Chapter 8
Residual Confounding Lurking in Big Data: A Source of Error

John Danziger and Andrew J. Zimolzak

Take Home Messages

- Any observational study may have unidentified confounding variables that influence the effects of the primary exposure, therefore we must rely on research transparency along with thoughtful and careful examination of the limitations to have confidence in any hypotheses.
- Pathophysiology is complicated and often obfuscates the measured data with many observations being mere proxies for a physiological process and many different factors progressing to similar dysfunction.

8.1 Introduction

Nothing is more dangerous than an idea, when you have only one…

—Emile Chartier

Big Data is defined by its vastness, often with large highly granular datasets, which when combined with advanced analytical and statistical approaches, can power very convincing conclusions [1]. Herein perhaps lies the greatest challenge with using big data appropriately: understanding what is not available. In order to avoid false inferences of causality, it is critical to recognize the influences that might affect the outcome of interest, yet are not readily measurable.

Given the difficulty in performing well-designed prospective, randomized studies in clinical medicine, Big Data resources such as the Medical Information Mart for Intensive Care (MIMIC) database [2] are highly attractive. They provide a powerful resource to examine the strength of potential associations and to test

whether assumed physiological principles remain robust in clinical medicine. However, given their often observational nature, causality can not be established, and great care should be taken when using observational data to influence practice patterns. There are numerous examples [3, 4] in clinical medicine where observational data had been used to determine clinical decision making, only to eventually be disproven, and in the meantime, potentially causing harm. Although associations may be powerful, missing the unseen connections leads to false inferences. The unrecognized effect of an additional variable associated with the primary exposure that influences the outcome of interest is known as confounding.

8.2   Confounding Variables in Big Data

Confounding is often referred to as a "mixing of effects" [5] wherein the effects of the exposure on a particular outcome are associated with an additional factor, thereby distorting the true relationship. In this manner, confounding may falsely suggest an apparent association when no real association exists. Confounding is a particular threat in observational data, as is often the case with Big Data, due to the inability to randomize groups to the exposure. The process of randomization essentially mitigates the influence of unrecognized influences, because these influences should be nearly equally distributed to the groups. However, more frequently observational data is composed of patient groups that have been distinguished based on clinical factors. For example, with critical care observational data, such as MIMIC, such "non-random allocation" has occurred simply by reaching the intensive care unit (ICU). There has been some decision process by an admitting team, perhaps in the Emergency Department, that the patient is ill enough for the ICU. That decision process is likely influenced by a host of factors, some of which are identifiable, as in blood pressure and severity of illness, and others that are not, as in "the patient just looks sick" intuition of the provider.

8.2.1   The Obesity Paradox

As an example of the subtlety of this confounding influence, let's tackle the question of obesity as a predictor of mortality. In most community-based studies [6, 7], obesity is associated with poorer outcomes: obese patients have a higher risk of dying than normal weighted individuals likely mediated by an increased incidence of diabetes, hypertension, and cardiovascular disease. However, amongst patients admitted to the ICU, obesity is a strong survival benefit [8, 9], with multiple studies elucidated better outcomes amongst obese critically ill patients than normal weighted critical ill patients.

There are potentially many explanations for this paradoxical association. On one hand, it is plausible that critically ill obese patients have higher nutritional stores and are better able to withstand the prolonged state of cachexia associated with critical illness than normal weighted patients. However, let's explore some other possibilities. Since obesity is typically defined by the body mass index (BMI) upon admission to the ICU, it is possible that unrecognized influences on body weight prior to hospitalization that independently affect outcome might be the true reason for this paradoxical association. For example, fluid accumulation, as might occur with congestive heart failure, will increase body weight, but not fat mass, resulting in an inappropriately elevated BMI. This fluid accumulation, when resulting in pulmonary edema, is generally considered a marker of illness severity and a warrants a higher level of care, such as the ICU. Thus, this fluid accumulation would prompt the emergency room team to admit the patient to the ICU rather than to the general medicine ward. Now, heart failure is typically a reasonably treatable disease process. Diuretics are an effective widely used treatment, and likely can resolve the specific factor (i.e. fluid overload) that leads to ICU care. Thus, such a patient would seem obese, but might not be, and would have a reasonable chance of survival. Compare that to another such patient, who developed cachexia from metastatic cancer, and lost thirty pounds prior to presenting to the emergency room. That patient's BMI would have dropped significantly over the few weeks prior to illness, and his poor prognosis and illness might lead to an ICU admission, where his prognosis would be poor. In the latter scenario, concluding that a low BMI was associated with a poor outcome may not be strictly correct, since it is often rather the complications of the underlying cancer that lead to mortality.

8.2.2 Selection Bias

Let's explore one last possibility relating to how the obesity paradox in critical care might be confounded. Imagine two genetically identical fraternal twins with the exact same comorbidities and exposures, presenting with cellulitis, weakness, and diarrhea, both of whom will need frequent cleaning and dressing changes. The only difference is that one twin has a normal weight, whereas the other is morbidly obese. Now, the emergency room team must decide which level of care these patients require. Given the challenges of caring for morbidly obese patient (lifting a heavy leg, turning to change), it is plausible that obesity itself might influence the emergency room's choice regarding disposition. In that case, there would be a tremendous selection bias. In essence, the obese patient who would have been generally healthy enough for a general ward ends up in the ICU due to obesity alone, where the observational data begins. Not surprisingly, that patient will do better than other ICU patients, since he was healthier in the first place and was admitted simply because he was obese.

Such selection bias, which can be quite subtle, is a challenging problem in non-randomly allocated studies. Patients groups are often differentiated by their

illness severity, and thus any observational study assessing the effects of related treatments may fail to address underlying associated factors. For example, a recent observational Big Data study attempted to examine whether exposure to proton pump inhibitors (PPI) was associated with hypomagnesemia [10]. Indeed, in many thousands of examined patients, PPI users had lower admission serum magnesium concentrations. Yet, the indication for why the patients were prescribed PPIs in the first place was not known. Plausibly, patients who present with dyspepsia or other related gastrointestinal symptoms, which are major indications for PPI prescription, might have lower intake of magnesium-containing foods. Thus, the conclusion that PPI was responsible for lower magnesium concentrations would be conjecture, since lower dietary intake would be an equally reasonable explanation.

8.2.3 Uncertain Pathophysiology

In addition to selection bias, as illustrated in the obesity paradox and PPI associated hypomagnesemia examples, there is another important source of confounding, particularly in critical care studies. Given that physiology and pathophysiology are such strong determinants of outcomes in critical illness, the ability to fully account for the underlying pathophysiologic pathways is extraordinarily important, but also notoriously difficult. Consider that clinicians caring for patients, standing at the patient's bedside in direct examination of all the details, sometimes cannot explain the physiologic process. Recognizing diastolic heart failure remains challenging. Accurately characterizing organ function is not straightforward. And if the caring physician can't delineate the underlying processes, how can observational data, so removed from the patient? It can't, and this is a huge source of potential mistakes. Let's consider some examples.

In critical care, the frequent laboratory studies that are easily measured with precise reproducibility make a welcoming target for cross sectional analysis. In the literature, almost every common laboratory abnormality has been associated with a poor outcome, including abnormalities of sodium, potassium, chloride, bicarbonate, blood urea nitrogen, creatinine, glucose, hemoglobin, etc. Many of these cross sectional studies have led to management guidelines. The important question however is whether the laboratory abnormality itself leads to a poor patient outcome, or whether instead, the underlying patient pathophysiology that leads to the laboratory abnormality is the primary cause.

Take for example hyponatremia. There is extensive observational data linking hyponatremia to mortality. In response, there have been extensive treatment guidelines on how to correct hyponatremia through a combination of water restriction and sodium administration [11]. However, the mechanistic explanation for how chronic and/or mild hyponatremia might cause a poor outcome is not totally convincing. Some data might suggest that potential subtle cerebral edema might lead to imbalance and falls, but this is not a completely convincing explanation for the association of admission hyponatremia with in-hospital death.

Fig. 8.1 Concept map of the association of kidney function, as determined by the glomerular filtration rate, as a determinant of cardiovascular morality

Many cross-sectional studies have not addressed the underlying reason for hyponatremia in the first place. Most often, hyponatremia is caused by sensed volume depletion, as might occur in liver disease and heart disease. Sensed volume is a concept describing the body's internal measure of intravascular volume, which directly affects the body's sodium avidity, and which under certain conditions affects its water avidity. Sensed volume is quite difficult to determine clinically, and there are no billing or diagnostic codes to describe it. Therefore, even though sensed volume is the strongest determinant of serum sodium concentrations in large population studies, it is not a capturable variable, and thus it cannot be included as a covariate in adjusted analyses. Its absence likely leads to false conclusions. As of now, despite a plethora of studies showing that hyponatremia is associated with poor outcomes, we collectively can not conclude whether it is the water excess itself, or the underlying cardiac or liver pathophysiologic abnormalities that cause the hyponatremia, that is of greater importance.

Let us consider another very important example. There have been a plethora of studies in the critical care literature linking renal function to a myriad of outcomes [12, 13]. One undisputed conclusion is that impaired renal function is associated with increased cardiovascular mortality, as illustrated in Fig. 8.1.

However, this association is really quite complex, with a number of important confounding issues that undermine this conclusion. The first issue is how accurately a serum creatinine measurement reflects the glomerular filtration rate (GFR). Calculations such as the Modification of Diet in Renal Disease (MDRD) equation were developed as epidemiologic tools to estimate GFR [14] but do not accurately define underlying renal physiology. Furthermore, even if one considers the serum creatinine as a measure of GFR, there are multiple other aspects of kidney functions beyond the GFR, including sodium and fluid balance, erythropoietin and activated vitamin D production, and tubular function, none of which are easily measurable, and thus cannot be accounted for.

However, in addition to confounding due to an inability to accurately characterize "renal function," significant residual confounding due to unaccounted pathophysiology is equally problematic. In relation to the association of renal function with cardiovascular mortality, there are many determinants of cardiac function that simultaneously and independently influence both the serum creatinine
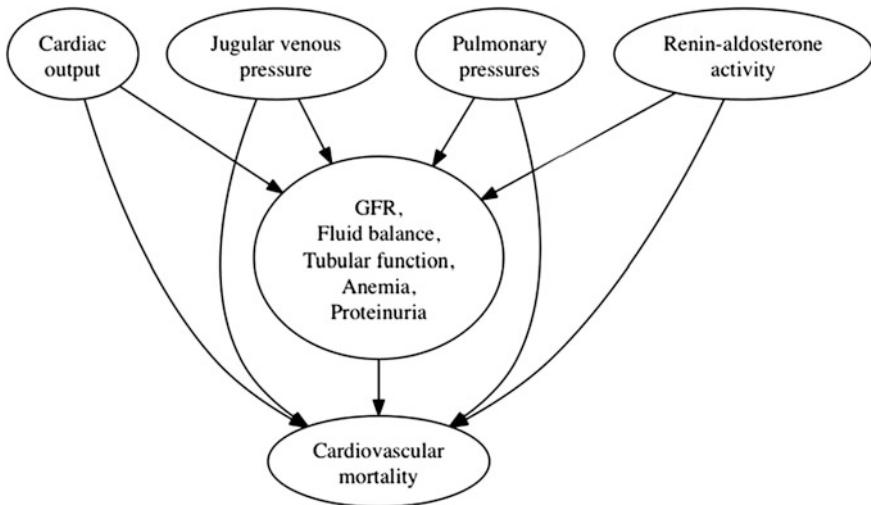
Fig. 8.2 Concept map of the association of renal function and cardiovascular mortality revealing more of the confounding influences

concentration and cardiovascular outcomes. For example, increased jugular venous pressures are a strong determinant of cardiac outcome and influence renal function through renal vein congestion. Cardiac output, pulmonary artery pressures, and activation of the renin-angiotensin-aldosterone axis also likely influence both renal function and cardiac outcomes. The concept map is likely more similar to Fig. 8.2.

Since many of these variables are rarely measured or quantified in large epidemiologic studies, significant residual confounding likely exists, and potential bias by failing to appreciate the complexity of the underlying pathophysiology is likely.

Multiple statistical techniques have been developed to account for residual confounding to non-randomization and to underlying severity of illness in critical care. Propensity scores, which attempt to better capture the factors that lead to the non-randomized allocation (i.e. the factors which influence the decision to admit to the ICU or to expose to a PPI) are used widely to minimize selection bias [15]. Adjustment using variables that attempt to capture severity of illness, such as the Simplified Acute Physiology Score (SAPS) [16], or the Sequential/ Sepsis-related Organ Failure Assessment (SOFA) score [17], or comorbidity adjustment scores, such as Charlson or Elixhauser [18, 19], remain imprecise, as does risk adjustment with area under the receiver operating characteristic curve (AUROC). Ultimately, significant confounding cannot be adjusted away by the most sophisticated statistical techniques, and thoughtful and careful examination of the limitations of any observational study must be transparent.

8.3   Conclusion

In summary, tread gently when harvesting the power of Big Data, for what is not seen is exactly what may be of most interest. Be clear about the limitations of using observational data, and suggest that most observational studies are hypothesis generating and require more well designed studies to better address the question at hand.

References

 1. Bourne PE (2014) What big data means to me. J Am Med Inf Assoc. 21(2):194–194
 2. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L-W, Moody G et al (2011) Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database. Crit Care Med 39(5):952–960
 3. Patel CJ, Burford B, Ioannidis JPA (2015) Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J Clin Epidemiol 68 (9):1046–1058
 4. Tzoulaki I, Siontis KCM, Ioannidis JPA (2011) Prognostic effect size of cardiovascular biomarkers in datasets from observational studies versus randomised trials: meta-epidemiology study. BMJ 343:d6829
 5. Greenland S (2005) Confounding. In: Armitage P, Colton T (eds) Encyclopedia of biostatistics, 2nd edn.
 6. National Task Force on the Prevention and Treatment of Obesity (2000) Overweight, obesity, and health risk. Arch Intern Med 160(7):898–904
 7. Berrington de Gonzalez A, Hartge P, Cerhan JR, Flint AJ, Hannan L, MacInnis RJ et al (2010) Body-mass index and mortality among 1.46 million white adults. N Engl J Med 363 (23):2211–2219
 8. Hutagalung R, Marques J, Kobylka K, Zeidan M, Kabisch B, Brunkhorst F et al (2011) The obesity paradox in surgical intensive care unit patients. Intensive Care Med 37(11):1793–1799
 9. Pickkers P, de Keizer N, Dusseljee J, Weerheijm D, van der Hoeven JG, Peek N (2013) Body mass index is associated with hospital mortality in critically ill patients: an observational cohort study. Crit Care Med 41(8):1878–1883
10. Danziger J, William JH, Scott DJ, Lee J, Lehman L-W, Mark RG et al (2013) Proton-pump inhibitor use is associated with low serum magnesium concentrations. Kidney Int 83(4): 692–699
11. Verbalis JG, Goldsmith SR, Greenberg A, Korzelius C, Schrier RW, Sterns RH et al (2013) Diagnosis, evaluation, and treatment of hyponatremia: expert panel recommendations. Am J Med 126(10 Suppl 1):S1–S42

12. Apel M, Maia VPL, Zeidan M, Schinkoethe C, Wolf G, Reinhart K et al (2013) End-stage renal disease and outcome in a surgical intensive care unit. Crit Care 17(6):R298
13. Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS et al (2010) Chronic kidney disease prognosis consortium. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet 375(9731):2073–2081
14. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D (1999) A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of diet in renal disease study group. Ann Intern Med 130(6):461–470
15. Gayat E, Pirracchio R, Resche-Rigon M, Mebazaa A, Mary J-Y, Porcher R (2010) Propensity scores in intensive care and anaesthesiology literature: a systematic review. Intensive Care Med 36(12):1993–2003
16. Le Gall JR, Lemeshow S, Saulnier F (1993) A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA 270(24):2957–2963
17. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H et al (1996) The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. On behalf of the working group on sepsis-related problems of the European society of intensive care medicine. Intensive Care Med 22(7):707–710
18. Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40 (5):373–383
19. Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity measures for use with administrative data. Med Care 36(1):8–27