



# HHS Public Access

Author manuscript

*Pediatr Crit Care Med.* Author manuscript; available in PMC 2022 May 01.

Published in final edited form as:

*Pediatr Crit Care Med.* 2021 May 01; 22(5): 496–498. doi:10.1097/PCC.0000000000002702.

## Statistical Note: Confounding and Causality in Observational Studies

**Christopher Horvat, MD MHA**

Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA

No teacher in medicine is better than experience. Observational studies that examine large repositories of data, such as electronic health record (EHR) databases and multicenter registries, afford a vantage of collective experience beyond what any individual can gain at the bedside. Recent years have seen a surge of observational literature propelled by increasing accessibility to EHR databases and a growing familiarity and availability of observational research tools such as the open-source R programming language and software libraries that facilitate advanced modeling.(1) Observational insights through pattern identification and statistical tests of association share both a bedside and mathematical space with causation. However, association and causation are not the same. Awareness of the potential role of confounding is essential when working to distill meaningful insights from observational data and studies, a task that can seem particularly daunting with the bloom of increasingly esoteric “machine learning” analyses throughout the medical literature, including *Pediatric Critical Care Medicine*.

In this issue of the *Journal*, a study by Beshish and colleagues describes an association between hyperoxemia during cardiopulmonary bypass and operative mortality among children with heart disease.(2) In the accompanying editorial, Dr. Peters highlights the natural temptation that we all have to infer causal relationships from patterns proximate to one another in time and space.(3) Indeed, several studies have examined the relationship between hyperoxemia and outcome among critically ill children and identified associations between supra-physiologic oxygen levels and mortality.(2, 4-12) It is a research question with a biologically plausible basis in that reactive oxygen species are recognized to wreak molecular havoc in cells and tissue. It is also a question readily investigated in data-rich intensive care unit registries and databases, in which multiple observations of  $F_iO_2$ ,  $P_aO_2$ ,  $S_pO_2$  (fractional inspired oxygen, arterial partial pressure of oxygen, pulse oximetry oxyhemoglobin saturation, respectively) abound and can be analyzed in relation to outcome. But the fact remains that sicker patients are in general exposed to greater amounts of oxygen, either out of necessity or as part of resuscitation bundles. Hence, oxygen has been deemed a highly suspicious culprit, but just because oxygen is dressed in stripes does not make it a convict! So how does one derive meaningful information from observational studies while disentangling associations from confounding to form actionable insights?

One can start by considering the basic definition of confounding, which is “a blurring or mixing of effects.”(13, 14) A confounder has an association with both the disease, and/or predictor, and/or exposure, and the outcome of interest – but is not an effect of the exposure or a factor in the outcome or property of the underlying disease. For example, a large forest

fire causes substantially more damage than a small bush fire and requires far more fire fighters to combat the blaze. Without knowledge of the causal mechanisms at play, one might identify that a greater number of fire fighters is associated with more catastrophe and could erroneously infer a causal relationship along the lines of more fire fighters results in more damage.

Since causal relationships are the relevant targets for treatments and interventions, we need strategies to control for confounding. Let us consider these using the example of pediatric hyperoxemia studies. As already mentioned, illness severity is an important confounder in many studies examining the relationship between hyperoxemia and outcome. A common technique to control for illness severity is multivariable modeling, often incorporating some version of a validated measure of disease severity, such as the pediatric risk of mortality score (PRISM), pediatric index of mortality (PIM) or pediatric logistic organ dysfunction score (PELOD).(2, 4, 6, 8-11) Each of these scoring systems has a track record that includes external validation and robust predictive performance. However, the development of each of these scores has had to account for the competing aims of being sufficiently comprehensive representations of illness severity while also being calculable using checklists and commonly available data. Further, the relatively low rate of mortality in pediatric critical care means that these models are trained on imbalanced datasets, with more to teach about life than death. Accordingly, these models only account for a subset, and certainly not all, of data representing illness severity. Modeling last rites as a variable, for example, may be associated with mortality after adjusting for severity with an established score, but does not indicate a causal relationship between last rites and death. Instead, identifying an association is only the first step, and further steps are necessary to tease out potential causality. Fortunately, readers can ask some relatively straightforward questions of their own before drawing conclusions from observational data.

Bradford Hill outlined a set of classic criteria to assess causality in observed relationships: strength, consistency, specificity, temporality, biological gradient (dose-response), plausibility, and coherence.(14) *Strength* refers to the observed effect size of an association. In the context of hyperoxemia research, this refers to how much more likely death is for patients exposed to hyperoxemia compared to those with apparent normoxemia, commonly expressed as odds. Greater odds of death provides some indication that the relationship is more likely to be causal, though there is no established threshold for declaring a causal relationship. *Consistency* refers to the reproducibility of the finding. As the number of observational studies citing an association between hyperoxemia and mortality accumulate from a range of institutions and patient subpopulations, the finding appears increasingly consistent. *Specificity* refers to an association consistently observed in a select context, such as hyperoxemia being associated with death post-cardiac arrest. The *temporality* criterion necessitates the timing of the exposure to be sensibly related to outcome. For example, hyperoxemia defined by  $P_aO_2$  obtained during the episode of critical illness, rather than during apnea testing when assessing for brain death. The *biological gradient* criterion would require the magnitude of hyperoxemia, measured either by cumulative exposure, magnitude of an instance of exposure, or both, to have a proportionate effect on the odds of death. Several studies have demonstrated a plausible “U-shaped” relationship between oxygen level and outcome, with death more likely at both

the extreme highs and lows of  $P_aO_2$  values, though it is less clear if this is a relationship defined by polynomial smoothing rather than true dose-response.(6, 8, 11) *Plausibility* asks whether a biologically sound mechanism underlies the observed association. As noted already, substantial pre-clinical evidence, some findings dating back more than a century, contributes to the enthusiastic belief that hyperoxemia is causally linked to death.(15) *Coherence* would expect that the outcome of interest, mortality in this case, would be altered by removing exposure to hyperoxemia. Guidelines for the management of pediatric cardiac arrest recommend judicious use of oxygen in resuscitation, for example, though it is difficult to discern the impact of this single intervention amidst many other evolving aspects of care.(16)

With these criteria in mind, some additional steps can be taken in the design of observational studies to further interrogate potentially causal relationships. Well-constructed propensity scores, for example, can allow researchers to incorporate a greater number of potentially confounding variables in observational models without problematic overfitting. Following identification of a significant association, assessment for lurking confounders can be performed, which quantifies the prevalence and effect size of an unmeasured confounder that would be necessary to tip the model to fail to reject the null hypothesis of the observed association. Observational randomized analyses can also be performed to deal with uncontrollable confounders. This is commonly performed by dividing large datasets into a training set for model development and a testing set for model assessment. When smaller datasets preclude division into train and test groups, data splitting can be performed using k-fold cross validation. Intrinsically random, machine-learning approaches such as random trees, forests and fields and neural networks may be coupled with feature importance analyses to assess whether hyperoxemia is predictive of mortality in large, granular databases.(17) Such techniques provide added capabilities for dealing with unmeasured confounders in observational data, but such experiments are still subject to the limitations of the dataset on which they are being applied.

Ultimately, the gold standard to deal with confounding in clinical experimentation remains the randomized controlled trial. By randomizing comparable groups of patients to alternative treatment strategies, the intervention of interest ideally becomes the distinguishing characteristic of the experimental arms. There are, of course, myriad practical challenges to randomized controlled trials in pediatric critical care, including the relative rarity of diseases, cost, special ethical considerations, and challenges obtaining consent at the appropriate time, among others.(18) However, the same circumstances that have promoted the surge in observational studies in recent years, including the increasing usability of EHR data, may also pave the way for larger, most efficient trials in our field. As an increasing number of Bradford Hill criteria are met in observational studies assessing a link between hyperoxemia and outcome among critically ill children, the findings should not be viewed as irrefutable causal associations, but rather as supporting the need to invest in more conclusive studies of this important topic.

## References

1. R: What is R? [Internet]. [cited 2021 Jan 17] Available from: <https://www.r-project.org/about.html>

2. Beshish AG, Jahadi O, Mello A, et al. : Hyperoxia during cardiopulmonary bypass is associated with mortality in infants undergoing cardiac surgery. *Pediatr Crit Care Med* 2021; 22:AA–BB
3. Peters MJ: Linking hyperoxia and harm: consequence or merely subsequence? *Pediatr Crit Care Med* 2021; 22:CC–DD
4. Ferguson LP, Durward A, Tibby SM: Relationship between arterial partial oxygen pressure after resuscitation from cardiac arrest and mortality in children. *Circulation* 2012; 126:335–342 [PubMed: 22723307]
5. Guerra-Wallace MM, Casey FLI, Bell MJ, et al. : Hyperoxia and Hypoxia in Children Resuscitated From Cardiac Arrest. *Pediatr Crit Care Med* 2013; 14:e143 [PubMed: 23392367]
6. Raman S, Prince NJ, Hoskote A, et al. : Admission PaO<sub>2</sub> and Mortality in Critically Ill Children: A Cohort Study and Systematic Review. *Pediatr Crit Care Med* 2016; 17:e444–e450 [PubMed: 27509363]
7. van Zelle L, de Jonge R, van Rosmalen J, et al. : High cumulative oxygen levels are associated with improved survival of children treated with mild therapeutic hypothermia after cardiac arrest. *Resuscitation* 2015; 90:150–157 [PubMed: 25576438]
8. Numa A, Aneja H, Awad J, et al. : Admission Hyperoxia Is a Risk Factor for Mortality in Pediatric Intensive Care. *Pediatr Crit Care Med* 2018; 19:699–704 [PubMed: 29927878]
9. Ramgopal S, Dezfulian C, Hickey RW, et al. : Association of Severe Hyperoxemia Events and Mortality Among Patients Admitted to a Pediatric Intensive Care Unit. *JAMA Netw Open* 2019; 2:e199812 [PubMed: 31433484]
10. Ramgopal S, Dezfulian C, Hickey RW, et al. : Early Hyperoxemia and Outcome Among Critically Ill Children. *Pediatr Crit Care Med* 2020; 21:e129–e132 [PubMed: 31821205]
11. Pelletier JH, Ramgopal S, Au AK, et al. : Maximum Pao<sub>2</sub> in the First 72 Hours of Intensive Care Is Associated With Risk-Adjusted Mortality in Pediatric Patients Undergoing Mechanical Ventilation. *Crit Care Explor* 2020; 2:e0186 [PubMed: 32984827]
12. Del Castillo J, López-Herce J, Matamoros M, et al. : Hyperoxia, hypocapnia and hypercapnia as outcome factors after cardiac arrest in children. *Resuscitation* 2012; 83:1456–1461 [PubMed: 22841610]
13. Jager KJ, Zoccali C, MacLeod A, et al. : Confounding: What it is and how to deal with it. *Kidney Int* 2008; 73:256–260 [PubMed: 17978811]
14. Hill AB: The Environment and Disease: Association or Causation? *Proc R Soc Med* 1965; 58:295–300 [PubMed: 14283879]
15. Smith JL: The pathological effects due to increase of oxygen tension in the air breathed. *J Physiol* 1899; 24:19–35
16. Topjian Alexis A., Raymond Tia T., Atkins Dianne, et al. : Part 4: Pediatric Basic and Advanced Life Support: 2020 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation* 2020; 142:S469–S523 [PubMed: 33081526]
17. Gallicchio C, Martín-Guerrero J, Micheli A, et al.: Randomized Machine Learning Approaches: Recent Developments and Challenges. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2017 proceedings. Bruges (Belgium), 26–28 April 2017, [i6doc.com](http://www.i6doc.com) publ., ISBN 978-287587039-1. (available from <http://www.i6doc.com/en/>).
18. Duffett M, Choong K, Hartling L, et al. : Randomized controlled trials in pediatric critical care: a scoping review. *Crit Care* 2013; 17:R256 [PubMed: 24168782]