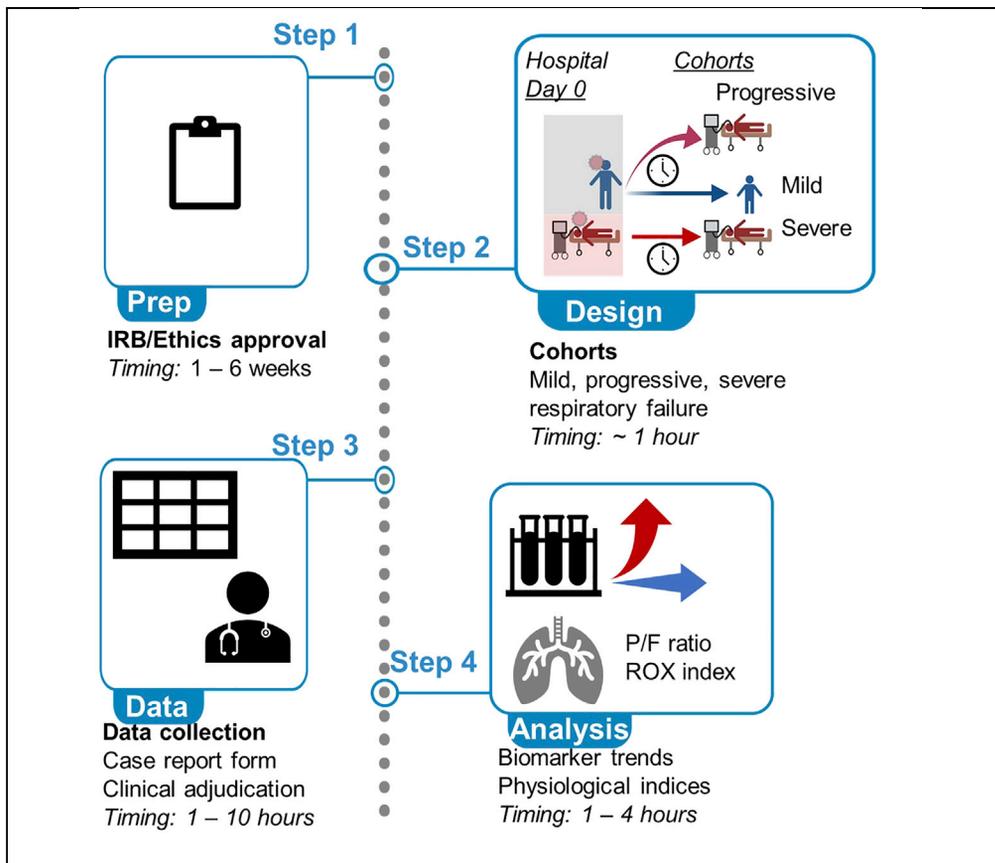


Protocol

Protocol for assessing and predicting acute respiratory decline in hospitalized patients



This protocol aids both new and experienced researchers in designing retrospective clinical and translational studies of acute respiratory decline in hospitalized patients. This protocol addresses (1) the basics of respiratory failure and electronic health record research, (2) defining patient cohorts as “mild, progressive, or severe” instead of “ICU versus non-ICU”, (3) adapting physiological indices, and (4) using biomarker trends. We apply these approaches to inflammatory biomarkers in COVID-19, but this protocol can be applied to any progressive respiratory failure study.

Conor P. Crowley,
Louis T. Merriam,
Alisa A. Mueller, ...,
Hibah Haider, Justin
D. Saliccioli, Edy Y.
Kim

crowconor@gmail.com
(C.P.C.)
ekim11@bwh.harvard.edu
(E.Y.K.)

Highlights

Basics of hypoxemic respiratory failure and electronic health records research

Design of patient cohorts that discriminate progressive respiratory failure

Physiological measurements of respiratory failure (P/F, ROX index)

Use of inflammatory biomarker trends to predict respiratory decline

Crowley et al., STAR Protocols

100545

2021 © 2021 The

Author(s).

[https://doi.org/10.1016/](https://doi.org/10.1016/j.xpro.2021.100545)

[j.xpro.2021.100545](https://doi.org/10.1016/j.xpro.2021.100545)



Protocol

Protocol for assessing and predicting acute respiratory decline in hospitalized patients

Conor P. Crowley,^{1,5,*} Louis T. Merriam,¹ Alisa A. Mueller,^{2,3} Tomoyoshi Tamura,^{1,3} Jeremy R. DeGrado,⁴ Hibah Haider,¹ Justin D. Saliccioli,^{1,3} and Edy Y. Kim^{1,3,6,*}

¹Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

²Division of Rheumatology, Brigham and Women's Hospital, Boston, MA 02115 USA

³Harvard Medical School, Boston, MA 02115, USA

⁴Department of Pharmacy, Brigham and Women's Hospital, Boston, MA 02115 USA

⁵Technical contact

⁶Lead contact

*Correspondence: crowconor@gmail.com (C.P.C.), ekim11@bwh.harvard.edu (E.Y.K.)
<https://doi.org/10.1016/j.xpro.2021.100545>

SUMMARY

This protocol aids both new and experienced researchers in designing retrospective clinical and translational studies of acute respiratory decline in hospitalized patients. This protocol addresses (1) the basics of respiratory failure and electronic health record research, (2) defining patient cohorts as “mild, progressive, or severe” instead of “ICU versus non-ICU”, (3) adapting physiological indices, and (4) using biomarker trends. We apply these approaches to inflammatory biomarkers in COVID-19, but this protocol can be applied to any progressive respiratory failure study.

For complete details on the use and execution of this protocol, please refer to Mueller et al. (2020).

BEFORE YOU BEGIN

Basics of respiratory failure for the non-specialist

Respiratory failure is defined as the inability of a patient's respiratory system to meet the body's oxygenation, ventilation or metabolic requirements. Respiratory failure in hospitalized patients has several etiologies, including chronic obstructive pulmonary disease, heart failure, pneumonia, and Acute Respiratory Distress Syndrome (ARDS). The key clinical challenge is predicting which hospitalized patients will remain stable and which patients will have progressive respiratory failure and/or ARDS. Defining predictors of respiratory decline using retrospective studies could help guide utilization of clinical resources or guide tailored therapeutic intervention.

Understanding respiratory failure. Performing studies of acute hypoxic respiratory failure requires an understanding of how these patients are diagnosed and managed. Emergency Department (ED) criteria for inpatient admission vary by institution and clinical context which can lead to disparities among studies. For example, during the COVID-19 pandemic, some overtaxed hospitals developed stricter criteria for admission. A patient's oxygen saturation (SpO₂) is measured by non-invasive pulse oximetry, typically a probe wrapped around the patient's finger or ear. If the SpO₂ is less than 89% while the patient is breathing room air without supplemental oxygenation, the patient will be considered to have respiratory failure and requires supplemental oxygen and hospital admission. Patients with SpO₂ greater than 89% may still be admitted if it is anticipated that the patient may clinically worsen. Other factors such as work of breathing, co-morbidities, advanced age,



non-respiratory illness, need for “inpatient only” treatments (e.g., i.v. antibiotics or antivirals) or psychosocial issues may also necessitate admission.

After admission to the hospital, the patient is triaged to either the intensive care unit (ICU) or a non-ICU medical floor. Absolute criteria that require ICU admission include need for intubation and invasive mechanical ventilation (IMV), vasopressor infusion for hypotension, or elevated nursing needs (e.g., due to altered mental status). Admitted patients with acute hypoxemic respiratory failure that do not require mechanical ventilation will need supplemental oxygen. The amount of O₂ and its delivery method will dictate the location and level of care, and these standards may differ across institutions. Methods of oxygen delivery and their estimated flow rates are summarized in [Table 2](#). Patients will be treated with a device that provides the least amount of supplemental oxygen delivery that normalizes their level of oxygen saturation and respiratory symptoms. For example, patients with severe respiratory decompensation may avoid intubation and mechanical ventilation by using special oxygen devices that deliver high amounts of oxygen, such as high flow nasal cannula (HFNC) or Non-Invasive Positive Pressure Ventilation (NIPPV, or “Bi-PAP”), in which a tight-fitting mask that ventilates a patient with positive air pressure in a manner similar to one mode of mechanical ventilation. Institutions may vary on whether patients on HFNC or NIPPV are cared for in the ICU, the non-ICU floor or a step-down unit that provides a level of care between these two options. Different hospitals triage patient locations differently, and the same hospital may triage patient locations differently over time. Thus, equating patient location, such as ICU versus non-ICU, may not adequately distinguish clinical severity of respiratory failure and can lead to systematic errors in the study’s interpretation. Methods to mitigate these concerns is discussed further in the section [step-by-step method details](#) “Develop patient cohorts that discriminate progressive respiratory failure (steps 6–10).”

P/F ratio to quantify hypoxemia. In addition to the categorical measures of acute hypoxemic respiratory failure, such as requirements for HFNC, NIPPV or IMV, several quantitative measures exist. Acute hypoxemic respiratory failure results in a gradient between the oxygen content of the air inspired by the patient (F_iO₂) and the resulting oxygen content of the patient’s blood (partial pressure of arterial oxygen, P_aO₂). This gradient is most commonly quantified as the ratio of P_aO₂ to F_iO₂, known as the P/F ratio. The P/F ratio allows comparison of patients treated with different concentrations of inspired oxygen. P_aO₂ can be measured directly by a laboratory test known as an Arterial Blood Gas, or ABG, from an arterial blood sample. For this purpose, critically ill patients often have temporary, indwelling arterial catheters. If frequent ABGs are not necessary, intermittent arterial puncture is performed. Patients that are improving or lack indwelling arterial catheters tend to have ABG drawn less frequently. Thus, studies benefit from estimating P_aO₂ by imputation from SpO₂, which all hospitalized patients will have recorded. [Table 1](#) shows empirically determined P_aO₂ and SpO₂ correlations ([Brown et al., 2017](#)). Patients on mechanical ventilation, high flow nasal cannula and Venturi face mask have an F_iO₂ directly selected by the clinician. For other types of supplemental oxygen, the O₂ flow rate and delivery device is selected, not the FiO₂. FiO₂ can be estimated using the formula $FiO_2 = 0.21 + 0.03 \times O_2 \text{ flow rate in L/min}$ ([Frat et al., 2015](#); [Coudroy et al., 2020](#)). Changes to the P/F ratio can reflect changes to clinical therapies or changes to the patient’s pathophysiology besides hypoxemia. For example, the P/F ratio can change if the clinician adjusts the end-expiratory pressure (PEEP) settings for a patient on NIPPV or mechanical ventilation, or if the clinician significantly adjusts the flow rate on HFNC. The P/F ratio can vary moment to moment due to other factors, including the patient’s posture and position in the bed; airway secretions and clearance; tachypnea; or synchrony with NIPPV and mechanical ventilation.

ROX index to quantify hypoxemia. Like the P/F ratio, the ROX index is another physiological index of hypoxemic respiratory failure ([Patel et al., 2020](#)). The ROX index integrates hypoxemia and respiratory rate as a proxy for work of breathing, and its calculation is described in the step-by-step methods. The ROX index was developed to predict respiratory failure in patients on HFNC. A lower index corresponds to worse respiratory failure. In this protocol, we suggest the innovative extension

Table 1. The imputation of P_aO_2 values from SpO_2 (Brown et al. 2017)

Measures SpO_2 (%)	Imputed PaO_2 (mmHg)
100	167
99	132
98	104
97	91
96	82
95	76
94	71
93	67
92	64
91	61
90	59
89	57
88	55
87	53
86	51

of the ROX index to non-intubated patients that are on other types of supplemental oxygen besides HFNC. Extension of the ROX index to oxygen delivery methods besides HFNC will require the methods to estimate F_iO_2 described in Tables 1 and 2.

ARDS. Patients can develop progressive respiratory decline culminating in Acute Respiratory Distress Syndrome (ARDS). In the Berlin criteria, ARDS is defined by acute timing chest imaging with bilateral opacities and lack of heart failure as the primary etiology (Ferguson et al., 2012). Etiologies of ARDS include respiratory infection, non-infectious lung inflammation) and non-respiratory illnesses (Table 4) After patients are admitted to the hospital, a key clinical challenge is predicting which hospitalized patients will remain stable and which patients will have progressive respiratory failure and ARDS. Defining predictors of respiratory decline could help guide utilization of clinical resources or guide tailored therapeutic intervention.

Basics of research utilizing electronic health records (EHR) for the non-specialist

Producing novel and reproducible clinical research using EHR has become increasingly important as these databases may allow researchers to draw inference about exposures without the need for time-intensive and resource-heavy prospective clinical trials. Despite their ever-increasing availability and great potential for research and clinical purposes, the ability to translate clinical electronic datasets into meaningful knowledge remains challenging. Researchers who endeavor to utilize

Table 2. Estimated fraction of inspired oxygen (F_iO_2) for O_2 delivery devices and their flow rates (Catterall et al., 1967; Frat et al., 2015; Hardavella et al., 2019; Coudroy et al., 2020)

Oxygen Device	Oxygen flow rate (L/min)	Estimated FiO_2 (%)
Nasal Cannula	1	24
	2	27
	3	30
	4	33
	5	36
	6	39
Simple Face Mask	6–10	44–50
Non-Rebreather Mask	10–20	Approximately 60–80
	20	60
	30	~70
	40	~80
High Flow Nasal Cannula	30–70	50–100

electronic and routinely collected data sources should be aware of common pitfalls so as to avoid the concern of amplifying low-level signals that do not have clinical importance or may not be generalizable beyond the source from which they are derived.

Investigators should become familiar with prior studies into best practices for reproducible clinical research from electronic medical records. For instance, the MIT Critical Data group have produced an open-access textbook on the [Secondary Analysis of Electronic Health Records](#) which provides step-by-step guidance and examples ([MIT Critical Data, 2016](#)). Knowledge of the pitfalls of research utilizing EHR will inform the decisions about study design.

Prior studies have found that the accuracy of different types of EMR data depend on research question and context. For example, administrative data, such as ICD-9 billing codes, were highly predictive of heart failure defined by clinical chart review ([Lee et al., 2005](#)). In contrast, similar administrative data had poor sensitivity for detecting inpatient adverse events, like deep vein thromboses or hospital-acquired pneumonia defined by clinical chart review ([Maass et al., 2015](#)).

Notably, these studies used chart review as the gold standard. Several types of error can affect chart review of EHR and are addressed in the section [troubleshooting](#) in “[problems 1, 2, 3 and 4](#).” A clinical study utilizing EHR requires several methodological decisions. First, the EHR data can be gathered by manual chart review or by electronic query. Manual chart review takes more time and had up to 10% error rate in one study ([Feng et al., 2020](#)), although such errors can be mitigated by protocol designs like using two independent researchers to duplicate chart review. However, electronic queries can result in datasets that require significant manual clean-up and pre-processing, or, at minimum, manual spot-checks of random subjects to ensure data quality. In addition, certain types of variables often require manual adjudication and input, such as free text notation of clinical decision making.

KEY RESOURCES TABLE

REAGENT or RESOURCE	Source	Identifier
Software and algorithms		
Prism software (GraphPad)	https://www.graphpad.com/scientific-software/prism/	N/A
R software (The R project)	https://www.r-project.org	N/A
Other		
Clinical treatment guidelines for standards of care	Varies by study. Our study of COVID-19 utilized the Brigham and Women’s Hospital COVID-19 Clinical Guidelines (https://www.covidprotocols.org)	N/A

STEP-BY-STEP METHOD DETAILS

Perform ethical review

⌚ Timing: Development of protocol will likely take 2–4 h. However, review and revision may take 1–6 weeks.

1. Submit the study protocol to the institutional review board (IRB) or ethics panel for comment, revision, and approval prior to starting the study.

Assess EHR dataset to formulate the research question.

⌚ Timing: Varying amounts of time will be required for reviewing EHR, however, likely 1–2 h will be required.

2. Formulate the research question:

In multi-disciplinary teams of clinicians (domain experts), data scientists, and biostatisticians, researchers must formulate the research question with an appraisal of the available data and its architecture.

3. Select the method of data entry:

Select manual, electronic query or a combination of both. Decide on protocols to ensure the accuracy of data, such random spot checks of data.

4. Pre-process the clinical dataset:

EMR data are of varying granularity and frequency and must be inspected and prepared in a pre-processing step before the case report forms can be completed. Pre-processing may require data modifications for missingness and outliers. Data-scientists and clinicians must work together to provide appropriate representation of the data forms.

5. Perform an exploratory data analysis to refine the initial research question:

Assess cohort demographics and the suitability of the dataset for the research question.

Develop patient cohorts that discriminate progressive respiratory failure

⌚ Timing: Less than one hour.

6. Secure the availability of domain experts to adjudicate clinical questions on study design, patient inclusion and exclusion, and data:

For studies of respiratory failure, domain experts include physicians that are board-certified subspecialists in Pulmonary Medicine, Critical Care Medicine, or Infectious Disease.

7. Review potential clinical questions regarding respiratory failure that affect study design and patient inclusion and exclusion criteria:

A key question is whether to identify, segregate, or exclude types of respiratory failure that cause critical respiratory illness but may not reflect ARDS pathophysiology. For example, an infection with SARS-CoV-2 could trigger an exacerbation of underlying asthma or COPD that leads to NIPPV or intubation without severe hypoxemia.

Note: By the definition of ARDS, respiratory failure primarily driven by heart failure is excluded in studies on ARDS.

8. Assess subjects for inclusion and exclusion; confirm each subject's diagnosis with criteria relevant to the research question:

For example, a study of COVID-19 pneumonia should exclude a patient who is hospitalized for appendicitis and only incidentally found to be SARS-CoV-2 positive by PCR and lacks respiratory symptoms.

9. Divide subjects in "mild," "progressive," or "severe" respiratory failure to distinguish patients with stable from progressive respiratory failure:

Table 3. Description of the cohorts of mild, progressive, and severe respiratory failure

Disease severity:	Description:
Mild	Non-critical illness and remains non-critical throughout the hospitalization. Adequate oxygen saturation on room air, nasal cannula, or simple face mask
Progressive	Initially non-critical illness however develops worsening oxygenation during hospitalization requiring high flow nasal cannula, non-invasive positive pressure ventilation, or mechanical ventilation
Severe	Critical illness on presentation requiring high flow nasal cannula, non-invasive positive pressure ventilation, or mechanical ventilation within 12 h of hospital admission

- a. “Mild”: Patients who initially present to the hospital with non-critical illness and remain non-critical throughout their hospital admission. That is, the patients only require room air, low flow nasal cannula, or face mask (excluding non-rebreathing mask [NRB]) during their hospitalization.
- b. “Progressive”: These patients have the same initial clinical presentation as the “mild” group—non-critical illness. However, “progressive” patients develop worsening respiratory failure during their hospital course and later require NRB, high-flow nasal cannula (HFNC), or non-invasive positive pressure ventilation (NIPPV) or invasive mechanical ventilation (IMV).
- c. “Severe”: These patients arrive at the hospital with critical illness and require NRB, HFNC, NIPPV or IMV within 12 h of presentation to the hospital (Table 3).

Note: Some studies of inpatient acute respiratory failure divides patients into cohorts of “ICU” or “non-ICU”, “severe or not severe disease” or focus on an association with mortality (Cummings et al., 2020; Petrilli et al., 2020; Zhang et al., 2020; Haljasmagi et al., 2020). However, this two-cohort model or the binary outcome of mortality may not elucidate the determinants of progressive respiratory failure. Predicting which patients with milder illness will remain stable and which patients will worsen to critical illness is a key clinical need to guide triage, intervention, and clinical studies. The three group model in this proposal is suited to capture the dynamic process of progressive respiratory failure.

Note: These categories are not to be confused with other classification schemes that use the term “severe,” such as “severe ARDS” (i.e., P/F ratio < 100) or “severe COVID-19 pneumonia” defined in the ACTT clinical trials (Beigel et al., 2020).

10. Consider sub-dividing the “progressive” and “severe” groups of subjects by their use of NRB, HFNC, NIPPV or IMV:

To capture progression within critical illness, the study can sub-divide the “progressive” and “severe” cohorts further into these patient sub-cohorts, listed in typical order of escalation of intensity of respiratory support:

 - a. “HFNC”: These patients required HFNC but did not require NIPPV or IMV at any point during their hospitalization.
 - b. “NIPPV”: These patients required NIPPV but did not require IMV at any point during their hospitalization. They may have required HFNC.
 - c. “IMV”: These patients required intubation for IMV at any point. They may have required HFNC or NIPPV.

Note: Grouping together patients that utilize any one of NRB, HFNC, NIPPV or IMV under a single definition, as in step 9, is reasonable. Any of these modalities of respiratory support are a clear escalation from milder respiratory failure that only requires nasal cannula or simple face mask. However, patients with ARDS can have progression of their respiratory deterioration within this larger definition of critical respiratory failure. For example, some patients will remain on HFNC, while other patients will worsen and require IMV after HFNC due to

Table 4. Berlin criteria for ARDS (Ferguson et al., 2012)

Berlin criteria for ARDS			
1. Acute onset respiratory decline			
2. Chest imaging revealing bilateral opacities without pleural effusion, lung collapse, nodules, or masses			
3. No Clinical evidence of significant heart failure			
Etiologies of ARDS			
<i>Pulmonary related illness:</i> Infection (bacterial or viral, aspiration, or interstitial lung disease		<i>Non-pulmonary related illness:</i> Sepsis, trauma, acute pancreatitis, burns, blood product transfusion, drug toxicity, or graft vs. host disease.	
Grading of ARDS			
ARDS Severity	Mild	Moderate	Severe
PF ratio:	300	200	100

worsening hypoxemia or work of breathing. Sub-cohorts within the “progressive” and “severe” categories would capture these clinical changes. Creating sub-divisions within critical illness quickly raises additional questions since patients can move between HFNC, NIPPV and IMV. For example, in a retrospective cohort study, the use of HFNC or NIPPV for a particular patient may be based on clinical characteristics, and the modes themselves may affect outcomes, as in a systematic review of HFNC and NIPPV in preventing IMV (Zhao et al., 2017). Should patients that are initially on HFNC or NIPPV but then intubated be distinguished from patients that go directly to IMV without HFNC or NIPPV? It can be hypothesized for a given study that extended periods of HFNC or NIPPV in a deteriorating patient risks greater patient-induced lung injury compared to earlier initiation of low tidal volume IMV; alternatively, delaying IMV, with its risks of complication such as sedation-induced delirium, could be beneficial. Thus, unless a study is well-powered to accommodate sub-divisions of critical illness, we recommend limiting the cohorts to “mild”, “progressive” and “severe” as described in step 9 as a simpler yet powerful approach to study the determinants of respiratory deterioration.

Develop a case report form (CRF)

⌚ Timing: 1–10 h depending on length of CRF and number of variables.

11. Develop a case report form to ensure standardized data collection across sites and staff:

An example of a case report form for a study of respiratory failure is in [Table S1](#).

Note: In steps 2–5, a multi-disciplinary team should have already considered the quality and availability of data before formulating the research question that drives the design of the CRF.

12. Perform a pilot test of the CRF on a subset of patients:

If a systematic barrier to completing the data fields in the CRF is found, revise the research question and/or CRF. See section [troubleshooting](#) for approaches to common issues in data acquisition, such as erroneous or missing clinical data.

Use physiological measures of respiratory failure

⌚ Timing: Roughly, 5 min per patient.

13. Calculate the P/F ratio:

$$P/F = P_aO_2 / F_iO_2$$

14. Calculate the ROX index for each subject and time point:

$$\text{ROX index} = (\text{SpO}_2 / \text{F}_i\text{O}_2) / \text{Respiratory Rate}$$

Note: Lower P/F ratio or lower ROX index correlates to worse hypoxemic respiratory failure.

Use inflammatory biomarker trends to predict respiratory decline

⌚ Timing: 1–4 h.

15. Pre-specify a statistical analysis of inflammatory and other biomarkers over time, instead of static values at a single time point, and their association to the primary and secondary end points that have clinical and/or pathophysiological significance.
16. Challenge any associations with sensitivity analyses to test the risks of confounding bias.

Note: This protocol instructs the comparison of “mild” versus “progressive” respiratory failure patient groups. This protocol’s model excels at discriminating factors associated with progression of respiratory failure—a dynamic process. Consequently, the utility of this model is maximized by examining the dynamics of biomarker trends over time rather than as static values at single time points.

EXPECTED OUTCOMES

The examples below detail the outcomes after applying this protocol to COVID-19 pneumonia in our study by Mueller et al.

Method 1: Develop patient cohorts that discriminate progressive respiratory failure (steps 6–10)

We applied the model of mild, progressive, and severe respiratory failure in this protocol to patients admitted to a hospital with COVID-19 pneumonia (Mueller et al., 2020). We first used the model of “ICU” versus “non-ICU” patients and we confirm previous studies that found elevated inflammatory biomarkers in the “ICU” cohort. This “ICU” cohort includes both the “progressive” and “severe” respiratory failure groups in this protocol’s model.

After applying this protocol’s model, we found that “mild” and “progressive” groups had largely overlapping ranges of inflammatory biomarkers at hospital admission (i.e., C-reactive protein [CRP], d-dimer and procalcitonin), with both “mild” and “progressive” groups distinct from the “severe” group. Thus, this protocol’s model of “mild” versus “progressive” respiratory failure captured important heterogeneity that was lost in a simpler model of “ICU” versus “non-ICU.”

Method 2: Use physiological measures of respiratory failure (steps 13 and 14)

Quantitative analysis of possible determinants of hypoxemic respiratory failure can be facilitated by use of physiological measures of respiratory failure, such as the P/F ratio and ROX index. In our study of patients hospitalized with COVID-19 pneumonia, CRP level associated with P/F ratio ($\rho = -0.54$, $p < 0.001$) (Mueller et al., 2020). We adapted the ROX index for non-intubated patients not on HFNC. We calculated the ROX index at time zero and 24 h after hospital presentation. By Area Under the ROC curve (AUROC) analysis, the ROX index had moderate predictive value for respiratory deterioration and intubation later in the hospital course and mildly under-perform the change in CRP. Hospital day 0 ROX index had AUROC 0.71 [0.58–0.83], and Day 1 ROX index had AUROC 0.68 [0.51–0.85]. Please see Table S2 for example dataset comparing CRP values over time to calculated ROX index.

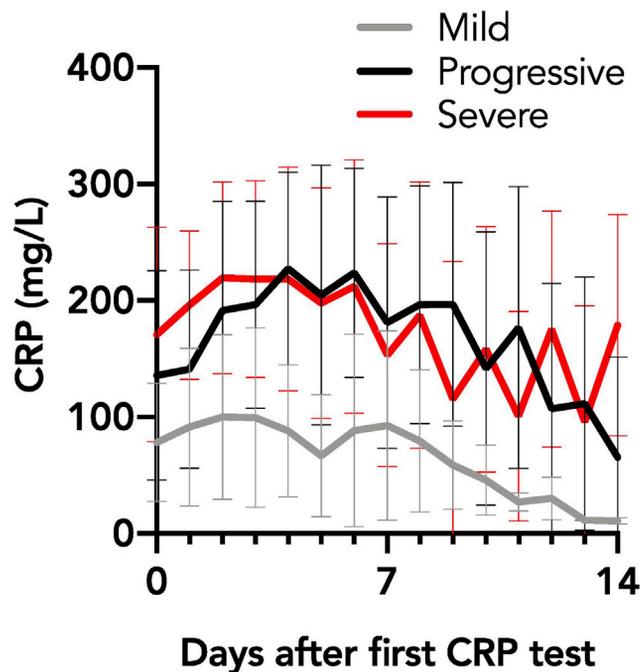


Figure 1. Trend of mean CRP values for mild, progressive, and severe cohorts, from first collected value to 14 days after admission

(Reproduced from Mueller, Tamura et al., 2020)

Method 3: Use inflammatory biomarker trends to predict respiratory decline (steps 15 and 16).

In our study of COVID-19 pneumonia, both “mild” and “progressive” groups of patients were initially admitted to the hospital with non-critical illness. By the definitions described in this protocol, the “mild” group remained stable, and the “progressive” group had respiratory decline later during their hospital course to critical illness. The admission CRP levels for the “progressive” group of patients had a statistically significant elevation compared to the “mild” group (113 [63–198] versus 74 [39–118] mg/L, $p = 0.03$) (Mueller et al., 2020). However, the “mild” and “progressive” groups had largely overlapping ranges for admission CRP levels, which limited the practical value and clinical significance of admission CRP for predicting later respiratory decline.

However, the “mild” and “progressive” groups had markedly distinct CRP trends over the first 72 h of their hospital course (Figures 1 and 2; Mueller et al., 2020). The CRP trend of “mild” patients remained flat, while “progressive” patients had an early rise in CRP levels. We compared how admission CRP level (a single, static value) compared to rise in CRP (change in CRP over the first 72 h) in predicting later intubation. By AUROC analysis, change in CRP (0.74 [0.59–0.88], $p = 0.02$, cutoff value for change in CRP = 13) modestly outperformed the single value of CRP at admission (0.68 [0.55–0.80], $p = 0.008$, cutoff value for admission CRP = 146).

QUANTIFICATION AND STATISTICAL ANALYSIS

The following describes key considerations for statistical analysis. This approach was used in our study of COVID-19 detailed in “expected outcomes”.

1. Assess whether the data have a normal distribution:
 - a. Test for a normal distribution using the Shapiro-Wilk test, D’Agostino and Pearson test, Kolmogorov-Smirnov test, or Anderson-Darling test. Typically, one sets the significance (α) at 0.05.

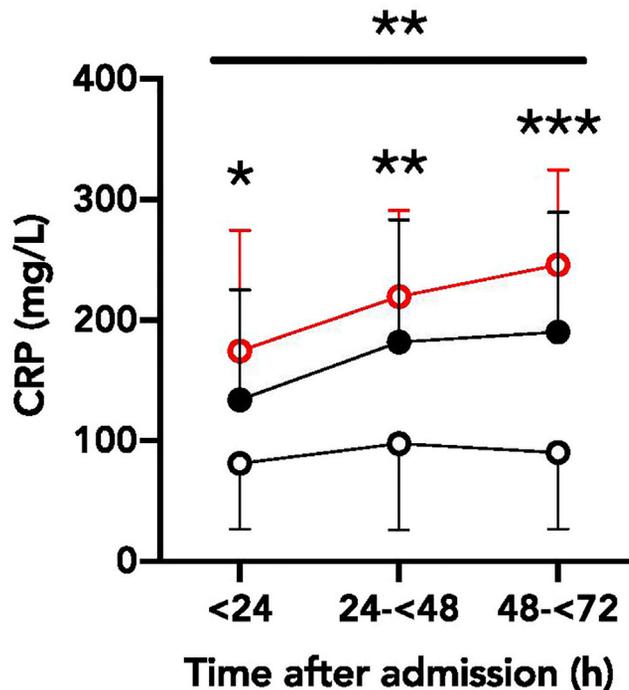


Figure 2. CRP values for mild, progressive, and severe cohorts collected 0–24 h, 24–48 h, and 48–72 h after admission to the hospital
(Reproduced from Mueller, Tamura et al., 2020)

Note: When data have a normal distribution, a parametric test can be utilized. However, when data do not follow normal distribution, a non-parametric test of significance should be used.

Note: Since the central limit theorem dictates that datasets tend towards a normal distribution with large N, studies with small N are most vulnerable to non-normal distributions. For example, the distribution of inflammatory biomarkers in several retrospective cohort studies of COVID-19 were non-normally distributed.

2. Assess the null hypothesis with the proper statistical test by considering the number of comparisons and whether the distribution is normal.
3. For descriptive statistics, calculate the mean with standard deviation for normal distributions or median values with interquartile ranges for non-normal distributions.
4. For comparisons of continuous variables, if two groups, use the Student's t test (if normal) or Mann-Whitney U test (if non-normal). For comparison of continuous variables among three groups, such as the mild, progressive, and severe groups in our study, use analysis of variance (ANOVA) with Turkey's multiple comparison or the Kruskal-Wallis test with Dunn's multiple comparison. Carefully note whether one way or two-way comparisons are required.
5. For comparisons of binary variables, use the chi-square or fisher exact test.
6. To assess correlation between two variables, use the Spearman rank correlation coefficient.
7. To analyze repeated measures, use a mixed effects model with Sidak's multiple comparison.
8. To assess the predictive performance of variables for a binary outcome, utilize Area Under the Receiver Operating Characteristics curve (AUROC) with 95% confidence interval and the optimal cutoff value per Younden's J statistic.
9. In our study (Mueller et al., 2020), statistical analysis was performed using R (version 3.6.1, The R Project) and Prism (version 8.4.1, GraphPad).

LIMITATIONS

Method 1: Develop patient cohorts that discriminate progressive respiratory failure (steps 6–10)

Early versus late progressive respiratory decline. The “mild, progressive, severe” respiratory failure approach to patient cohorts is a key improvement in the granularity of patient groups compared to “ICU” versus “non-ICU” models while still maintaining simplicity. However, the “progressive” patient group in this protocol does not distinguish between patients with respiratory decline early in their hospital course, such as the first four days of hospitalization, and patients with later decline in their hospital course, such as one week or greater after admission. Disease pathophysiology and windows of therapeutic intervention likely to differ between patients with early or later respiratory decline.

Method 2: Use physiological measures of respiratory failure (steps 13 and 14)

Quantification of physiological processes. The ROX index uses respiratory rate as a proxy for work of breathing. Inevitably, respiratory rate does not capture the full complexity of the physiological processes in “increased work of breathing,” which includes the interplay of lung compliance, resistance, strength of respiratory muscles, use of accessory respiratory muscles and respiratory drive.

Method 3: Use inflammatory biomarker trends to predict respiratory decline (steps 15 and 16)

Bias in availability of lab values sent by clinicians in retrospective studies. Retrospective, observational studies depend on lab tests or physiological parameters measured by clinicians. Thus, the types of variables are limited, and there may be bias in the data, as addressed in the section [troubleshooting](#). This issue applies to all clinical data points but is a key problem in assessing biomarkers, as only a limited range are available to clinicians and may be inconsistently ordered. For example, in our study of COVID-19 pneumonia, we had a full dataset of CRP measurements but many missing subjects for IL-6 and no data on IL-1 and other cytokines that may be upstream of CRP ([Mueller et al., 2020](#)).

TROUBLESHOOTING

Problem 1

Retrospective cohort study (steps 1–5)

All retrospective cohort studies are vulnerable to multiple biases, such as selection bias, misclassification bias, and confounding bias, with the ultimate caveat that association, not causation, is established.

Potential solution 1

Minimizing biases requires multiple approaches. To reduce selection bias, prospective randomization is ideal. For a retrospective cohort, the study subjects should be defined in a manner that minimizes bias, such as all consecutive patients over a time period without major changes in clinical practice. One approach to reducing misclassification bias is rigorous examination of the accuracy of the clinical data before initiation of the study. For example, if a study is examining the association between a patient’s prior use of statin and their respiratory outcomes, the investigators must assess whether home medications are consistently and accurately recorded. A particularly problematic misclassification bias would be introduced if the accuracy of the clinical records varied in a non-random way with clinical care. In this hypothetical example, home medications may be more often omitted, with patients incorrectly labeled as a statin non-user, when a hospital is busiest during the Covid-19 surge. Thus, rates of misclassification may associate with other changes to clinical care that occurred during the busiest periods of the Covid-19 surge. The text of the manuscript, particularly the conclusion and discussion, should be rigorous in acknowledgment of the limitations of a retrospective cohort study due to bias and confounding variables.

Problem 2

Missing data (steps 2–5)

Both retrospective and prospective studies can suffer from missing or unevenly sampled data. Retrospective cohort studies rely on data recorded by clinical staff and so are vulnerable to non-random missing data. For example, nurses may record clinical data less frequently or less accurately for patients that are less ill or recovering. Thus, missing data can lead to exclusion of subjects and introduce bias in the dataset.

Potential solution 2

Several strategies can be employed to mitigate the effect of missing data. If the data are missing completely at random (MCAR), investigators can perform listwise deletion, in which a subject is removed completely from the analysis. Investigators can alternatively employ pairwise deletion, in which the subject is deleted from analyses that depend on the missing data, but the subject is included in other analyses for which data are available. Investigators should pre-specify their definition and strategy for missing data that triggers subject deletion. Alternatively, the data may be missing at random (MAR), meaning that the likelihood of missing data is linked to a patient characteristic, but random within that subgroup of patients with that characteristic. The missing data can be considered random for analyses within that subgroup. For example, patients admitted from the study site's emergency department may tend to have different inflammatory biomarkers sent and different initial clinical care than patients that presented to a different hospital's emergency department and were directly admitted to an inpatient floor of the study site. Here, treating missing biomarker values could be treated as random within the subgroup of patients that were admitted from the emergency department, but it could introduce systematic error to treat missing lab values as random across the entire cohort.

If the missing data are random, an alternative approach to avoid exclusion of subjects is imputation to replace the missing data value with a calculated value. Older imputation techniques include replacing the missing value with the mean of non-missing observations (mean substitution) or with values from a regression model (regression substitution). Newer methods include the multiple imputation method (Gold and Bentler, 2000), maximum likelihood estimation (Enders and Peugh, 2004), and full information maximum likelihood (FIML) method (Enders and Bandalos, 2001). Full discussion of these statistical methods is beyond the scope of this protocol focused on respiratory decline, and investigators should refer to primary statistical sources. In brief, the effect of imputation can be assessed with the multiple imputation method, in which a missing data point is replaced with slightly different values multiple times, which allows estimate of the standard error introduced by imputation. In maximum likelihood estimation methods, no data are imputed, but the original dataset, with missing values, is analyzed. In FIML, a likelihood function is calculated for each subject based on the observed data to model the function most likely to have led to the observed dataset. Maximum likelihood methods generate estimates of standard error. It bears repeating that these imputation methods assume that the missing data are random and not associated with subject characteristics. Unfortunately, in a retrospective study, missing data can be non-random. For example, the more labile or critically ill patients may have the most data missing from the chart if the clinical providers were more occupied with direct clinical care and had less time to record data. This protocol is focused on challenges specific to studies of respiratory failure, so investigators are urged to refer to general biostatistical resources for in-depth discussion of data techniques (MIT Critical Data, 2016).

Problem 3

Errors in clinical data measurement and recording (steps 2–5)

Studies that rely on physiological data recorded by clinicians can suffer from inaccuracy that varies by type of parameter. The quality of charted clinical data varies by location, data type, and mode of

data entry, such as manual input by healthcare staff versus automated pull of data from clinical monitoring equipment to the electronic medical record. The institution and location within the institution can have different norms. In addition, the status of the institution and sub-locations may change over time in a non-random manner, such as with increased clinical volume during a surge of COVID-19 infections. Furthermore, even precisely recorded data may not accurately reflect the patient's clinical status. For example, a SpO₂ by pulse oximeter may not accurately reflect oxygenation due to poor placement and can have racial bias due to skin pigmentation (Feiner et al., 2007; Sjoding et al., 2021).

Potential solution 3

Some errors in clinical data can be detected by research staff with clinical expertise. These research staff are aware of the pitfalls of different types of clinical measurements and can assess the validity of the data point by clinical context. For example, to calculate the ROX index for our study, staff with clinical experience knew that the respiratory rate is much more prone to inaccurate measurement and charting than the SpO₂. For SpO₂, experienced researchers evaluated whether an outlier low SpO₂ value reflected true respiratory decline or a technical artifact by examining clinical context, such as an isolated low SpO₂ value without any response by clinical staff paired with a nursing note in the clinical flowchart that noted a likely reason for artifact at that time, such as the patient's agitation that led to poor measurements.

Problem 4

Rapid variability in clinical measurements over time (steps 2–5)

A patient's clinical measurements, even if exhibiting an overall trend, can have hour to hour variability that makes data collection challenging.

Potential solution 4

First, the investigators should determine which unit of time reflects the study's hypothesis and practical constraints of data availability and generation. In our comparison of inflammatory biomarkers and physiology indices in Covid-19, since laboratory biomarkers were measured once per hospital day, we assessed the ROX index daily. Second, for clinical measurements that vary throughout the study's time unit, the investigators should select a consistent protocol tailored to the clinical measurement. For example, the ROX index, based on SpO₂ and respiratory rate, varies throughout the day, so for our study we used the day's lowest value. Alternatively, a study could select the day's highest value or the value at a set time each day. The selection of a set time can be informed by the clinical site. For example, an institution may know that respiratory rates are most accurately assessed during the nursing assessment at the beginning of their shift.

Problem 5

Differences between clinical institutions and clinical practice (steps 2–5)

In multi-center studies, differences among clinical institutions can introduce error or complicate the study even if data collection is standardized across sites. Key differences among centers include characteristics of the patient population, clinical practices that affect patient care, or clinical practices that affect data collection. Some differences will be noted by standard assessments, such as patient demographics, scores of severity of illness on admission, or obvious aspects of clinical care such as percentage of HFNC, NIPPV and IMV. Other differences may be more subtle and not typically assessed. For example, the families of patients may elect to withdraw care and end life-supporting measures at different rates and timing due to factors unique to each center, such as the local beliefs of the community or the structure of palliative care consultation.

Potential solution 5

The investigators should pre-specify clinical variables that will be assessed for association with the primary and secondary endpoints, such as patient demographics and characteristics of the clinical care. Clinical variables should be informed by discussion with clinicians and clinical researchers at each site, who will be most familiar with local factors in the patient population and clinical practice patterns that could confound outcomes.

Problem 6

Type II error (steps 15 and 16)

These studies are vulnerable to type II error, in which an underpowered study incorrectly supports the null hypotheses.

Potential solution 6

Careful consideration of power calculations to drive sample size is key to avoiding type II error. For these calculations, selection of the power (typically 80% to 95%) and significance (typically $\alpha = 1\%$ – 5%) are individual to each study. A major challenge is predicting effect size, particularly in novel clinical situations. A pilot retrospective cohort study can be invaluable to assess data quality, troubleshoot data gathering and estimate effect size. Effect size and required sample size will be affected by the definition of groups. Moving from two groups (e.g., “ICU” versus “non-ICU”) to the three groups suggested in this protocol (“mild,” “progressive,” and “severe”) may increase the total number of patients required for the study in many cases. However, in some cases, since the three groups in this protocol may associate clinical pathophysiology, dividing into three groups may increase the effect size seen in pilot experiments and lead to power calculations that require a smaller sample size than expected.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Edy Kim MD, PhD (ekim11@bwh.harvard.edu).

Materials availability

No new reagents or materials were generated as part of this study.

Data and code availability

Patient data reviewed in this study is not publicly available due to restrictions on patient privacy and data sharing. There was no new code developed as part of this study.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xpro.2021.100545>.

ACKNOWLEDGMENTS

This work was supported by NIAMS T32 AR007530-35 (to A.A.M.).

AUTHOR CONTRIBUTIONS

A.A.M., T.T., and J.D.S. developed these protocols. A.A.M., T.T., J.R.D., C.P.C., and H.H. tested these protocols. C.P.C., A.A.M., and E.Y.K. prepared this manuscript. E.Y.K. supervised this study.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Beigel, J., Tomashek, K., Dodd, L., Mehta, A., Zingman, B., Kalil, A., Hohmann, E., Chu, H., Luetkemeyer, A., Kline, S., et al. (2020). Remdesivir for the treatment of COVID-19 – final report. *N. Engl. J. Med.* *383*, 1812–1826.
- Brown, S.M., Duggal, A., Houe, P.C., Tidswell, M., Khan, A., Exline, M., Park, P.K., Schoenfeld, D.A., Liu, M., Grissom, C.K., et al. (2017). Nonlinear imputation of Pa₂/FiO₂ from SpO₂/FiO₂ among mechanically ventilated patients in the ICU: a prospective observational study. *Crit. Care Med.* *45*, 1317–1324.
- Catterall, M., Kazantzis, G., and Hodges, M. (1967). The performance of nasal catheters and a face mask in oxygen therapy. *Lancet* *289*, 415–417.
- Coudroy, R., Frat, J.-P., Girault, C., and Thille, A. (2020). Reliability of methods to estimate the fraction of inspired oxygen in patients with acute respiratory failure breathing through non-rebreather reservoir bag oxygen mask. *Thorax* *75* (9), 805–807.
- Cummings, M., Baldwin, M., Abrams, D., Jacobson, S., Meyer, B., Balough, E., Aaron, J., Claassen, J., Rabbani, L., and Hastie, J. (2020). Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet* *395*, 1763–1770.
- Enders, C.K., and Bandalos, D.L. (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Structural Equation Modeling* *8*, 430–457.
- Enders, C.K., and Peugh, J.L. (2004). Using an EM Covariance Matrix to Estimate Structural Equation Models With Missing Data: Choosing an Adjusted Sample Size to Improve the Accuracy of Inferences. *Structural Equation Modeling* *11*, 1–19.
- Feiner, J., Severinghaus, J., and Bickler, P. (2007). Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. *Anesth. Analg.* *105*, S18–S23.
- Feng, J., Anoushiravani, A., Tesoriero, P., Ani, L., Meftah, M., Schwarzkopf, R., and Leucht, P. (2020). Transcriptions error rates in retrospective chart reviews. *Orthopedics* *1*, e404–e408.
- Ferguson, N.D., Fan, E., Camporota, L., Antonelli, M., Anzueto, A., Beale, R., Brochard, L., Brower, R., Esteban, A., Gattinoni, L., et al. (2012). The Berlin definition of ARDS: an expanded rationale, justification, and supplementary material. *Intensive Care Med.* *38*, 1573–1582.
- Frat, J.P., Thille, A., Mercat, A., Girault, C., Ragot, R., Perbet, S., Prat, G., Boulain, T., Morawiec, E., Cottreau, A., et al. (2015). High flow oxygen through nasal cannula in acute hypoxemic respiratory failure. *N. Engl. J. Med.* *372*, 2185–2196.
- Gold, M., and Bentler, P. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling* *7*, 319–355.
- Haljasmagi, L., Salumets, A., Rumm, A., Jurgenson, M., Krassohina, E., Remm, A., Sein, H., Kareinen, L., Vapalati, O., and Sironen, T. (2020). Longitudinal proteomic profiling reveals increased early inflammation and sustain apoptosis proteins in severe COVID-19. *Sci. Rep.* *10*, 20533.
- Hardavella, G., Karampinis, I., Frille, A., Sreter, K., and Rousalova, I. (2019). Oxygen devices and delivery systems. *Breathe* *15*, 108–116.
- Maass, C., Kuske, S., Lessing, C., and Schrappe, M. (2015). Are administrative data valid when measuring patient safety in hospitals? A comparison of data collection methods using a chart review and administrative data. *Int. J. Qual. Health Care.* *27*, 305–313.
- MIT Critical Data (2016). *Secondary Analysis of Electronic Health Records* (Springer).
- Lee, S., Donovan, L., Austin, P., Gong, Y., Liu, P., Rouleau, J., and Tu, J. (2005). Comparison of coding heart failure and comorbidities in administrative and clinical data use in outcomes research. *Med. Care* *43*, 182–188.
- Mueller, A., Tamura, T., Crowley, C., Degrado, J., Haider, H., Jezmir, J., Keras, G., Penn, E., and Massaro, A. (2020). Inflammatory biomarker trends predict respiratory decline in COVID-19 patients. *Cell Rep. Med.* *1*, 100144.
- Patel, M., Chowdhury, J., Mills, N., Marron, R., Gangemi, A., Dorey-Stein, Z., Yousef, I., Zheng, M., Tragesser, L., Giurinatano, J., et al. (2020). ROX index predicts intubation in patients with COVID-19 pneumonia and moderate to severe hypoxemic respiratory failure receiving high flow nasal therapy. *medRxiv*. <https://doi.org/10.1101/2020.06.30.20143867>.
- Petrilli, C., Jones, S., Yang, J., Rajagopalan, H., O'Donnel, L., Chernyak, Y., Tobin, K., Cerfolio, R., Francois, F., and Horwitz, L. (2020). Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* *369*, m1966.
- Sjoding, M., Dickson, R., Iwashyna, T., Gay, S., and Valley, T. (2021). Racial bias in pulse oximetry measurement. *N. Engl. J. Med.* *383*, 2477–2478.
- Zhang, X., Tan, Y., Ling, Y., and Lu, G. (2020). Viral and host factors related to the clinical outcome of COVID-19. *Nature* *583*, 437–440.
- Zhao, H., Wang, H., Sun, F., Lyu, S., and An, Y. (2017). High flow nasal cannula oxygen therapy is superior to conventional oxygen therapy but not to noninvasive mechanical ventilation on intubation rate: a systematic review and meta-analysis. *Crit. Care.* *21*, 184.