



Scholarly Commons

Faculty Publications

2018

Big Data Analytics: What Can Go Wrong

Sharona Hoffman

Case Western University School of Law, sharona.hoffman@case.edu

Follow this and additional works at: https://scholarlycommons.law.case.edu/faculty_publications

 Part of the [Health Law and Policy Commons](#)

Repository Citation

Hoffman, Sharona, "Big Data Analytics: What Can Go Wrong" (2018). *Faculty Publications*. 2024.
https://scholarlycommons.law.case.edu/faculty_publications/2024

This Article is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholarly Commons.

BIG DATA ANALYTICS: WHAT CAN GO WRONG

SHARONA HOFFMAN*

Abstract: It is not uncommon to read that long-held beliefs about medical treatments have been dislodged by new studies. For example, there is now doubt as to whether women should undergo annual mammograms, previously a cornerstone of cancer screening. Hormone replacement therapy for menopausal women, once considered highly suspect in light of worrisome research findings, is now being reconsidered as a beneficial therapy. These reversals trouble and confuse many Americans.

This Article explores why medical research findings can be erroneous and what can go wrong in the process of designing and conducting research studies. It provides readers with essential analytical tools and scientific vocabulary. The challenges of medical research include data quality deficiencies; selection, confounding, measurement, and confirmation biases; inadequate sample sizes; sampling errors; effect modifiers; and causal interactions, among others. All of these can cause researchers to mistake mere associations for causal relationships and to reach conclusions that are invalid and cannot be replicated in subsequent studies.

Erroneous research findings can mislead legislators, regulators, and lawyers who use them for purposes of policy-making or litigation. Thus, understanding the pitfalls of big data analysis is important not only for scientists but also for anyone working with or reading about research studies, that is, for attorneys, health policy professionals, and the public at large.

I. INTRODUCTION

Big data analysis can be a valuable addition to the research toolkit and fill many knowledge gaps.¹ It is unclear, however, whether the proliferation of research endeavors will consistently advance scientific truths.

Big data research is relevant to health law and policy because it can be used in litigation by both plaintiffs and defendants.² Based on studies that reveal

* Edgar A. Hahn Professor of Law and Professor of Bioethics, Co-Director of Law-Medicine Center, Case Western Reserve University School of Law; B.A., Wellesley College; J.D., Harvard Law School; LL.M. in Health Law, University of Houston; S.J.D. in Health Law, Case Western Reserve University. This article is based in part on Sharona Hoffman & Andy Podgurski, *The Use and Misuse of Biomedical Data: Is Bigger Really Better?*, 39 AM. J. L. & MED. 497 (2013) and SHARONA HOFFMAN, *Medical Big Data Quality and Analysis Concerns*, in ELECTRONIC HEALTH RECORDS AND MEDICAL BIG DATA: LAW AND POLICY (Cambridge U. Press 2016).

1. Sharona Hoffman & Andy Podgurski, *The Use and Misuse of Biomedical Data: Is Bigger Really Better?*, 39 AM. J. L. & MED. 497, 499 (2013).

2. DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: STANDARDS, STATISTICS, AND RESEARCH METHODS 338-42 (student ed., 2008); Bernard Marr, *How Big Data Is Disrupting Law Firms and the Legal Profession*, FORBES (Jan. 20, 2016), <https://www.forbes.com/sites/bernardmarr/2016/01/20/how-big-data-is-disrupting-law-firms-and-the-legal-profession/#328c59227c23> [<https://perma.cc/4MXA-N45S>] (stating that lawyers “are starting to

statistical associations between illnesses and exposure to certain substances, plaintiffs may file tort cases against product manufacturers. For their part, defendants may use different studies to undermine plaintiffs' claims and argue that something other than their products caused the conditions in question.³

In addition, responsible legislators and policy-makers often rely on research findings for purposes of formulating laws, regulations, and policies. When the studies they review are wrong, government authorities can impose needless burdens on the public or, worse yet, cause harm.⁴

Two notorious examples illustrate the potential influence of erroneous research outcomes. A 1998 study published in the prestigious journal, *Lancet*, suggested that the measles, mumps, rubella (MMR) vaccination caused autism.⁵ While the study was later retracted,⁶ the belief that vaccinations can lead to autism gained traction⁷ and still needs to be explicitly rebutted on the Centers for Disease Control and Prevention's website.⁸ It is not inconceivable that such a study could have convinced state legislatures to relax vaccination requirements before it was disavowed.

In 2009, the *Journal of Psychiatric Research* published a study that claimed to have discovered a causal link between abortion and psychiatric disorders.⁹ The researchers reviewed "national data sets with reproductive history and mental health variables" to formulate their findings.¹⁰ Abortion foes widely cited the study,¹¹ and several states in fact enacted laws requiring women who want

think about how this technology could be applied to the fundamental research and case preparation which is the core of their job.").

3. FAIGMAN ET AL., *supra* note 2, at 339-40 (explaining that epidemiological evidence has already played an important role in many mass tort cases); Steve C. Gold, *The More We Know, the Less Intelligent We Are? – How Genomic Information Should, and Should Not, Change Toxic Tort Causation Doctrine*, 34 HARV. ENVTL. L. REV. 369, 412-17 (2010) (discussing genes and other toxins as alternate causes of plaintiffs' injuries).

4. Hoffman & Podgurski, *supra* note 1, at 500-01.

5. Andrew J. Wakefield et al., *Ileal-Lymphoid-Nodular Hyperplasia, Non-Specific Colitis, and Pervasive Developmental Disorder in Children*, 351 LANCET 637, 641 (1998).

6. Simon H. Murch et al., *Retraction of an Interpretation*, 363 LANCET 750, 750 (2004).

7. Jack Healy & Michael Paulson, *Vaccine Critics Turn Defensive over Measles*, N.Y. TIMES (Jan. 30, 2015), <https://www.nytimes.com/2015/01/31/us/vaccine-critics-turn-defensive-over-measles.html?mcubz=0> [<https://perma.cc/ZJ32-DWC7>].

8. *Measles, Mumps, and Rubella (MMR) Vaccine Safety Studies*, CDC, <http://www.cdc.gov/vaccinesafety/Vaccines/MMR/MMR.html> [<https://perma.cc/FT5D-HJ54>] (last updated Aug. 28, 2015).

9. Priscilla K. Coleman et al., *Induced Abortion and Anxiety, Mood, and Substance Abuse Disorders: Isolating the Effects of Abortion in the National Comorbidity Survey*, 43 J. PSYCHOL. RES. 770, 773 (2009).

10. *Id.*

11. Sharon Begley, *Journal Disavows Study Touted by U.S. Abortion Foes*, REUTERS (Mar. 7, 2012, 3:15 PM), <http://www.reuters.com/article/2012/03/07/us-usa-abortion-psychiatry-idUSTRE8261UD20120307> [<https://perma.cc/T2WQ-U2C8>] (stating that the study had been

abortions to receive counseling that warns them of potential long-term mental health problems.¹² The study was discredited in 2012 when scientists found that its design was deeply flawed.¹³ The original researchers did not compare women with unplanned pregnancies who had abortions to those who did not and failed to focus only on mental health problems that manifested *after* abortions.¹⁴ Thus, legislative initiatives were based on invalid findings.

There are many other examples of research findings that have gained extensive media coverage, but all too frequently, they are later questioned or contradicted.¹⁵ These reversals can be perplexing and frustrating.¹⁶ A 2010 article in *The Atlantic* bemoaned the phenomenon of conflicting, highly publicized research outcomes:

And they sometimes make headlines, as when in recent years large studies or growing consensuses of researchers concluded that mammograms, colonoscopies, and PSA tests are far less useful cancer-detection tools than we had been told; or when widely prescribed antidepressants such as Prozac, Zoloft, and Paxil were revealed to be no more effective than a placebo for most cases of depression; or when we learned that staying out of the sun entirely can actually increase cancer risks; or when we were told that the advice to drink lots of water during intense exercise was potentially fatal; or when, last April, we were informed that taking fish oil, exercising, and doing puzzles doesn't really help fend off Alzheimer's disease, as long claimed. Peer-reviewed studies have come to opposite conclusions on whether using cell phones can cause brain cancer, whether sleeping more than eight hours a night is healthful or dangerous, whether taking aspirin every day is more likely to save your life or cut it short, and whether routine angioplasty works better than pills to unclog heart arteries.¹⁷

This Article explores why medical research findings can be erroneous and

“widely cited by legislators and advocates to argue that abortion raises a woman's risk of mental illness and to push for laws requiring providers” to inform women of this danger).

12. *Counseling and Waiting Periods for Abortion*, GUTTMACHER INST.: STATE POLICIES IN BRIEF (2018), http://www.guttmacher.org/statecenter/spibs/spib_MWPA.pdf [<https://perma.cc/49By-RL9A>].

13. Ronald C. Kessler & Alan F. Schatzberg, *Commentary on Abortion Studies of Steinberg and Finer (Social Science & Medicine 2011; 72:72–82) and Coleman (Journal of Psychiatric Research 2009;43:770–6 & Journal of Psychiatric Research 2011;45:1133–4)*, 46 J. PSYCHOL. RES. 410, 410–411 (2012).

14. *Id.* at 410.

15. David H. Freedman, *Lies, Damned Lies, and Medical Science*, THE ATLANTIC (Nov. 2010), <https://www.theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/> [<https://perma.cc/K9LG-YGNE>].

16. Rebekah H. Nagler, *Adverse Outcomes Associated with Media Exposure to Contradictory Nutrition Messages*, 19 J. HEALTH COMM. 24, 24 (2014).

17. Freedman, *supra* note 15.

what can go wrong in the process of designing and conducting research studies. It provides readers with essential analytical tools and scientific vocabulary. Understanding the pitfalls of big data analysis is important not only for scientists but also for anyone working with or reading about research studies, that is, for attorneys, health policy professionals, and the public at large.

Big data analysis can face many challenges. First, as discussed in Part II, the data that researchers review can be fraught with errors, incomplete, fragmented, or otherwise flawed.¹⁸ Second, Part III analyzes a variety of biases that can distort research findings, including selection bias, confounding bias, measurement bias, and confirmation bias.¹⁹ Part IV explores additional research obstacles.²⁰ For example, sample sizes can be too small to produce reliable research results, sampling errors may exist, treatment effects can vary between individuals because of effect modifiers, and researchers might fail to discover causal interactions among multiple causal factors.²¹ As explained in Part V, all of these problems can cause researchers to mistake mere associations for causation and to reach outcomes that cannot be replicated and confirmed in subsequent studies. Thus, research errors and debunked studies are common phenomena. Finally, Part VI concludes.

II. DATA QUALITY

Medical big data analysis often draws upon electronic health records (EHRs). A key hurdle to conducting record-based research is that existing EHRs often contain errors, are incomplete, or suffer from other shortcomings. The information in EHRs is initially collected for medical treatment and billing purposes, so it may be ill-suited for research.²² Moreover, data are often contaminated by a myriad of inaccuracies that can compromise research outcomes. Though improved data-capture technology may remedy many shortcomings in the future,²³ these deficiencies should be a significant concern to contemporary researchers. This section discusses a number of potential data quality problems.

A. Data Can be Entered Incorrectly

Clinicians inputting EHR data can mistype words and numbers or copy and paste information from a prior date into a current visit's entry in order to save time without editing and updating it appropriately. They can also select the wrong EHR codes and menu items or be forced to choose among diagnosis or treatment codes that are not specific or detailed enough to describe the patient's condition

18. *See infra* Part II.

19. *See infra* Part III.

20. *See infra* Part IV.

21. *See infra* Part IV.

22. Hoffman & Podgurski, *supra* note 1, at 518-20.

23. *Id.* at 527-30.

accurately.²⁴

Some research suggests that error rates in EHRs are considerably higher than they were in paper records.²⁵ In one study, researchers examined progress notes made between August 2011 and July 2013 at a Michigan hospital that implemented an EHR system in 2012.²⁶ They found an error rate of 24.4 percent in EHRs compared to a 4.4 percent error rate in paper records.²⁷ It is noteworthy, however, that the study was conducted when the EHR system was first adopted and staff had to learn the new technology.²⁸ Moreover, the researchers found that medical residents, who are presumably more tech-savvy than the prior generation of physicians, had far fewer errors and omissions than their older colleagues, so that error rates might decline appreciably in the future.²⁹ It is also significant that EHRs have largely eliminated the serious problem of illegible handwriting in medical records, unless their contents are transcribed from handwritten notes.³⁰

Data inaccuracies can distort research outcomes. Even a small error rate, as low as one to five percent, can significantly impact estimates of mortality and adverse events.³¹ As incidents of hacking and cybercrimes proliferate, data analysts also cannot ignore the possibility that hackers will access EHR medical databases and purposefully alter and taint them.³²

24. Sharona Hoffman & Andy Podgurski, *E-Health Hazards: Provider Liability and Electronic Health Record Systems*, 24 BERKELEY TECH. L. J. 1523, 1544-45 (2009) (discussing input errors); Taxiarchis Botsis et al., *Secondary Use of EHR: Data Quality Issues and Informatics Opportunities*, AMIA SUMMITS TRANSLATIONAL SCI. PROC. 1, 3 (2010); Jessica S. Ancker et al., *Root Causes Underlying Challenges to Secondary Use of Data*, AMIA ANN. SYMP. PROC. ARCHIVE 57, 61 (2011); Hoffman & Podgurski, *supra* note 1 at 518-20 (discussing coding); Fouzia F. Ozair, *Ethical Issues in Electronic Health Records: A General Overview*, 6 PERSP. CLINICAL RES. 73, 75-76 (Apr-Jun 2015).

25. Jack McCarthy, *Doctors Make More Note-Taking Mistakes with EHRs than Paper Records*, *JAMIA Study Finds*, HEALTHCARE IT NEWS (July 8, 2016, 7:58 AM), <http://www.healthcareitnews.com/news/doctors-make-more-note-taking-mistakes-ehrs-paper-records-jamia-study-finds> [<https://perma.cc/W8WG-55KU>].

26. *Id.*

27. *Id.*

28. *Id.*

29. *Id.* (noting that the error rates were 5.3% versus 17.3%, and the omission rates were 16.8% versus 33.9%).

30. WIN PHILLIPS & YANG GONG, DEVELOPING A NOMENCLATURE FOR EMR ERRORS: PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON HUMAN-COMPUTER INTERACTION, PART IV: INTERACTING IN VARIOUS APPLICATION DOMAINS 587, 589, 591 (Julie A. Jacko ed., 2009).

31. George Hripcsak et al., *Bias Associated with Mining Electronic Health Records*, 6 J. BIOMED. DISCOVERY & COLLABORATION 48, 52 (2011); Steve Gallivan & Christina Pagel, *Modelling of Errors in Databases*, 11 HEALTH CARE MGMT. SCI. 35, 39 (2008); Christina Pagel & Steve Gallivan, *Exploring Potential Consequences on Mortality Estimates of Errors in Clinical Databases*, 20 IMA J. MGMT. MATHEMATICS 385, 391 (2009).

32. Sander Greenland, *Multiple Bias Modelling for Analysis of Observational Data*, 168 J.

B. Data Can Be Incomplete or Fragmented

If EHR data are incomplete or fragmented among several different records, data analysis will be unreliable. Researchers may often find that EHRs do not include all of the information they need for their studies.³³ Clinicians collect data for treatment and billing purposes and thus they generally do not have research studies in mind.³⁴ For example, a review of EHR data from the New York-Presbyterian Hospital clinical data warehouse revealed that when patients died of pneumonia in the emergency room, clinicians “spent little time documenting symptoms so that in the electronic health record, the patient appeared to be healthy other than the death.”³⁵ Thus, researchers may find these records to be relatively unilluminating and a poor fit for research.

Data concerning treatment outcomes are particularly likely to be missing because doctors and hospitals often do not follow up with patients.³⁶ For example, a patient who visits an emergency room for a few hours and is discharged may choose not to seek further care or may later go to a doctor with a different EHR system so that the hospital record provides no evidence of how the individual ultimately fared. The absence of information could mean that the emergency room treatment cured the patient or, by contrast, that she did not improve or even deteriorated and sought care elsewhere.

Data fragmentation often exists because different facilities have EHR systems that are not interoperable.³⁷ Interoperability is “the ability of different information technology systems and software applications to communicate, to exchange data accurately, effectively, and consistently, and to use the information that has been exchanged.”³⁸ Patients who are treated at multiple medical centers may have their records divided among different EHR systems so that their record cannot easily be put together into an integrated whole.³⁹ EHR fragmentation is a serious

ROYAL STAT. SOC’Y: SERIES A 267, 267-68 (2005); Craig H. Mallinckrodt et al., *Assessing and Interpreting Treatment Effects in Longitudinal Clinical Trials with Missing Data*, 53 BIOL. PSYCH. 754, 755 (2003).

33. Craig Newgard et al., *Electronic Versus Manual Data Processing: Evaluating the Use of Electronic Health Records in Out-of-Hospital Clinical Research*, 19 ACAD. EMERGENCY MED. 217, 224 (2012).

34. M. Alan Brookhart et al., *Confounding Control in Healthcare Database Research Challenges and Potential Approaches*, 48 (Suppl 6) MED. CARE S114, S115 (2010).

35. Hripcsak et al., *supra* note 31, at 50.

36. Newgard et al., *supra* note 33, at 225.

37. Julia Adler-Milstein, *Moving Past the EHR Interoperability Blame Game*, NEJM CATALYST (July 18, 2017), <http://catalyst.nejm.org/ehr-interoperability-blame-game/> [https://perma.cc/LGF7-D9YW].

38. Kevin Heubusch, *Interoperability: What it Means, Why it Matters*, 77 AHIMA 26 (Jan. 2006), <http://library.ahima.org/doc?oid=60942#.WZXnGVGGM2w> [https://perma.cc/9C8G-GYF4].

39. Botsis et al., *supra* note 24, at 4 (stating that the EHR system that was mined for purposes

obstacle for research because analysts looking at records from a particular institution or database see only pieces of patients' records, and key medical information may be missing. Furthermore, researchers who examine de-identified records from multiple facilities may mistake the records of a single patient for records of several different individuals.

C. Data Can Be Difficult to Extract and Lack Harmonization

As other scholars have noted, “[t]he realities of doing research across different institutions using the EHR, even when working with sites that have an EHR from the same vendor, are daunting.”⁴⁰ Different entities may use different versions of EHRs that collect data in different ways. Entities may customize their EHRs to change menu items or terminology, and health care providers have different data documentation policies and habits.⁴¹ These problems and others can lead to a lack of semantic harmonization that adversely affects the usability of records for research purposes. Semantic harmonization is “the process of collating ... data into a singular consistent logical view” so that data are not subject to different interpretations.⁴²

For example, clinicians may be able to choose among a variety of fields in which to document a particular detail, so that researchers expecting to find it in one field may be confused and miss it altogether. Likewise, different medical offices use terms, phrases, and abbreviations differently. To illustrate, the abbreviation “MS” can mean “mitral stenosis,” “multiple sclerosis,” “morphine sulfate,” or “magnesium sulfate.”⁴³ Therefore, a reader looking at records from different sources might be unable to understand what “MS” means in a particular instance.⁴⁴

Another challenge is free text narrative in EHRs. EHR systems allow users to enter both coded information in structured fields and natural-language free text

of the study did not contain records of patients who were transferred to dedicated cancer centers because of the severity of their disease or who had initially been treated elsewhere).

40. Kathryn H. Bowles et al., *Conducting Research Using the Electronic Health Record Across Multi-Hospital Systems: Semantic Harmonization Implications for Administrators*, 43 J. NURSING ADMIN. 355, 356 (2013).

41. *Id.*

42. James A. Cunningham et al., *Nine Principles of Semantic Harmonization*, 2016 AMIA ANN. SYMP. PROC. ARCHIVE 451, 451 (2016).

43. Christopher G. Chute, *Medical Concept Representation*, in MEDICAL INFORMATICS: KNOWLEDGE MANAGEMENT AND DATA MINING IN BIOMEDICINE 170 tbl.6-1 (Hsinchun Chen et al. eds., 2005).

44. See The Clinical Cancer Genome Task Team of the Global Alliance for Genomics and Health, *Sharing Clinical and Genomic Data on Cancer—The Need for Global Solutions*, 376 NEW. ENG. J. MED. 2006, 2006 (2017) (stating that “[i]ncompatible data formats and a shortage of interoperable data-harmonizing informatics tools also compromise researchers’ ability to mine multiple data sets” as does “the absence of a single standardized cancer ontology (a machine-readable set of defined descriptors of clinical manifestations)”).

notes about patients.⁴⁵ Important information may appear in free text rather than in the structured fields, and information written in narrative form is much more difficult to extract accurately from EHRs for research purposes.⁴⁶ As an example, a patient's worsening asthma may be linked to smoking, but researchers may be unaware of this connection if exacerbated asthma is coded, but smoking history is described only in free-text clinical notes.⁴⁷ Similarly, clinicians may record family histories or details about adverse reactions to drugs in narrative form.⁴⁸ Experts may employ natural-language processing tools to extract data from free-text narrative, but these techniques are still developing and are often imperfect.⁴⁹

Diagnoses, measurements or medical histories that are not standardized or are inaccessible because they are not available in structured form, can significantly hinder research efforts.⁵⁰ Similarly, medical vocabulary that is not harmonized may cause analysts to misconstrue or be unable to understand medical records.

D. Data Can Be Distorted by Software Glitches

Defects in software that is used in EHR systems or is used for data analysis may also generate troubling errors.⁵¹ This is particularly likely when scientists or their assistants develop complex software without consulting skilled software developers.⁵² Inexperienced programmers are likely to create incorrect software

45. S. Trent Rosenbloom et al., *Data from Clinical Notes: A Perspective on the Tension between Structure and Flexible Documentation*, 18 J. AM. MED. INFORMATICS ASS'N 181, 181-82 (2011).

46. The Clinical Cancer Genome Task Team of the Global Alliance for Genomics and Health, *supra* note 44, at 2006 (noting the "lack of effective methods for extracting data from electronic health records"); Hilary Townsend, *Natural Language Processing and Clinical Outcomes: The Promise and Progress of NLP for Improved Care*, 84 AHIMA 44, 44-45 (Mar. 2013), http://library.ahima.org/doc?oid=106198#.Wo21_qjwa70 [<https://perma.cc/B47V-RVWT>].

47. Naren Ramakrishnan et al., *Mining Electronic Health Records*, 43 COMPUTER 95, 97 (2010).

48. Isaac S. Kohane, *Using Electronic Health Records to Drive Discovery in Disease Genomics*, 12 NATURE REV. GEN. 417, 420 (2011).

49. Mike Miliard, *EHR Natural Language Processing Isn't Perfect, but It's Really Useful*, HEALTHCARE IT NEWS (May 18, 2017), <http://www.healthcareitnews.com/news/ehr-natural-language-processing-isnt-perfect-its-really-useful-tool> [<http://perma.cc/Z46T-VYRS>].

50. Andrea L. Benin et al., *How Good Are the Data? Feasible Approach to Validation of Metrics of Quality Derived from an Outpatient Electronic Health Record*, 26 AM. J. MED. QUALITY 441, 441 (2011).

51. Mi Ok Kim et al., *Problems with Health Information Technology and Their Effects on Care Delivery and Patient Outcomes: A Systematic Review*, 24 J. AM. MED. INFORMATICS ASS'N, 246, 250-51 (2017).

52. Rebecca Sanders & Diane Kelly, *Dealing with Risk in Scientific Software Development*, 25 IEEE SOFTWARE 21, 27 (2008); Diane F. Kelly, *A Software Chasm: Software Engineering and Scientific Computing*, 24 IEEE SOFTWARE 118, 118-20 (2007); Les Hatton, *The Chimera of Software Quality*, 40 COMPUTER 104, 104 (2007).

and then test it inadequately.⁵³ However, commercially developed medical research software is not immune to software glitches.⁵⁴ Ideally, scientists should closely collaborate with software experts to develop and thoroughly validate software for medical research.

III. THE PROBLEM OF BIASES

Even without data blemishes and software defects, analyzing data and drawing correct causal inferences can be extremely challenging. This section analyzes problems of bias affecting big data studies, in particular, selection bias, confounding bias, measurement bias, and confirmation bias. Anyone conducting research or interpreting its results must have a deep understanding of these concepts.

A. Selection Bias

Selection bias relates to the way in which data subjects are selected for study and generally occurs when selection is not randomized.⁵⁵ Selection bias may arise if individuals have the opportunity to opt out of being included in a database or clinical study.⁵⁶ The subset of individuals that remain may not be representative of the patient population of interest because people of a particular sex, ancestry, or socioeconomic group opted out disproportionately.⁵⁷ Selection bias can skew assessments of measures such as disease prevalence or exposure risk because the study's conclusions differ systematically from the true values of these measures for the target population. That is, the researchers' findings will not be generalizable from the study's particular subjects to the larger population about which analysts wish to draw conclusions.⁵⁸

53. Kelly, *A Software Chasm*, *supra* note 52, at 118.

54. Sanders & Kelly, *supra* note 52, at 25; Nicole K. Henderson-MacLennan et al., *Pathway Analysis Software: Annotation Errors and Solutions*, 101 *MOLECULAR GENETICS & METABOLISM* 134, 137-38 (2010).

55. *What researchers mean by . . . selection bias*, INST. FOR WORK & HEALTH, <https://www.iwh.on.ca/wrmb/selection-bias> [<https://perma.cc/L355-V84M>] (last visited Aug. 18, 2017).

56. KENNETH J. ROTHMAN ET AL., *MODERN EPIDEMIOLOGY* 136 (2008); DAVID L. FAIGMAN ET AL., *MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY* §5:16 (2011).

57. SHARYL J. NASS ET AL., *BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH* 209 (2009); Franklin G. Miller, *Research on Medical Records Without Informed Consent*, 36 *J. L. MED. & ETHICS* 560, 560 (2008).

58. HERBERT I. WEISBERG, *BIAS AND CAUSATION: MODELS AND JUDGMENT FOR VALID COMPARISONS* 93-94 (2010).

B. Confounding Bias

Confounding bias can be a significant problem in causal effect studies.⁵⁹ Causal effect studies typically seek to measure a particular treatment's average beneficial effect on patients or a particular exposure's average harmful effect on individuals. Confounding occurs when there is a common cause of the treatment/exposure variable *and* the outcome variable.⁶⁰

To illustrate confounding, I offer the following hypothetical. Assume a physician chooses among different treatment options based on the severity or duration of a patient's disease, and these factors also influence the outcome of treatment.⁶¹ Thus, patients at a later stage of illness receive Treatment A and those at an earlier stage receive Treatment B. At the same time, sicker patients, as a general matter, will not respond as well as healthier patients to treatments because they are frailer or their conditions are more complicated. Unless researchers appropriately adjust for the disease stage factor, called a "confounding variable" or "confounder," during statistical data analysis, the confounder may produce a false association between the treatment variable and the outcome variable, which will distort findings regarding the causal effects of treatments. In other words, researchers' conclusions regarding how effective the two medications are may be wrong because of the confounding variable: the illness severity of patients receiving the different therapies. Treatment A may appear to be inferior to treatment B not because it actually is less effective but because the patients receiving treatment A are in a late stage of the disease and would not thrive no matter what therapy was administered.

Socioeconomic factors and patient lifestyle choices may also constitute confounders. Financially disadvantaged (and underinsured) individuals may select less expensive treatments, even if those are not the best choices for them, because those are the only affordable options. Poverty may also separately lead to ill health, *e.g.* because of poor nutrition, stress, and lack of leisure time for exercise. On the flip side, the benefits of preventive care, such as cholesterol-lowering medication, might appear to be greater than they truly are because health-conscious individuals who seek the intervention also take care of themselves with exercise, low-fat diets, and other health-promoting behaviors. Thus, these patients' impressive outcomes are associated with a multitude of factors rather than solely with the preventive measure.⁶²

Analysts must strive to ascertain, accurately measure, and adjust for all potential confounding variables in order to reduce or eliminate confounding

59. Sander Greenland, *Quantifying Biases in Causal Models: Classical Confounding vs. Collider-Stratification Bias*, 14 EPIDEMIOLOGY 300, 306 (2003).

60. Miguel A. Hernan et al., *A Structural Approach to Selection Bias*, 15 EPIDEMIOLOGY 615, 618 (2004).

61. Bruce M. Psaty & David S. Siscovick, *Minimizing Bias Due to Confounding by Indication in Comparative Effectiveness Research*, 304 J. AM. MED. ASS'N 897, 897 (2010).

62. Brookhart et al., *supra* note 34, at S115.

bias.⁶³ This is no simple task in medical studies because it is often very difficult to discover which variables are potential confounders. Medical care often depends on a patchwork of factors relating to the patients' health care facilities, health insurance, and clinicians, and to the character and circumstances of the patients themselves, so that various influences remain hidden or obscured.⁶⁴

In clinical studies,⁶⁵ researchers ideally randomize treatment assignment to subjects in order to prevent confounding.⁶⁶ Thus, in my hypothetical, both treatment A and treatment B would be given to patients all along the disease severity spectrum to determine their efficacy regardless of disease stage. By contrast, in a record-based study (also called an observational study), analysts cannot control treatment assignment, because the treatments have already been given and recorded in patients' medical files. Researchers must, therefore, strive to identify and obtain the values of confounding variables and adjust for them during data analysis.⁶⁷

C. Measurement Bias

Study results can be further compromised by measurement biases arising from errors in measurement and data collection. Measuring devices might not be calibrated properly or might not be sensitive enough to detect small differences in relevant variables.⁶⁸ Biological samples might be stored for different lengths of time or in different conditions, which can cause measurement bias when they are analyzed.⁶⁹ To the extent that research is based on patients' own accounts and memories, outcomes might be distorted because subjects' responses are influenced by the questioners' skills, patience, and apparent sympathy or by how important and relevant patients perceive the study's topic to be.⁷⁰ In addition, patients' memories may be impaired, or they may lie in response to questions that they find to be uncomfortable or embarrassing.

For example, a study of patients' use of opioid pain medication may have to

63. ROTHMAN ET AL. *supra* note 56, at 58.

64. Brookhart et al., *supra* note 34, at S114.

65. In clinical studies, investigators conduct experiments using human subjects. By contrast, observational studies involve review of existing records rather than controlled experiments. Sharona Hoffman & Andy Podgurski, *Balancing Privacy, Autonomy, and Scientific Needs in Electronic Health Records Research*, 65 SMU L. REV. 85, 98-102 (2012) (contrasting clinical trials and observational studies) <https://ssrn.com/abstract=1923187> [<https://perma.cc/9X4W-766Q>].

66. Jaelyn L.F. Bosco et al., *A Most Stubborn Bias: No Adjustment Method Fully Resolves Confounding by Indication in Observational Studies*, 63 J. CLINICAL EPIDEMIOLOGY 64, 64 (2010) (explaining that "confounding is best controlled by a randomized design").

67. *Id.* at 64-65.

68. Gael P. Hammer et al., *Avoiding Bias in Observational Studies*, 106 DEUTSCHES ÄRZTEBLATT INT'L 664, 665 (2009), <https://www.aerzteblatt.de/int/archive/article?id=66288&src=search> [<https://perma.cc/6RUQ-BXAN>].

69. *Id.* at 665.

70. *Id.*

rely on patients' own reports concerning how many pills they took each day. This is because prescription records will only reveal to researchers how many prescriptions were filled, not how many pills were ingested and how often. Yet, patients who are in severe pain may not record or recall the details of their medication intake. In addition, in light of extensive publicity regarding the opioid crisis,⁷¹ individuals may tend to under-estimate or under-report the amount of medication they took because they do not want to be considered to be opioid addicts.

In fact, the treatment or outcome itself can affect measurement errors.⁷² One of the side effects of opioid use is cognitive impairment.⁷³ Thus, in the above example, patients may misremember how much medication they took in part because the opioids themselves caused them to suffer cognitive deficits.

D. Confirmation Bias

Confirmation bias can be defined as “an agent’s tendency to seek, interpret and use evidence in a manner biased toward confirming her existing beliefs or hypotheses.”⁷⁴ Medical researchers are under significant pressure to obtain grant funding and publications.⁷⁵ Academics thus undertake projects with particular results in mind, fervently hoping for dramatic findings and headline-grabbing outcomes.⁷⁶ These goals, which some consider to be conflicts of interest, can lead analysts to make intentional or unconscious errors such as manipulating data, distorting results, or making stronger claims than the true findings justify.⁷⁷

Confirmation bias is also at play when researchers return time and again to a database, testing and retesting data in order to find publishable results. If initial results are not satisfactory, they look for other associations until one emerges.⁷⁸

71. See, e.g., Maya Salam, *The Opioid Epidemic: A Crisis Years in the Making*, N.Y. TIMES, (Oct. 26, 2017), https://www.nytimes.com/2017/10/26/us/opioid-crisis-public-health-emergency.html?_r=0 [<https://perma.cc/2388-AFQ8>].

72. ROTHMAN ET AL., *supra* note 56, at 137-38.

73. DONALD TEATER, NAT'L SAFETY COUNCIL, THE PSYCHOLOGICAL AND PHYSICAL SIDE EFFECTS OF PAIN MEDICATIONS 3 (2017), <http://www.nsc.org/RxDrugOverdoseDocuments/900006497-ADV-Rx-Side-Effects-WhitePaper.pdf> [<https://perma.cc/5Q79-ZY3Y>].

74. Gary Charness & Chetan Dave, *Confirmation Bias with Motivated Beliefs*, 104 GAMES AND ECON. BEHAV. 1, 1 (2017).

75. Freedman, *supra* note 15.

76. *Id.* See also Sridharan Kannan & Sivaramakrishnan Gowri, *Contradicting/Negative Results in Clinical Research: Why (Do We Get These)? Why Not (Get These Published)? Where (to Publish)?*, 5 PERSP. CLINICAL RES. 151, 151 (2014).

77. Freedman, *supra* note 15; Alok Jha, *False Positives: Fraud and Misconduct are Threatening Scientific Research*, THE GUARDIAN (Sept. 13, 2012), <https://www.theguardian.com/science/2012/sep/13/scientific-research-fraud-bad-practice> [<https://perma.cc/7ZWV-Z4QB>].

78. Freedman, *supra* note 15 (stating that researchers “run everything through the mill, one at a time, and they start finding associations, and eventually conclude that vitamin X lowers the risk of cancer Y, or this food helps with the risk of that disease.”).

Experts caution researchers who are conducting subgroup analysis that they should develop a small number of hypotheses a priori, before beginning their study.⁷⁹ By contrast, testing a large number of subgroup hypotheses, especially those developed post hoc, is likely to reveal spurious subgroup differences.⁸⁰ For example, in one instance, a randomized controlled trial (“RCT”) of a treatment for septic⁸¹ patients failed to show that the therapy had a statistically significant benefit.⁸² A subgroup analysis, however, seemed to demonstrate that the treatment was effective for individuals with a particular type of infection.⁸³ Yet, a later, large RCT failed to reproduce the subgroup result and proved the positive conclusion to be false.⁸⁴ The root of the problem was likely that the researchers’ initial, erroneous subgroup analysis was one of fifteen subgroup hypotheses that they tested. One expert explains the issue as follows: “the odds are that in any large database ... there will be a few apparent connections that are in fact merely flukes, not real health effects—it’s a bit like combing through long, random strings of letters and claiming there’s an important message in any words that happen to turn up.”⁸⁵

IV. OTHER DATA ANALYSIS PITFALLS

This section focuses on several additional potential research errors. These consist of using a sample size that is too small, ignoring effect modifiers, and disregarding causal interactions.

A. Sample Size

Researchers must ensure that their sample size is large enough to yield credible results.⁸⁶ Sample size affects statistical power, that is, “the probability that a statistical test will indicate a significant difference when there truly is one.”⁸⁷ If the sample size is too small, it is more likely that any observed

79. Xin Sun et al., *How to Use a Subgroup Analysis: Users’ Guides to the Medical Literature*, 311 JAMA 405, 408 (2014) (explaining that subgroup analysis is designed to identify groups of patients that “respond differently to treatment than other groups.”).

80. *Id.* at 408.

81. *Definition of Sepsis*, SEPSIS ALL. (last visited August 24, 2017), <http://www.sepsis.org/sepsis/definition/> [<https://perma.cc/HT9G-P2ZT>] (defining sepsis as “the body’s overwhelming and life-threatening response to infection that can lead to tissue damage, organ failure, and death.”).

82. Xin Sun et al., *supra* note 79, at 408 (stating that the treatment was platelet-activating factor receptor antagonist).

83. *Id.* (stating that the subgroup was individuals with gram-negative bacterial infection).

84. *Id.*

85. Freedman, *supra* note 15.

86. David Jean Biau et al., *Statistics in Brief: The Importance of Sample Size in the Planning and Interpretation of Medical Research*, 466 CLINICAL ORTHOPEDICS & RELATED RES. 2282, 2283 (2008).

87. John Eng, *Sample Size Estimation: How Many Individuals Should Be Studied?*, 227 RADIOLOGY 309, 310 (2003).

differences among subgroups (or other findings) are false and attributable purely to chance.⁸⁸ For example, assume that a medication causes severe nausea and vomiting for one in ten patients. If researchers look at only ten medical records, it is very possible that they will encounter only patients who tolerated the drug well and will wrongly conclude that the drug does not cause adverse side effects for anyone who takes it.

Determining the correct sample size for a particular study is a complicated matter, and it behooves analysts to consult statisticians.⁸⁹ Sample size can be particularly challenging for studies targeting rare diseases or rare genetic variations.⁹⁰ However, anyone relying on research outcomes for medical or health policy purposes should know to scrutinize sample size and question studies that include only a small number of subjects.

B. Sampling Error

Any sample of data or subjects that researchers study may behave differently or have different characteristics from the larger population from which the sample is drawn because the particular collection of a study participants is just one of many potential groups that researchers could have formed from the population of interest.⁹¹ This can lead to sampling error.⁹² Sampling error is different from selection bias because it does not involve systematic bias, such as younger individuals disproportionately opting out of a study; it simply occurs because of random variability of subjects.⁹³

Small samples are particularly vulnerable to sampling error.⁹⁴ For example, assume that investigators wish to determine whether Cleveland, Ohio has a higher cancer rate than other locations in the United States. Assume also that they review a sample of one hundred medical records and determine that ten percent of

88. Biau et al., *supra* note 86, at 2282-83; Katherine S. Button et al., *Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience*, 14 NATURE REVIEWS NEUROSCIENCE 365, 365 (2013).

89. See Eng *supra* note 87, at 309-12; Mohamad Amin Pourhoseingholi et al., *Sample Size Calculation in Medical Studies*, 6 GASTROENTEROLOGY & HEPATOLOGY FROM BED TO BENCH 14, 16 (2013).

90. J.H. van der Lee et al., *Efficient Ways Exist to Obtain the Optimal Sample Size in Clinical Trials in Rare Diseases*, 61 J. CLINICAL EPIDEMIOLOGY 324, 324 (2008); see also Seunggeun Lee et al., *Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies*, 91 AM. J. HUM. GENETICS 224, 224 (2012).

91. *Id.*; see MIGUEL A. HERNÁN & JAMES M. ROBINS, CAUSAL INFERENCE 119 (2017) (ebook).

92. Douglas G. Altman & J. Martin Bland, *Uncertainty and Sampling Error*, 349 BMJ 1 (2014), <http://www.bmj.com/content/bmj/349/bmj.g7064.full.pdf> [<https://perma.cc/MZ2N-9379>].

93. See HERNÁN & ROBINS, *supra* note 91, at 119.

94. Stephen Tyrer & Bob Heyman, *Sampling in Epidemiological Research: Issues, Hazards and Pitfalls*, 40 BJPSYCH BULLETIN 57, 57 (2016).

patients have cancer. They thus would conclude that Cleveland has a cancer rate that is double that of the national average which is approximately five percent.⁹⁵ The problem is that this very small sample may not tell the investigators anything about the true cancer rate in Cleveland. It may just happen by chance that an unusually large portion of the patients whose records were reviewed had cancer. Thus, it would be inappropriate to generalize the results of this very limited study to the Cleveland population at large.

C. Effect Modifiers

Research outcomes can be skewed by effect modifiers, which are characteristics of patients or of the study that change treatment effects.⁹⁶ Thus, “the magnitude of the association between an exposure and an outcome varies across strata of some other factor.”⁹⁷ To illustrate, the degree to which alcohol consumption affects blood pressure varies with age, gender, and smoking status, and therefore these three attributes are effect modifiers.⁹⁸ Effect modification merits study for its own sake. Moreover, investigators examining the relationship between alcohol and blood pressure will reach mistaken conclusions if they ignore the effect modifiers and do not address them in their analysis.

D. Causal Interactions

Causal interactions occur when two or more factors interact to produce an effect.⁹⁹ The causal factors do not have to operate at the same time, but they all contribute to the outcome at issue.¹⁰⁰ For example, in some cases, smoking and alcohol consumption interact to cause head and neck cancers.¹⁰¹ Researchers have

95. See *Cancer Stat Facts: Cancer of Any Site*, NAT'L CANCER INST., <https://seer.cancer.gov/statfacts/html/all.html> [<https://perma.cc/WET3-LANU>] (stating that in “2014, there were an estimated 14,738,719 people living with cancer of any site in the United States”). That same year the U.S. population was approximately 317,000,000. Robert Schlesinger, *The 2014 U.S. and World Populations*, U.S. NEWS & WORLD REP. (Dec. 31, 2013, 3:15 PM), <https://www.usnews.com/opinion/blogs/robert-schlesinger/2013/12/31/us-population-2014-317-million-and-71-billion-in-the-world> [<https://perma.cc/GQY3-Z9QE>].

96. Jeroen P. Jansen & Huseyin Naci, *Is Network Meta-Analysis as Valid as Standard Pairwise Meta-Analysis? It All Depends on the Distribution of Effect Modifiers*, 11 BMC MED. 159, 159 (2013); See Tyler J. VanderWeele, *On the Distinction between Interaction and Effect Modification*, 20 EPIDEMIOLOGY 863, 863 (2009).

97. David C. Bellinger, *Effect Modification in Epidemiologic Studies of Low-Level Neurotoxicant Exposures and Health Outcomes*, 22 NEUROTOXICOLOGY & TERATOLOGY 133, 133 (2000) (footnote omitted).

98. *Id.*

99. Kenneth J. Rothman & Sander Greenland, *Causation and Causal Inference in Epidemiology*, 95 AM. J. PUB. HEALTH S144, S145 (2005); Tyler J. VanderWeele & Mirjam J. Knol, *A Tutorial on Interaction*, 3 EPIDEMIOLOGY METHODS 33, 33 (2014).

100. Rothman & Greenland, *supra* note 99, at S145.

101. *Id.* at S146.

found that seventy-five percent of head and neck cancer cases can be attributed to smoking and sixty-seven percent to drinking alcohol among those who engage in both behaviors.¹⁰² While these figures may seem puzzling because they add up to more than one-hundred percent, they are accurate because smoking and alcohol consumption can interact together to cause cancer. In fact, cancer is generally caused by a multitude of factors, including diet, smoking, occupational exposures (e.g. asbestos) genetic factors, and others.¹⁰³ In addition, there can be synergy among the different causal factors such that their interaction amplifies the effect of the separate factors.¹⁰⁴ Although either smoking or asbestos exposure can cause lung cancer, the risk of lung cancer is much higher for individuals who both smoked and breathed asbestos.¹⁰⁵ Researchers focusing on one cause of an outcome must be aware of potential interactions and multi-causal triggers.

V. THE PREVALENCE OF RESEARCH MISTAKES

The pitfalls discussed thus far in the article can lead to invalid findings. How often do researchers reach wrong conclusions? According to some experts, more often than not. Dr. John Ioannidis, who has dedicated his career to studying the credibility of medical research, estimates that “80 percent of non-randomized studies (by far the most common type) turn out to be wrong, as do 25 percent of supposedly gold-standard randomized trials, and as much as 10 percent of the platinum-standard large randomized trials.”¹⁰⁶ This section probes the prevalence of research mistakes.

A. Association v. Causation

If researchers succumb to some of the pitfalls described in the Article, they may mistake mere *association* for actual *causation*.¹⁰⁷ They may identify associations between certain exposures and particular outcomes but wrongly conclude that there is an actual causal relationship between them.¹⁰⁸ To illustrate,

102. *Id.*; See also ROTHMAN ET AL., *supra* note 56, at 71.

103. Rothman & Greenland, *supra* note 99, at S146; VanderWeele & Knol, *supra* note 99, at 33.

104. VanderWeele & Knol, *supra* note 99, at 35.

105. *Id.*

106. Freedman, *supra* note 15; see also John P. A. Ioannidis, *Why Most Published Research Findings Are False*, 2 PLOS MED. 0696 (2005), <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124> [<https://perma.cc/4GDL-E6FP>].

107. See, e.g., Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 PROC. ROYAL SOC'Y MED. 295, 295–300 (1965); Arvid Sjölander, *The Language of Potential Outcomes*, in CAUSALITY: STATISTICAL PERSPECTIVES AND APPLICATIONS 6, 9 (Carlo Berzuini et al. eds., 2012).

108. See Stephen Choi et al., *The Power of Proxy Advisors: Myth or Reality?* 59 EMORY L.J. 869, 879–85 (2010) (discussing the difference between correlation and causation); Csaba P. Kovcsdy & Kamyar Kalantar-Zadeh, *Observational Studies vs. Randomized Controlled Trials: Avenues to Causal Inference in Nephrology*, 19 ADVANCED CHRONIC KIDNEY DISEASE 11, 11–12

data analysis may suggest that drinking coffee increases an individual's risk of lung cancer.¹⁰⁹ The association would likely disappear, however, when analysts adjust for the fact that, generally, coffee drinkers are more likely to smoke cigarettes than non-coffee consumers,¹¹⁰ and it is cigarette smoking that, in truth, increases lung cancer risk.¹¹¹ Researchers who confuse the association between coffee and lung cancer for a causal relationship would stumble because of confounding bias.¹¹² Scientists have found that a genetic variant influences both caffeine consumption and smoking, which in turn increases the risk of lung cancer, and thus the genetic factor is a confounder.¹¹³ Moreover, smoking itself may cause individuals to drink more coffee because nicotine increases the metabolism of caffeine, so that smokers lose the stimulating effects of caffeine more quickly.¹¹⁴ Smoking is thus a confounder that causes both increased coffee drinking and lung cancer. Without awareness of the genetic and smoking confounders, researchers could reach erroneous conclusions about the cause of lung cancer.

As a second example, selection bias can cause analysts to mistake association for causation. Suppose researchers are interested in the question of whether folic acid, taken during pregnancy, can prevent fetal heart defects. Assume too that the investigators include only babies that are born live and exclude from the study fetuses that are miscarried. Researchers might conclude that folic acid is linked to a reduction in the number of fetal heart defects. However, no causal link can be assumed because fetuses that were miscarried were not studied.¹¹⁵ Some of these fetuses, in fact, may not have survived because of heart defects. Researchers who posit that folic acid reduces fetal heart defects based on these results would have failed to establish an actual causal relationship between the treatment and outcome.

B. Inability to Replicate Study Findings

Research errors can make it impossible for researchers to replicate studies and confirm results. Although it may be far more exciting to make new discoveries,¹¹⁶

(2012).

109. Bellinger, *supra* note 97, at 133.

110. Jorien L. Treur et al., *Associations between Smoking and Caffeine Consumption in Two European Cohorts*, 111 *ADDICTION* 1059, 1059 (2016) (stating that “[t]here appears to be a positive association between smoking and caffeine consumption in the Netherlands and the United Kingdom”).

111. *Id.*

112. *See supra* Part III.B.

113. Treur et al. *supra* note 110, at 1066; Akshat Rathi, *There's a Reason Why Smokers Drink More Coffee*, *QUARTZ* (Mar. 7, 2017), <https://qz.com/926328/scientists-can-explain-why-smokers-drink-more-coffee-genetics-metabolism-and-blunted-taste-buds/> [<https://perma.cc/9DKM-8CFU>].

114. Treur et al. *supra* note 110, at 1066; Rathi, *supra* note 113.

115. HERNÁN & ROBINS, *supra* note 91, at 95.

116. Michael Price, *To Replicate or Not To Replicate?*, *SCI.* (Dec. 2, 2011, 10:00 AM),

the work of reproducing prior study results is essential to scientific integrity and the validity of research claims.¹¹⁷

Replicability is a very significant problem for both clinical and observational research. To replicate a clinical trial, a new group of investigators conducts the experiment again, and in the case of a record-based study, new analysts examine a different data set and try to obtain the same findings.¹¹⁸ But if the initial study outcome is erroneous because of data quality, biases, and/or the other shortcomings discussed above, it will not be reproducible. On the other hand, if replication studies are themselves tainted by any of these problems, they will fail even if the original results are legitimate.

In 2005, Dr. John Ioannidis published a study in *JAMA*, a top medical journal, focusing on forty-nine of the most highly regarded research findings published during the prior thirteen years.¹¹⁹ Of these, forty-five studies claimed to have discovered effective interventions, and thirty-four of these findings were retested in order to validate results. Astonishingly, upon retesting, fourteen of the claims (forty-one percent) were shown to be incorrect or exaggerated.¹²⁰

In a highly publicized 2015 article, the journal *Science* featured “a large-scale, collaborative effort to obtain an initial estimate of the reproducibility of psychological science.”¹²¹ Investigators attempted to reproduce the outcomes of one-hundred recent peer-reviewed psychology experiments, but were successful in only thirty-nine cases.¹²²

Commentators bemoan the “replication crisis,” and journalists have written articles with headlines such as “Is Science Broken?”¹²³ Others are less pessimistic

<http://www.sciencemag.org/careers/2011/12/replicate-or-not-replicate> [https://perma.cc/L5WT-PSGL]; Timothy M. Errington et al., *An Open Investigation of the Reproducibility of Cancer Biology Research*, ELIFE, (Dec. 10, 2014), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4270077/> [https://perma.cc/Y2DS-JVDW] (asserting that “careers are made by producing exciting new results at the frontiers of knowledge, not by verifying prior discoveries”).

117. Open Science Collaboration, *Estimating the Reproducibility of Psychological Science*, 349 SCI. 943, 943 (2015); Matthew C. Makel & Jonathan A. Plucker, *Facts Are More Important Than Novelty: Replication in the Education Sciences*, 43 EDUC. RESEARCHER 304, 304 (2014); Jens B. Asendorpf et al., *Recommendations for Increasing Replicability in Psychology*, 27 EUR. J. PERSONALITY 108, 108 (2013) (“Replicability of findings is at the heart of any empirical science.”).

118. Makel & Plucker, *supra* note 113, at 304 (emphasizing “the importance of third-party, direct replications in helping education research improve its ability to shape education policy and practice”); Arturo Casadevall & Ferric C. Fang, *Reproducible Science*, 78 INFECTION & IMMUNITY 4972, 4972-75 (2010).

119. John P. A. Ioannidis, *Contradicted and Initially Stronger Effects in Highly Cited Clinical Research*, 294 JAMA 218, 218 (2005) (“The size of the sample studied is a major determinant of the risk of reporting false-negative findings”).

120. *Id.*

121. Open Science Collaboration, *supra* note 117, at 943.

122. *Id.*

123. Daniel Engber, *Is Science Broken Or Is It Self Correcting?*, SLATE (Aug. 21, 2017, 6:00 AM), http://www.slate.com/articles/health_and_science/science/2017/08/

about replication failures and state that “[w]e should just be looking at an accumulating evidence paradigm, where we’re getting closer and closer to truth.”¹²⁴ Still others argue that replications might fail because they themselves are flawed, that is, the later studies are not conducted under the same conditions by people with the same level of expertise as the original investigators.¹²⁵

Nevertheless, repeated replication failures have distressed the scientific community and all who care about the medical research endeavor. Many agree that “[r]eplication is central to the progress of science: if others cannot reproduce the evidence backing a scientific claim, then the claim loses status as scientific knowledge.”¹²⁶ The Internet can enable exchange of data protocols, and software, and such data sharing should greatly facilitate replication efforts. Researchers should take advantage of these opportunities and also do their utmost to ensure that their studies are free of errors and biases so that they can be duplicated by skilled analysts.

VI. CONCLUSION

This Article has analyzed numerous challenges that researchers face in analyzing big data (and other health information). These include data quality deficiencies; selection, confounding, measurement, and confirmation biases; inadequate sample sizes; sampling errors; effect modifiers; and causal interactions. All of these can cause researchers to mistake mere associations for causal relationships and to reach conclusions that are invalid and cannot be replicated in subsequent studies. Erroneous research findings can mislead legislators, regulators, and lawyers, who use them for purposes of policy-making or litigation.

Formulating solutions for these research obstacles is beyond the scope of this paper, and I have done so in prior work.¹²⁷ Suffice it to say that well-developed

science_is_not_self_correcting_science_is_broken.html [https://perma.cc/5HY7-JYNN]; Joel Achenbach, *Many Scientific Studies Can't Be Replicated. That's a Problem*, WASH. POST (Aug. 27, 2015), https://www.washingtonpost.com/news/speaking-of-science/wp/2015/08/27/trouble-in-science-massive-effort-to-reproduce-100-experimental-results-succeeds-only-36-times/?utm_term=.ab0793214a51 [https://perma.cc/GTT9-6PPL].

124. Christie Aschwanden, *Failure Is Moving Science Forward: The replication crisis is a sign that science is working*, FIVETHIRTYEIGHT (Mar. 24, 2016), <https://fivethirtyeight.com/features/failure-is-moving-science-forward/> [https://perma.cc/ZXS6-9PYT].

125. Mina Bissell, *Reproducibility: The Risks of the Replication Drive*, NATURE (Nov. 20, 2013), <http://www.nature.com/news/reproducibility-the-risks-of-the-replication-drive-1.14184> [https://perma.cc/W5JZ-FHY6] (arguing that “[p]eople trying to repeat others' research often do not have the time, funding or resources to gain the same expertise with the experimental protocol as the original authors, who were perhaps operating under a multi-year federal grant and aiming for a high-profile publication.”); Errington et al., *supra* note 116 (discussing whether failure to replicate always means “that the original result was a false positive”).

126. Errington et al., *supra* note 116.

127. See Hoffman & Podgurski, *supra* note 1, at 527-38; SHARONA HOFFMAN, *Medical Big*

techniques can allow researchers to adjust for biases and identify true causal factors.¹²⁸ However, the pitfalls of big data analysis are numerous, and anyone conducting, reviewing, or relying upon it must be aware of their existence. We all must know what questions to ask and what potential flaws to scrutinize, whether we be scientists, attorneys, policy-makers, or members of the interested public.¹²⁹

Data Quality and Analysis Concerns, in ELECTRONIC HEALTH RECORDS AND MEDICAL BIG DATA: LAW AND POLICY 160-67 (Cambridge U. Press 2016).

128. Hammer et al., *supra* note 68, at 667; Kovesdy & Kalantar-Zadeh, *supra* note 108, at 15.

129. David Faraoni & Simon Thomas Schaefer, *Randomized Controlled Trials vs. Observational Studies: Why Not Just Live Together?*, BMC ANESTHESIOLOGY (Oct. 21, 2016), <https://bmcanesthesiol.biomedcentral.com/articles/10.1186/s12871-016-0265-3> [<https://perma.cc/5TM5-9CFJ>] (urging that “it remains our accountability to scrutinize methods, controls and conclusions drawn in all the papers we read”).