

**Prognostic association of routinely measured biomarkers in patients  
admitted to critical care: a systematic review**

Yize I. Wan<sup>\*a, b</sup>, Adam Brayne<sup>a, b, c</sup>, Ryan W. Haines<sup>a, b</sup>, Zudin A. Puthuchearu<sup>a, b</sup>, John R.  
Prowle<sup>a, b</sup>

*<sup>a</sup>Adult Critical Care Unit, The Royal London Hospital, Barts Health NHS Trust, London, UK*

*<sup>b</sup>William Harvey Research Institute, Queen Mary University of London, London, UK*

*<sup>c</sup>Northern Devon Healthcare NHS Trust, Barnstaple, Devon, UK*

\*Corresponding author:

Yize I. Wan, PhD

Critical Care Research Office,

The Royal London Hospital,

London E1 1BB,

United Kingdom

Email: [yize.wan@qmul.ac.uk](mailto:yize.wan@qmul.ac.uk)

Tel: +44 20 3594 0351

Author's Contributions: All authors reviewed and approved the final manuscript. YW and JP were responsible for the study concept and YW, RH, and JP for study design. Data extraction and analysis was undertaken by YW, AB, and RH. YW, RH, ZP, and JP wrote the manuscript.

Word count: 4320 (excluding tables)

Abstract word count: 195

# **Prognostic association of routinely measured biomarkers in patients admitted to critical care: a systematic review**

## **Abstract**

**Purpose:** To examine reported prognostic associations of routine blood measurements in the intensive care unit.

**Materials and Methods:** We searched PubMed, EMBASE through 28<sup>th</sup> May 2020 to identify all studies in adult critical care investigating associations between parameters measured **routinely** in whole blood, plasma or serum, and length of stay or mortality. Registration: PROSPERO; CRD42019122058.

**Results:** A total of 128 studies, reporting 28 different putative prognostic biomarkers, met eligibility criteria. Those most frequently examined were red cell distribution width, neutrophil-to-lymphocyte ratio, C-reactive protein, and platelet count. A higher red cell distribution width, a lower platelet count, and a higher neutrophil-to-lymphocyte ratio were consistently associated with both increased mortality and length of stay. A lower level of albumin was consistently associated with greater mortality. **C-reactive protein was inconsistent.** Most studies (n=110) used regression modelling with wide variation in variable selection and covariate-adjustment; none externally validated the proposed predictive models.

**Conclusions:** Simple regression models have so far proved inadequate for the complexity of data available from routine blood sampling in critical care. Adoption of a direct causal framework may help better assess mechanistic processes, aid design of future studies, **and guide clinical decision making using routine data.**

Keywords: systematic reviews; intensive care; biomarker; mortality; length of stay

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of interest: The authors report no declarations of interest.

## **Introduction**

Panels of laboratory blood tests are routinely used for diagnosis and assessment of hospitalized patients (Badrick, 2013). In patients admitted to critical care, blood samples are frequently taken throughout the course of admission generating a large quantity of repeated measures data in every patient, only a little of which is directly utilized in clinical decision making (Ezzie et al., 2007). Large numbers of observational studies have investigated the association of routine laboratory test results to clinical outcomes within different subgroups of critically ill patients, including sepsis, acute kidney injury, and acute respiratory distress syndrome (Gerlach, 2018, Heilmann et al., 2019, Kelly et al., 2018, Opal and Wittebole, 2020, Vincent and Teixeira, 2014, Koyner et al., 2019, Srisawat and Kellum, 2020, Ware et al., 2010). Furthermore, there has been a growing number of studies examining less frequently used indices in routine blood panels, or measures derived from combinations of routinely measured blood tests, including red cell distribution width and neutrophil-to-lymphocyte ratio which have been proposed as novel surrogate measures of pathophysiological pathways (Gluck et al., 2018, Hwang et al., 2017, Güell et al., 2019, Bazick et al., 2011, Kim et al., 2013, Purtle et al., 2014).

There is a long history of including haematological and biochemical parameters in models to estimate expected mortality and morbidity in critically ill patients (Vincent and Moreno, 2010, Breslow and Badawi, 2012a, Breslow and Badawi, 2012b). As mortality from critical illness continues to reduce, identification of high risk subgroups may be important to enable continued improvement in outcomes (Sjoding and Cooke, 2015). Furthermore, it is becoming accepted that simple survival does not constitute intensive care unit (ICU) success. An increasingly recognized syndrome of persistent critical illness is associated with significant ongoing morbidity, and poor post-ICU outcomes (Hermans et al., 2019, Jeffcote et al., 2019). However these phenotypes are poorly understood (Viglianti et al., 2019). Consequently, while there is

considerable interest in understanding biological signatures of critical illness phenotypes from routine blood investigations, interpretation and clinical application of these potential biomarkers remains ambiguous.

Newer approaches using routinely collected data have used machine learning techniques for prediction of ICU morbidity and mortality with the aim of improving prognostic models and classifying sub-populations (Shillan et al., 2019). Studies have most commonly used supervised methods such as support vector machines, random forests, and neural networks. **However**, such hypotheses driven approaches may tend to reinforce pre-existing concepts rather than identifying novel groups. Furthermore, the majority of published studies have analysed data on relatively small patient populations, have no external validation, and lack complete reporting of predictive accuracy. **There is currently limited statistical literature on how to evaluate the performance of these methods in clinical practice, particularly when more than one biomarker is involved or when outcome variables are classified using more than two categories. In these cases, traditional measures such as receiver operating characteristic (ROC) curves and the area under the ROC curve (AUC) are not applicable (Li et al., 2019).** Thus, despite an apparent wealth of information, it is difficult to clearly determine true clinical associations and clinical implications of these models (Ware, 2017, Tyler et al., 2018). In this systematic review, we aimed to examine and collate reported associations between routine blood test results carried out in critical care and outcomes using ICU length of stay and mortality up to 90 days to identify useful candidates for future large-scale outcomes modelling either alone or in combination.

## **Materials and methods**

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guidelines (Moher et al., 2009). The study protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42019122058).

### ***Study selection***

A study was eligible if it aimed to test the association between one or more routinely measured biomarker(s) and length of stay, and/or mortality. These outcomes were defined as 1) ICU or hospital length of stay, 2) mortality on the ICU or in hospital or within 90 days of ICU admission. Candidate biomarkers were any indices available from blood tests measured as part of a routine daily panel in international ICUs. Studies were restricted to those carried out in adults. Prospective and retrospective studies were included. Case reports and case series, as well as non-research publications such as literature reviews, editorials and correspondences were excluded. Only **full text articles** in English were included.

### ***Search strategy***

Searches were carried out using two separate databases MEDLINE (PubMed) and Excerpta Medica dataBASE (EMBASE) from inception through 28<sup>th</sup> May 2020. Search strategies are detailed in the online supplement, briefly these involved combinations of blood test name or terms for abnormalities in the measured parameter, terms indicative of critical illness and terms indicative of outcomes of relevance. Reference lists from prior systematic reviews were searched manually for additional publications of relevance.

### ***Data extraction***

Data extraction was performed independently by three independent investigators (YW, AB, RH) and in duplicate using predefined data abstraction forms. A random check of 10% of cases were performed to ensure accuracy and any disagreements were resolved by a fourth investigator (JP). Abstracted data included study characteristics, demographic data, biomarker data and outcome data.

### ***Quality of evidence and risk of bias***

Quality of evidence and individual study risk of bias within studies was assessed using the Newcastle-Ottawa Scale (NOS) criteria (Wells et al., 2013). Scores were given based on selection of study groups (four points), comparability of groups (two points), and ascertainment of exposure and outcomes (three points). Statistical methods and models were assessed including statistical model type, method for variable selection for adjustment, and model performance metrics. In addition, we recorded studies that used an *a priori* statistical analysis plan and those that adhered to the appropriate STROBE reporting guidelines. We assessed citations and geographical location of studies. Study citations were reviewed using Pubmed, Crossrefs, and Google Scholar.

### ***Data synthesis***

As we anticipated insufficient methodological heterogeneity between study cohorts due to large variations in biomarker measurements and outcome measurements, calculating overall measures of effect was not considered appropriate. Consequently, a number of summary assessments were undertaken to critically appraise studies and collate the best available evidence. For biomarkers showing consistent effects across multiple studies, we reviewed proposed mechanisms and underlying hypotheses reported in manuscripts. We then developed a directed acyclic graph (DAG) using current evidence and scientific reasoning to examine the

causal association between reported biomarkers and outcomes (Textor et al., 2016). We used the DAGitty R package and website (<http://dagitty.net>) to inspect the DAG and identify suggested variables to adjust for in future modelling.



## Results

### *Search results and study characteristics*

Results of the literature search are shown in Figure 1, of 799 records we included 128 studies meeting our eligibility criteria for qualitative synthesis. Identified studies analysed different biomarkers either individually or using a composite measure of two related biomarkers (Table 1). The majority of study populations were focused within mixed ICU admissions (n=74), followed by medical admissions (n=33) then surgical admissions (n=21) (online E2). Study designs were a combination of cohort studies (n=112) and case-control studies (n=16). All studies described a retrospective method of data collection.

[Figure 1 near here] [Table 1 near here]

### *Biomarker associations*

There was a total of 28 routinely measured parameters examined within identified studies, 20 with one to four records, 3 with five to nine records, and 5 with ten or more records. Red cell distribution width (RDW) was the biomarker most frequently described (Table 1). Definitions of biomarker varied in terms of timing and frequency of measurements used. Single measures within 72 hours of admission were most common (68%) followed by 26% of studies using repeated measurements, and 6% of studies not clearly reporting when measures were taken. As studies used different biomarker cut offs, we were unable to compare size of associations between studies. Data reporting differed in using continuous variables (41.9%), categorical variables (46.0%), and 12.1% utilized increases between selected time-points.

[Table 2 near here]

For studies reporting the most frequently examined 5 biomarkers, relevant data extractions are summarized in Table 2 (remaining biomarkers are presented in online E1). Amongst defined

primary outcome variables, 10.5% of studies reported length of stay either on ICU and/or in hospital, the remaining 89.5% reported mortality. The majority of studies reported either death in ICU or in-hospital or mortality at 28 or 30-days. Seventeen studies reported mortality to 90 days and two studies also reported longer term mortality to 1 and 5-years respectively. Patient inclusion criteria varied widely including sub-groups of patients admitted with intra-abdominal sepsis, sub-arachnoid haemorrhage, cardiac surgery, pneumonia, and renal failure. We were able to report overall trends for the most commonly investigated biomarkers. The biomarkers showing most consistently reported effects were RDW, platelet count, neutrophil-to-lymphocyte ratio (NLR), and albumin. A higher RDW, a lower platelet count, and a higher NLR was associated with both mortality and increased ICU length of stay (Figure 2). A lower level of albumin was associated with mortality. All other reported biomarkers showed opposing directions of effect between different studies.

[Figure 2 near here]

### ***Study impact***

Using maximum number of citations, 23.4% of studies were cited by ten or fewer other studies, 20.2% of studies were cited by more than 50 other studies (Table 1). Within only the 128 studies assessed in this review, 26 studies were cited by another study included in this review providing evidence of knowledge accumulation (online E5). However, of the top 5 cited papers, none were cited by a prospective study or randomized controlled trial.

### ***Study methodology***

All studies used a retrospective methodology and the chosen diagnostic test accuracy varied per study and target biomarker. Of studies reporting methods of statistical analysis, the majority (n=110) used logistic or linear regression, Cox proportional hazards survival modelling, or

both. Of 110 studies which adjusted for confounders, covariates were selected based on bivariate screening (n=41), stepwise selection (n=21), author based decision based on plausibility (n=26), Akaike information criterion (AIC) model fitting (n=1), and univariate p value threshold (n=1), the rest used unclear selection criteria (n=20).

The most common method reported for model selection and performance was the Area under the ROC Curve (AUC) (n=61, 49.2%) and ranged from 0.57 to 0.91. **Risk thresholds were not clearly reported.** When multiple variables were included in a model, four studies reported criteria-based methods to guide model selection (2 used the log-likelihood and 2 used AIC). None of the studies reported pre-specified statistical analysis plans. None of the studies used an external validation cohort to test the finalized model. Only one followed STROBE reporting guidelines.

### ***Exploratory analysis***

In an exploratory approach we constructed a directed acyclic graph (DAG) for RDW and CRP, two of the most consistently reported biomarkers (Figure 3). For RDW, the DAG suggested three adjustment sets to test the direct effect of reactive erythropoiesis in response to the effects of prolonged critical illness. These comprised nutritional intake, cell hypoxia, and inflammation, which might be approximated using surrogate markers, respectively Malnutrition Universal Screening Tool (MUST) score, lactate, and CRP. Similarly, for CRP, an adjustment set to test the direct effect of prolonged inflammation on prolonged ICU stay included severity of illness and catabolism which could be approximated using admission SOFA/APACHE II scoring and urea-to-creatinine ratio. Full code and methods for DAG construction including references to mechanistic pathways are detailed in the online supplement.

[Figure 3 near here]

## Discussion

### *Summary of results*

We performed a systematic review on routinely collected biomarkers and associations with prolonged length of stay and mortality following critical illness. Despite identifying a large number of independent studies constituting a significant amount of data, due to non-uniformity in variable definition, collection and reporting, interpretation was limited. Consequently, we were unable to increase the utility of these biomarkers by establishing consistent cut-offs in homogenous patient-groups. Strength of evidence for any reported biomarker remains low precluding the use of specific algorithms to guide clinical decision making, which remains based on heuristic interpretation of multiple data points by clinicians at the bedside.

For the majority of reported biomarkers, the certainty of evidence for **associations** with outcomes was low or moderate, primarily due to imprecision in effect size and risk of bias. There was no evidence of pre-specification of statistical analyses, conversely there was evidence of selective reporting, two important standards suggesting low-quality evidence. There was limited use of longitudinal data despite the potential to greatly enrich predictions with appropriate analysis. Failure to consider longitudinal effects may account for reports of opposing directions of association for the same biomarker even within similar study populations.

Some biologically plausible biomarkers have shown consistent effect in determining poor outcomes following critical illness. However, despite generating interest, citations and having coherent pathophysiological hypotheses, there has been little clinical uptake of decision models based on these parameters. All the studies we identified focused on determining statistical significance without evaluating underlying mechanisms or formally assessing utility of the

measured parameter in altering clinical practice. However, populations, cut-offs, outcomes measures, and adjustment for confounders varied considerably between studies. As a result, clear interpretation and meaningful application of these reported association remains lacking.

### ***Biological frameworks for prospective justification of variable selection***

In an exploratory analysis, we used a DAG methodology to explore RDW and CRP as examples of formalising variable selection for inclusion into a study model *a priori*, as an alternative more powerful technique to simple baseline characteristic adjustments in multivariable analysis. This method allows examination of proposed casual pathways. For example, a potential link between reactive erythropoiesis to low baseline levels and prolonged inflammation with prolonged duration of critical illness and increased mortality is pre-specified. Visualisation of multiple variables which are likely to be inter-related allows investigators to pre-specify an adjustment set, reducing both inadequate correction for confounding and collider bias, both of which may mask real effects (Wunsch et al., 2006). This also gives an explicit statement of proposed mechanisms and avoids variable selection by bias-prone methods (Lederer et al., 2019). Not all biomarkers follow the same trajectory and likely demonstrate different predictive value at different time points of disease. Improved understanding of underlying biological mechanisms will also strengthen the basis for setting biomarker thresholds and time of assessment thus improving the ability to detect true effects if present (Leisman et al., 2020).

### ***Strengths and limitations***

To our knowledge, this is the first systematic review evaluating routine blood tests as biomarkers within critical care populations. The strengths of this review include a comprehensive literature search, adherence to our pre-registered protocol, and focus on studies

which have specifically tested one or more biomarker(s) as their primary objective. However, we lacked sufficient data to explore subgroup effects, and significant study heterogeneity prevented formal meta-analysis and restricted further quantitative analysis of data. **There may be some variation in which biomarkers are routinely measured between ICUs.** Significant potential for publication bias exists, as evidenced by serial citation patterns and timescale of publications. Due to the breadth of our search terms, variations in the structure of study reporting, and exclusions of non-English articles, it is possible that some studies were missed.

### ***Recommendations***

In order to improve the quality and utility of future research, we propose three methodological recommendations for biomarker studies examining cohorts in critical illness. These are, to identify a prospective causal framework facilitating adequate selection of covariates and adjustment variables (Grimes and Schulz, 2002), use of a prospective analysis plan to reduce bias and multiple testing (Berger et al., 2012), and use of an external validation cohort to test accuracy and generalisability of proposed predictive models (Altman and Royston, 2000). Studies should also adhere to structured reporting guidelines such as STROBE or TRIPOD to improve comparability and transparency (von Elm et al., 2008, Collins et al., 2015). It should also be noted that many studies were carried out in similar populations, in particular a number used the freely accessible, single centre, MIMIC database. This highlights the need to increase diversity of study populations across different centres and countries to ensure generalizability. Where **possible**, patient level data sharing should be facilitated to increase power to assess association across multiple studies and allow validation of findings between studies. These principles are of key importance with increasing use of machine learning approaches, as findings will still require ongoing prospective evaluation to elucidate mechanistic understanding.

## *Conclusions*

Simple regression models with imprecise variable selection and adjustment have so far proved inadequate to derive consistent clinically applicable models from the complexity of routine blood test data. Previous research has had limited applicability, compounded by a lack of external validation. Machine learning approaches continue to face similar methodological challenges. Adoption of a direct causal framework may be needed *a priori* to better assess causal associations, increase biological understanding, and maximize learning from these studies. **Improving current methodology in observational biomarker research in this field may help to identify candidates for use in predictive modelling and guide clinical decision making using routine data.**



## References

- ALTMAN, D. G. & ROYSTON, P. 2000. What do we mean by validating a prognostic model?  
*Stat Med*, 19, 453-73.
- BADRICK, T. 2013. Evidence-based laboratory medicine. *Clin Biochem Rev*, 34, 43-6.
- BAZICK, H. S., CHANG, D., MAHADEVAPPA, K., GIBBONS, F. K. & CHRISTOPHER, K. B. 2011. Red cell distribution width and all-cause mortality in critically ill patients.  
*Crit Care Med*, 39, 1913-21.
- BERGER, M. L., DREYER, N., ANDERSON, F., TOWSE, A., SEDRAKYAN, A. & NORMAND, S. L. 2012. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. *Value Health*, 15, 217-30.
- BRESLOW, M. J. & BADAWI, O. 2012a. Severity scoring in the critically ill: part 1-- interpretation and accuracy of outcome prediction scoring systems. *Chest*, 141, 245-252.
- BRESLOW, M. J. & BADAWI, O. 2012b. Severity scoring in the critically ill: part 2: maximizing value from outcome prediction scoring systems. *Chest*, 141, 518-527.
- COLLINS, G. S., REITSMA, J. B., ALTMAN, D. G. & MOONS, K. G. 2015. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *J Clin Epidemiol*, 68, 134-43.
- EZZIE, M. E., ABEREGG, S. K. & O'BRIEN, J. M. 2007. Laboratory testing in the intensive care unit. *Crit Care Clin*, 23, 435-65.
- GERLACH, A. T. 2018. Sepsis Biomarkers...The Long and Winding Road. *Crit Care Med*, 46, 1194-1195.
- GLUCK, E., NGUYEN, H. B., YALAMANCHILI, K., MCCUSKER, M., MADALA, J., CORVINO, F. A., ZHU, X. & BALK, R. 2018. Real-world use of procalcitonin and

- other biomarkers among sepsis hospitalizations in the United States: A retrospective, observational study. *PLoS One*, 13, e0205924.
- GRIMES, D. A. & SCHULZ, K. F. 2002. Bias and causal associations in observational research. *Lancet*, 359, 248-52.
- GÜELL, E., MARTÍN-FERNANDEZ, M., DE LA TORRE, M. C., PALOMERA, E., SERRA, M., MARTINEZ, R., SOLSONA, M., MIRÓ, G., VALLÈS, J., FERNÁNDEZ, S., CORTÉS, E., FERRER, V., MORALES, M., YÉBENES, J. C., ALMIRALL, J. & BERMEJO-MARTIN, J. F. 2019. Impact of Lymphocyte and Neutrophil Counts on Mortality Risk in Severe Community-Acquired Pneumonia with or without Septic Shock. *J Clin Med*, 8.
- HEILMANN, E., GREGORIANO, C. & SCHUETZ, P. 2019. Biomarkers of Infection: Are They Useful in the ICU? *Semin Respir Crit Care Med*, 40, 465-475.
- HERMANS, G., VAN AERDE, N., MEERSSEMAN, P., VAN MECHELEN, H., DEBAVEYE, Y., WILMER, A., GUNST, J., CASAER, M. P., DUBOIS, J., WOUTERS, P., GOSELINK, R. & VAN DEN BERGHE, G. 2019. Five-year mortality and morbidity impact of prolonged versus brief ICU stay: a propensity score matched cohort study. *Thorax*, 74, 1037-1045.
- HWANG, S. Y., SHIN, T. G., JO, I. J., JEON, K., SUH, G. Y., LEE, T. R., YOON, H., CHA, W. C. & SIM, M. S. 2017. Neutrophil-to-lymphocyte ratio as a prognostic marker in critically-ill septic patients. *Am J Emerg Med*, 35, 234-239.
- JEFFCOTE, T., FOONG, M., GOLD, G., GLASSFORD, N., ROBBINS, R., IWASHYNA, T. J., DARVALL, J., BAGSHAW, S. M. & BELLOMO, R. 2019. Patient characteristics, ICU-specific supports, complications, and outcomes of persistent critical illness. *J Crit Care*, 54, 250-255.

KELLY, B. J., LAUTENBACH, E., NACHAMKIN, I., COFFIN, S. E., GERBER, J. S., FUCHS, B. D., GARRIGAN, C., HAN, X., BILKER, W. B., WISE, J., TOLOMEO, P., HAN, J. H. & PROGRAM, C. F. D. C. A. P. C. P. E. 2018. Combined Biomarkers Predict Acute Mortality Among Critically Ill Patients With Suspected Sepsis. *Crit Care Med*, 46, 1106-1113.

KIM, C. H., PARK, J. T., KIM, E. J., HAN, J. H., HAN, J. S., CHOI, J. Y., HAN, S. H., YOO, T. H., KIM, Y. S., KANG, S. W. & OH, H. J. 2013. An increase in red blood cell distribution width from baseline predicts mortality in patients with severe sepsis or septic shock. *Crit Care*, 17, R282.

KOYNER, J. L., ZARBOCK, A., BASU, R. K. & RONCO, C. 2019. The impact of biomarkers of acute kidney injury on individual patient care. *Nephrol Dial Transplant*.

LEDERER, D. J., BELL, S. C., BRANSON, R. D., CHALMERS, J. D., MARSHALL, R., MASLOVE, D. M., OST, D. E., PUNJABI, N. M., SCHATZ, M., SMYTH, A. R., STEWART, P. W., SUISSA, S., ADJEI, A. A., AKDIS, C. A., AZOULAY, É., BAKKER, J., BALLAS, Z. K., BARDIN, P. G., BARREIRO, E., BELLOMO, R., BERNSTEIN, J. A., BRUSASCO, V., BUCHMAN, T. G., CHOKROVERTY, S., COLLOP, N. A., CRAPO, J. D., FITZGERALD, D. A., HALE, L., HART, N., HERTH, F. J., IWASHYNA, T. J., JENKINS, G., KOLB, M., MARKS, G. B., MAZZONE, P., MOORMAN, J. R., MURPHY, T. M., NOAH, T. L., REYNOLDS, P., RIEMANN, D., RUSSELL, R. E., SHEIKH, A., SOTGIU, G., SWENSON, E. R., SZCZESNIAK, R., SZYMUSIAK, R., TEBOUL, J. L. & VINCENT, J. L. 2019. Control of Confounding and Reporting of Results in Causal Inference Studies. Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. *Ann Am Thorac Soc*, 16, 22-28.

- LEISMAN, D. E., HARHAY, M. O., LEDERER, D. J., ABRAMSON, M., ADJEI, A. A., BAKKER, J., BALLAS, Z. K., BARREIRO, E., BELL, S. C., BELLOMO, R., BERNSTEIN, J. A., BRANSON, R. D., BRUSASCO, V., CHALMERS, J. D., CHOKROVERTY, S., CITERIO, G., COLLOP, N. A., COOKE, C. R., CRAPO, J. D., DONALDSON, G., FITZGERALD, D. A., GRAINGER, E., HALE, L., HERTH, F. J., KOCHANNEK, P. M., MARKS, G., MOORMAN, J. R., OST, D. E., SCHATZ, M., SHEIKH, A., SMYTH, A. R., STEWART, I., STEWART, P. W., SWENSON, E. R., SZYMUSIAK, R., TEBOUL, J. L., VINCENT, J. L., WEDZICHA, J. A. & MASLOVE, D. M. 2020. Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals. *Crit Care Med*, 48, 623-633.
- LI, J., GAO, M. & D'AGOSTINO, R. 2019. Evaluating classification accuracy for modern learning approaches. *Stat Med*, 38, 2477-2503.
- MOHER, D., LIBERATI, A., TETZLAFF, J. & ALTMAN, D. G. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*, 6, e1000097.
- OPAL, S. M. & WITTEBOLE, X. 2020. Biomarkers of Infection and Sepsis. *Crit Care Clin*, 36, 11-22.
- PURTLE, S. W., MOROMIZATO, T., MCKANE, C. K., GIBBONS, F. K. & CHRISTOPHER, K. B. 2014. The association of red cell distribution width at hospital discharge and out-of-hospital mortality following critical illness\*. *Crit Care Med*, 42, 918-29.
- SHILLAN, D., STERNE, J. A. C., CHAMPNEYS, A. & GIBBISON, B. 2019. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Crit Care*, 23, 284.

- SJODING, M. W. & COOKE, C. R. 2015. Chronic critical illness: a growing legacy of successful advances in critical care\*. *Crit Care Med*, 43, 476-7.
- SRISAWAT, N. & KELLUM, J. A. 2020. The Role of Biomarkers in Acute Kidney Injury. *Crit Care Clin*, 36, 125-140.
- TEXTOR, J., VAN DER ZANDER, B., GILTHORPE, M. S., LISKIEWICZ, M. & ELLISON, G. T. 2016. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol*, 45, 1887-1894.
- TYLER, P. D., DU, H., FENG, M., BAI, R., XU, Z., HOROWITZ, G. L., STONE, D. J. & CELI, L. A. 2018. Assessment of Intensive Care Unit Laboratory Values That Differ From Reference Ranges and Association With Patient Mortality and Length of Stay. *JAMA Netw Open*, 1, e184521.
- VIGLIANTI, E. M., KRUSER, J. M. & IWASHYNA, T. 2019. The heterogeneity of prolonged ICU hospitalisations. *Thorax*, 74, 1015-1017.
- VINCENT, J. L. & MORENO, R. 2010. Clinical review: scoring systems in the critically ill. *Crit Care*, 14, 207.
- VINCENT, J. L. & TEIXEIRA, L. 2014. Sepsis biomarkers. Value and limitations. *Am J Respir Crit Care Med*, 190, 1081-2.
- VON ELM, E., ALTMAN, D. G., EGGER, M., POCOCK, S. J., GÖTZSCHE, P. C., VANDENBROUCKE, J. P. & INITIATIVE, S. 2008. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*, 61, 344-9.
- WARE, L. B. 2017. Biomarkers in Critical Illness: New Insights and Challenges for the Future. *Am J Respir Crit Care Med*, 196, 944-945.
- WARE, L. B., KOYAMA, T., BILLHEIMER, D. D., WU, W., BERNARD, G. R., THOMPSON, B. T., BROWER, R. G., STANDIFORD, T. J., MARTIN, T. R.,

MATTHAY, M. A. & NETWORK, N. A. C. T. 2010. Prognostic and pathogenetic value of combining clinical and biochemical indices in patients with acute lung injury. *Chest*, 137, 288-96.

WELLS, G., SHEA, B., O'CONNELL, D., PETERSON, J., WELCH, V., LOSOS, M. & TUGWELL, P. 2013. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.

WUNSCH, H., LINDE-ZWIRBLE, W. T. & ANGUS, D. C. 2006. Methods to adjust for bias and confounding in critical care health services research involving observational data. *J Crit Care*, 21, 1-7.

## Figure legends

Figure 1. PRISMA flowchart for the systematic review and meta-analysis.

Figure 2. Summary of studies for biomarkers with 5 or more records. RDW: red cell distribution width, NLR: neutrophil-to-lymphocyte ratio, CRP: C-reactive protein, Na: sodium. Effect on outcomes defined as positive if increased and negative if decreased mortality or increased length of stay. See online supplement E7 for references and E8 for key to study names.

Figure 3. Proposed directed acyclic graphs relating red cell distribution with (RDW) and C-reactive protein (CRP) to mechanisms underlying prolonged critical illness. See online supplement E4 for full code and references to mechanistic pathways.