Citation for published version (APA):
Just because things aren’t different, doesn’t mean they are the same:

biomarker patterns in ARDS

Manu Shankar-Hari PhD MSc¹,², Gordon D Rubenfeld MD MSc³
¹Guy’s and St Thomas’ NHS Foundation Trust, ICU support Offices, 1st Floor, East Wing, St Thomas’ Hospital, SE1 7EH, UK; ²Division of Infection, Immunity and Inflammation, Kings College London, SE1 9RT; ³Interdepartmental Division of Critical Care Medicine, Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, D5 03, Toronto, Ontario, M4N 3M5

*Corresponding author and reprints

Gordon D. Rubenfeld, MD MSc FRCPC
Professor of Medicine
Interdepartmental Division of Critical Care Medicine, University of Toronto
Chief, Program in Trauma, Emergency and Critical Care, Sunnybrook Health Sciences Center
2075 Bayview Avenue, Room D108c
Toronto, ON M4N 3M5
TEL: 416-480-6100 x2895
FAX: 416-480-4999

Financial support: None

Acknowledgements

Manu Shankar-Hari is supported by the UK National Institute for Health Research Clinician Scientist Award (NIHR-CS-2016-16-011). The views expressed in this publication are those of the author(s) and not necessarily those of the National Health Service, the National Institute for Health Research, or the United Kingdom Department of Health.
Is an inferior wall myocardial infarction the same disease as an anterior wall myocardial infarction? Is breast cancer the same disease as ovarian cancer? These are fundamental questions whose answers require insights into the pathophysiology of the disease and the efficacy of different therapies. These are questions whose answers change over time, for example, as our understanding of cancer has advanced from morphologic sameness to cell replication mechanism sameness, some breast cancers are, in important ways, the same as some ovarian cancers(1). Acute respiratory distress syndrome (ARDS) is a common problem in critically ill patients, representing 10% of all ICU admissions and nearly 25% of patients requiring mechanical ventilation(2). Biologically, ARDS is characterised as acute lung inflammation, associated with increased pulmonary vascular permeability, increased lung weight and acute loss of aerated lung tissue(3). Clinically, ARDS is, as it has been at least since 1967, a syndrome of acute hypoxemic respiratory failure with bilateral pulmonary infiltrates that cannot be attributed solely to heart failure and usually occurs in the setting of injury or infection(4).

The question posed by García-de-Acilu, et al, in this issue of CCM is: *Is ARDS managed with high flow nasal cannula oxygen (HFNC) the same as ARDS managed with endotracheal ventilation?*(5) This question has limited implications for clinicians as it does not address the question of the efficacy of HFNC in managing ARDS and there are no unique pharmacologic therapies to offer patients on HFNC if we decide they have ARDS. It is primarily of interest to researchers who may want support to enrol patients on HFNC in trials of therapy for ARDS and mechanistic studies of the syndrome. The question arises because patients on HFNC and other forms of supplemental oxygen are caught in a loophole in the most recent attempt to formalize the definition of ARDS which requires a PaO₂/FiO₂ < 300
on a PEEP or CPAP of at least 5 cm water(4). Because of variability in measuring the FiO₂ and the delivered PEEP on HFNC (6), as well as the potential differences in the chest radiograph due to delivered tidal volume and respiratory rate, it is possible that some patients classified as having ARDS on HFNC might not meet these criteria if they had been intubated. Note, we specifically say “might not meet these criteria” instead of “might not have ARDS” as the developers of the Berlin Definition explicitly note that we don’t have a laboratory test for “having ARDS”(4).

One approach is to simply argue that the method by which a physician chooses to deliver oxygen cannot change the underlying mechanism of what she is treating. HFNC may select a group of patients with less severe ARDS if the delivered FiO₂ and PEEP are both lower than estimated, but there is no reason to suspect these patients have a different mechanism causing their acute hypoxemic respiratory failure assuming they meet other diagnostic criteria. Those persuaded by this theoretical argument will find the paper by García-de-Acilu M et al superfluous and its findings intuitive. However, providing an empiric answer to the question of whether HFNC-ARDS is the same as intubated-ARDS is interesting, if for no other reason, because it requires us to address two fundamental issues: one biologic and the other statistical.

In ARDS, the immune responses that lead to the structural and functional disruption of the alveolar endothelial and epithelial barrier are well described(3). Briefly, leukocytes sense presence of pathogens by detecting pathogen associated molecular patterns (PAMPs) such as lipopolysaccharide and tissue damage associated molecular patterns (DAMPs) such as mitochondrial nucleic acids with pattern recognition receptors, to generate a multitude of immune activation mediators through inflammasome and signalosome complexes (3, 7). Thus, the
alveolar barrier disruption is the final common pathway resulting from a number of host responses that are not unique to ARDS, as similar host responses are seen in sepsis and trauma (3, 8, 9).

Blood provides an accessible window to measure and evaluate biomarkers, which are indicators of normal biologic processes, pathogenic processes, or pharmacologic responses to therapeutic interventions. Recently, Terpstra et al reported a systematic review that categorized biomarkers into those associated with either ARDS diagnosis or risk of death once ARDS is diagnosed (10). The diagnostic markers with strongest association were biomarkers of epithelial dysfunction such as Krebs von den Lungen-6 (KL-6), soluble receptor for advanced glycation end products (RAGE) and endothelial dysfunction such as von Willebrand factor (vWF). The ARDS prognosis markers in this study were markers of either inflammation (interleukins IL-4, IL-2, IL-1beta), or endothelial dysfunction (angiopoi etin-2) or epithelial dysfunction (KL-6) (10). Even though some of the associations are strong, the diagnostic or prognostic performance of individual biomarkers is poor and there is considerable overlap between ARDS and at-risk populations (11). There are no accepted cut-offs for a biomarker or panel of biomarkers that accurately separate ARDS from non-ARDS hypoxemic respiratory failure. There has been some progress in using panels of clinical variables combined with biomarkers to identify endotypes of ARDS (12).

In addition to figuring out what biologic measures convincingly identify patients with ARDS, there are also two interesting statistical concerns in this study. Patients receiving HFNC might be different than other patients with ARDS for reasons that have nothing to do with the mechanism of their ARDS. These confounding variables likely relate to the reasons that physicians might choose to
manage certain patients without intubation. It isn’t surprising that patients managed with HFNC tended to have lower severity of illness as measured by SOFA and APACHE II and a different ARDS risk factor distribution, however, it also wouldn’t be surprising that a difference in biomarkers in these patients might reflect these factors as much as any aspect of the mechanism for lung injury. Therefore, the approach taken by García-de-Acilu et al, to match on a propensity score, simply a technique to allow near matching on multiple confounding variables, was wise.

We use statistical hypothesis testing to account for the role of chance in scientific studies. Our standard procedure for this, the P value, tells us how unlikely the results of the study are if the null hypothesis of equivalence is true. The problem is, the research question in this study is proving the null hypothesis; and while small datasets can make equivalence unlikely, no amount of observed data can prove equivalence. It is well known that concluding equivalence from a non-statistically significant result is an error, an error of Type II. If it were not an error, equivalence studies would be easy, one would simply perform small under-powered experiments that will yield high P values. Nevertheless, questions of equivalence and non-inferiority are important and frequently addressed in clinical trials. These analyses require a crucial piece of information that is lacking in this ARDS equivalence study.

When trialists seek to demonstrate that a new therapy is equivalent or at least non-inferior to a standard therapy they must decide how much of a difference doesn’t matter. In the SYNERGY trial of enoxaparin versus unfractionated heparin for non-ST segment elevation acute coronary syndromes(13), investigators pre-specified that as long as enoxaparin did not increase the composite outcome of all-cause death or nonfatal myocardial infarction by more than 10% they would conclude that enoxaparin was non-inferior to heparin. That means that if heparin was actually
better by a 5% improvement in outcome, they would still conclude that enoxaparin was no worse. They could have decided 10% was too big a difference to miss and used 5% and would have needed a larger sample size. There are guidelines for selecting clinically persuasive and scientifically justified non-inferiority and equivalence margins, but they are debated and subject to interpretation (14).

How big a difference in biomarkers doesn’t matter? The important information in the García-de-Acilu study is not in the Figure 1 boxplots and P values comparing HFNC-ARDS with intubated-ARDS; it is in the third column of Table-3. This table tells us how big the differences were in the biomarkers and, more importantly, how big a difference could have been missed in this study. For RAGE, on average, patients with HFNC-ARDS had levels that were 280.56 pg/ml lower than matched patients who were intubated, but the data were consistent with patients having levels that were 969.41 pg/ml lower. In the largest study addressing the question, the difference in RAGE levels in ARDS patients compared to at-risk controls was 1093 pg/ml; a difference that arguably would have been missed in this study (15). If a difference of 969.41 pg/ml in RAGE, and the extremes of the confidence intervals around other biomarkers in Table 3, are inconsequential, then this study provides some convincing statistical support for the claim of similarity. If we actually don’t even know what constitutes an important difference in biomarkers, then the empiric claim for similarity is problematic.
References