Introduction

Although this may seem a paradox, all exact science is dominated by the idea of approximation.
Bertrand Russell

Like someone with a hammer who sees nails everywhere, writing a book about bias provokes a belief that bias pervades every aspect of biomedical research. By the same token, the author of a book with this subject matter must take particular care to avoid bias in writing such a work. In this regard, I am unsure as to how successful this book has been. I hope my readers will feel free to point out all deficiencies; I suspect they will not disappoint. To one degree or another, bias is of concern to, and has been examined by, researchers in all scientific specialties and some, notably psychology and epidemiology, have paid considerable attention to the role of bias in their research. Economists have also made significant contributions to the study of bias and provide useful and interesting examples. One consequence of all this attention is that various disciplines have derived multiple names for essentially the same bias, or the same name for different biases, and, as a result, numerous biases (well over 200) are named in the literature. In this book, I define the most common use of bias terms while trying to equate them to other named biases, some perhaps more familiar to the reader. My examination of bias has led to a conclusion that there is a hierarchy of bias; some cognitive biases (I call them first-order biases) are the source of different forms of more specific (second-order) biases. This concept is discussed more fully later in the work.

In 2009, Iain Chalmers and Paul Glasziou estimated that 85% of biomedical research was wasted,[1] a perhaps surprisingly large amount of waste but one that has never been seriously disputed. Many sources of research waste have been identified: unnecessary duplication (e.g., experiments being undertaken to answer questions already settled); poorly designed and conducted research; inadequate or absent synthesis of existing research; lack of protocols and explicit research hypotheses that fail to delineate hypothesis testing from hypothesis generation; limited research reproducibility with poor reward systems for conducting confirmatory research; research projects that go unreported or inadequately reported; and a burgeoning regulatory system that impedes many aspects of the research enterprise at excessive cost and of questionable benefit.[2] Through all of this, the insidious role of bias in biomedical research seems inescapable; examining this bias is the essential purpose of this present work.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 04:02:27, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009682763.002

Four Horsemen of the Bias Apocalypse

There are four essential constructs that befog all biomedical research. I have written about three of them in a previous book: *Risk, Chance and Causation: Investigating the Origins and Treatment of Disease*.[3] The cause of some diseases has been carefully elaborated and, while there is always more to understand about their disease pathology, enough is known to allow successful public health interventions to mitigate their worst effects. Here are two examples.

Tay–Sachs disease is caused by a single mutation in a gene on chromosome 15. The mutation prevents creation of hexosaminidase-A (Hex-A), an enzyme that normally perturbs the build-up of GM2 ganglioside, a fat that in the absence of Hex-A accumulates in the nerve cells of the brain. We carry two copies of every chromosome, one from each parent, and the child that has two copies of the Tay–Sachs mutation will develop the disease, which inexorably progresses to seizures, developmental regression, blindness, and paralysis. There is no cure, and death at age four or five is inevitable. The Tay–Sachs mutation can occur in anyone but it has evolved especially in Ashkenazi Jews from Eastern Europe, among whom 1 in 27 are carriers of the disease, they may have one copy of the mutation that does not result in the disease but if they mate with another carrier, each of their children has a 25% chance of receiving two copies of the mutation and so developing the disease.

Huntington's disease is caused by another mutation on the *HTT* gene on chromosome 4. This gene produces a brain protein that malfunctions when the gene is mutated, causing symptoms that subsequently lead to severe mental aberrations and eventual death, often between the ages of 30 and 50. Unlike Tay–Sachs, only one copy of the mutation is needed to produce the clinical symptoms of disease, which is called a dominant expression. (Tay–Sachs, requiring mutations in both copies of the gene, is an example of a recessive disease.)

Perhaps it would have been wiser not to introduce this book with two tragic conditions for which there is no cure, but they perfectly explain three of the four constructs that are susceptible to bias in biomedical research. The *cause* of each disease is indisputable: it is a specific mutation in one gene. To be sure, there are rare variations for each condition – late-onset Tay–Sachs and earlier-onset Huntingdon's disease – but for the great majority of cases, the cause of the condition is without question. The *risk* of each disease may be calibrated by knowing the prevalence of the mutations in the population and in the parents. The mutation itself is an example of a *chance* event occurring in an ancestor population thousands of years ago, although the same mutation still occurs de novo from time to time.

Success in identifying the genetic causes of Tay–Sachs and Huntington's has led to a sharp reduction in one of the diseases. Prenatal genetic screening for Tay–Sachs has resulted in major reductions in the prevalence of that disease, as the great majority of women who find themselves carrying a fetus with Tay–Sachs, unsurprisingly, choose to terminate their pregnancy. Prevention options are less clear for Huntington's, partly because of the more variable nature of the mutation, which involves repeated copies of a base, and some uncertainty as to whether disease will inevitably follow. Nonetheless, a family history of Huntington's will prompt prenatal screening in many couples, and a choice being made about the continuation of the pregnancy. Success in identifying the causes of these conditions had a massive impact on the direction of biomedical research for many decades and continues to the present time. If the genetic causes of Tay–Sachs and Huntington's could be uncovered, as well as many other important but relatively rare conditions, why not cancer, heart disease, depression, schizophrenia, asthma, or autism?

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 04:02:27, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009682763.002

Billions of dollars and years of scientific research have been spent on what has been a largely fruitless search for single genes that can explain the causes of what are now called complex chronic diseases. Instead, what has been importantly revealed are those genes, dozens in many cases, that contribute to increasing, usually in tiny increments, the *risk* of a specific disease. We now understand these "susceptibility genes" are responsible for supporting numerous biologic pathways that promote the initial onset of disease and for driving other pathways necessary for what will become their clinical presentation, subsequent course, and prognosis. More single-gene diseases are being, and will be, discovered, but they are for rare, albeit important, conditions; as for the common complex diseases, we now appreciate that much more work is needed to understand their full causal pathways. Nonetheless, there have been many important successes, such as the discovery of mutations that explain a great deal about the causes of adult macular degeneration, arguably the first successful discovery of a genetic cause for a complex disease, to be discussed later.

One of the best-known genetic causes of cancer are the *BRCA1* and *BRCA2* gene mutations for breast cancer discovered in 1994 on chromosomes 17 and 13, respectively.[4,5] The life-time increased risk of breast cancer with a *BRCA1* mutation is about five times, and a thirty-eight–times increase in risk for ovarian cancer. Parents with these mutations have a 50% chance of passing on the mutated gene to their children. Mutations in many other genes increase the risk for cancer and may explain the other half of inherited breast cancer. It is estimated that 1 in 9 women in the United States will develop breast cancer in their lifetime, but only about 20% of these will be inherited. Despite these strong genetic risks for breast cancer, they are not the only cause. Here we must look to possible environmental causes; many have been examined but few are agreed upon. Tobacco smoke and ionizing radiation (almost ubiquitous risk factors for all cancers) are the most accordant environmental risk factors. There is plenty of agreement for the reproductive risk factors: nulliparity (high breast cancer rates in nuns being an early clue), delayed age at first birth, early menarche, and delayed menopause, all of which relate to having higher estrogen levels. There is now great interest in looking at how genetic and environmental risk factors may interact to increase cancer risk: Does a particular genetic variant need an environmental condition to increase risk? As we will see in this book, there are numerous opportunities for bias to play a role in all areas of biomedical research, including the successful types of genetic research just discussed. However, with the study of environmental factors as risks for disease, we enter an area of biomedical research where *bias* is particularly likely to raise its ugly head.

After risk, chance, and causation, bias is our fourth horseman of the biomedical apocalypse, and the construct that is the focus of this book; it is the one most likely to plague all our research houses. Indeed, as important as an understanding of risk and chance are in the search for causal disease risk factors, it is bias that threatens the validity of all studies and experiments, regardless of the research methodology. In some circumstances, it is no exaggeration to say that bias poses an existential threat to entire fields of research This will become apparent in what follows. Appropriate research methods can mitigate many of the threats posed from bias, but no methodology is completely free from the possibility of bias, not least from bias inherent in scientists themselves.

While writing this book, it became clear that a broad perspective on bias was needed, starting with the susceptibility for bias among scientists, professionals among whom I am proud to be included. I take the, hopefully not too depressing, view that bias is in our DNA and must be unlearned, not just in our scientific lives but in everyday life. An innate

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 04:02:27, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009682763.002

susceptibility for bias is something all investigators and research teams must be alert to and trained to avoid in their research. Sadly, formal training to reduce personal bias is often lacking; some suggestions for addressing this form part of the book's conclusion.

Opportunities for bias in the research process start with limitations to the range of funding opportunities available to scientists, they include the nature of the hypotheses themselves, and who is chosen to be a research subject. Perhaps as expected in a book of this type, there is a panoply of important biases found in biomedical research methodology: in data collection, data analysis, and interpretation. The middle sections of the book cover bias in how data are synthesized and how research findings are promulgated. Special research areas invoke their own particular forms of bias; herein are reviewed: animal research, randomized controlled trials, genetic research, pandemic studies, and pharmacoepidemiology research. The book concludes with suggestions for avoiding bias when designing research projects and offers suggestions about anti-bias training for scientists.

The reader may wonder, given so many opportunities for bias, how so much reproducible and seemingly unbiased science is ever accomplished. One must appreciate what a huge amount of science is being done, thousands of experiments are conducted, and millions of statistical associations are analyzed in data sets daily; well over 1.5 million publications enter the PubMed database (just one example) each year. Even if a small proportion of research were biased, it would sum to a large absolute number of error-prone studies. Indeed, there is increasing evidence that many studies may be producing exaggerated or incorrect results.

Biomedical scientists understand that two major sources of error threaten their research: random error (chance) and systemic error (often labelled systematic error), the latter being how "bias" is broadly defined. Differences in these types of bias are explained in Chapter 1. Random error produces results that are less precise than may be desirable, leading to either false-positive or false-negative results (under a statistical testing paradigm, more about this later), but the imprecision is usually obvious from the variation or degree of confidence surrounding the result. Random error can be managed; it is wasteful of resources but rarely seriously misleading. Systemic error, on the other hand, is much more pernicious in its consequences. It can produce gross errors in a research result, sometimes reversing a true positive association to a negative one (or vice versa), and yet there may be little indication in study data that bias is even present. It is up to investigators to hypothesize how bias may be corrupting their calculated risk estimates and to undertake analyses that can document how much the presence of bias may have influenced their study result. Thus, detecting bias depends on the imagination of the investigator to initiate appropriate statistical analyses that will confirm or not whether bias is at play. To make matters even more interesting, and to paraphrase Shakespeare's Cassius, the fault may lie "not in our stars, but in ourselves."[6] Scientists may themselves be biased but not know it. This is no inconsequential matter, the book discusses several biased studies that have severely disrupted and misled public health and medical practice, and so waylaid the search for new biomedical knowledge.

Given the major role that systemic error (what we here call bias) plays in disrupting biomedical research, one would expect that it is the subject of considerable attention in graduate school research education and prominent in research proposals. Surprisingly, this is not the case. Postgraduate training involves taking many courses in statistical testing, such as using different types of regression analysis and modelling, which largely address random error, but less attention may be paid to how biased data are inadvertently collected and remains undetected and unmitigated. Graduate school courses that focus on the role of bias

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 04:02:27, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009682763.002

in research are surprisingly rare. Similarly, in grant proposals, a "power calculation," which addresses how a study, under proscribed conditions, can avoid or at least reduce random error, is mandatory. There is no equivalent expectation for a formal bias analysis, which declares how the proposed study will manage the potential biases threatening the study design and its conduct. To be sure, most scientific proposals, if they have any chance of being successful, attempt to address the threat of bias in the proposed work, but it is usually provided in an ad hoc and incomplete fashion. The book includes suggestions for how to address this need.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 04:02:27, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009682763.002