PART II

ANALYSIS OF THE PROBABLE AETIOLOGY OF NON-COMMUNICABLE DISEASES

Whether you can observe a thing or not depends on the theory which you use. It is the theory which decides what can be observed.

— Albert Einstein (1879–1955)

INTRODUCTION

> In the fields of observation chance favours only the prepared mind.
>
> Louis Pasteur, 7 December 1854 (dictum)

To paraphrase the dictum of Louis Pasteur, it can be said that in the fields of induction, chance favours only the prepared mind. To prepare the mind of the reader, I will call attention to some essential medical facts. These facts can be used according to Bayes' rule, which can be described in one sentence: by updating our initial beliefs (about something) with objective new information, we get a new and improved belief. (1)

Medical knowledge has advanced enormously since the fundamental discoveries of Louis Pasteur. However, despite the advanced medical knowledge of today, there are still a surprisingly great number of diseases with unknown aetiology. This is not just a scientifically astonishing fact, but it is, even more, an extremely disappointing fact to patients suffering from the disease entities and to the clinicians longing for a scientific foundation to treat the patients with the disease.

IS THERE A COMMON CAUSE OF DISEASES WITH STILL UNKNOWN AETIOLOGY?

It should be noted that a great number of diseases with still unknown aetiology correlate significantly even though many of them seem to be unrelated. Just one example is the significant excess in all-cause mortality among men and women with knee or hip osteoarthritis, which is particularly pronounced for cardiovascular- and dementia-associated mortality. (2) Another example is that in men with coronary heart disease, it is the sole condition in only 17% of cases. (3) In daily clinical practice, humans seem to be either completely healthy or they have a history of a number of diseases.

A theoretical approach could be to view the different degrees of the same diseases in different individuals and the mixture of different diseases within the same

---

1. McGrayne SB. The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines and emerged triumphant from two centuries of controversy. Yale University Press, 2011.

2. Nüesch E, Dieppe P, Reichenbach S, Williams S, Iff S, Jüni P. All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. *BMJ* 2011;342:d1165.

3. Weiss CO, Boyd CM, Yu Q, Wolff JL, Leff B. Patterns of prevalent major chronic disease among older adults in the United States. *JAMA* 2007;298:1160–2.

individuals in relation to the theory of deterministic chaos rather than random chaos. Instead of studying the irregularities by resorting to statistics, we could try to tease out the hidden patterns that characterize deterministic chaos. (4)

The unexplained high frequency of multimorbidity observed by medical clinicians on a daily basis is so conspicuous that it might indicate a common underlying aetiology—at least theoretically. It should not be forgotten that prior to the discovery of the tubercle bacillus by Robert Koch in 1882, tuberculosis was not one disease, but a number of disease entities. Similar to the theorized Higgs particle that is predicted to imbue elementary particles with mass according to the Standard Model of particle physics, it might be theorized that there should be a sort of particle imbuing the diseases with still unknown aetiology in humans.

Rather than developing a theory that predicts what sort of particle might cause these diseases in humans, medical doctors blindly search for correlations with diseases in medical databases without a theory to predict what is expected to be found according to the theory. When these findings are presented, the concepts risk factor, determinant, risk marker, correlation, and causality are much too often confused and the relationships that are proposed are actually rather spurious. Correlation does not imply causation, and there are five ways to illogically infer causation from correlation. (5) The most common one in medicine is probably to ignore that there might be an unknown common causal variable of the correlation.

For example, it has been found that poor school performance is linked to a variety of diseases in adulthood and that there is a correlation between higher levels of education and decreased morbidity and mortality rates. (6) Differences in lifestyle can explain this correlation only to some extent. But no reasonable theory has been proposed to explain the correlation between levels of education and morbidity and mortality. Theoretically, there might be a common cause of cognitive and physiological dysfunction that impairs the ability to educate oneself and to maintain physical health.

Furthermore, tobacco smoking is not the sole cause of lung cancer even though there is a correlation. Only a minority of smokers actually develop disease. Smoking is simply an especially important risk factor for developing lung cancer and also correlates to a great number of other diseases. One interpretation is simply to consider smoking as a risk factor for developing a number of diseases, but the correlation may as well be interpreted as a behavioural hallmark of people at risk to develop diseases. It should not be forgotten that people who smoke exhibit this behaviour despite the knowledge about the health dangers related to smoking.

OF WHAT NATURE COULD THE COMMON CAUSE BE?

To the extent that we actually know the fundamental cause of human disease and mortality, it is basically of a genetic or infectious nature, if we discount physical

4. Stuart I. 17 equations that changed the world. Profile Books, 2012.

5. Correlation does not imply causation. In: Wikipedia, The free encyclopedia. July 7, 2012. http://en.wikipedia.org/wiki/Correlation_does_not_imply_causation

6. Almquist YA. School performance as a precursor of adult health: exploring associations to disease-specific hospital care and their possible explanations. *Scand J Public Health* 2013;41:81–91.

injuries, poisons, vitamin deficiencies, and starvation. As a result of a lack of infectious explanations, genetic explanations have had a favouring wind during the last few decades to explain numbers of diseases. However, in the light of millions of years of evolution of life, it might seem improbable that genetics can account for the vast field of diseases with unknown aetiology that we still observe. No matter what the probability of hypotheses about genetics as causes of diseases might be, hypotheses should be tested. During the last two decades, immense resources have been spent on research that examines the links between genes and disease.

Technical advances in the field of human genetics have given us the opportunity to industrialize this type of medical research. The opportunity has willingly been grasped, and trawling for genetic factors associated with different diseases has generated a plentiful harvest. (7) Despite an exponential increase in the number of genes associated with different diseases gained by this hypothesis fishing industry, (8–10) most variants identified so far confer relatively small increments in risk and explain only a small proportion of familial clustering. (11,12) The utility of single genetic markers to improve cardiovascular risk prediction has shown disappointing results, even for the most promising marker, located in the 9p21 region. (13) The use of a multilocus genetic risk score to better capture the complex relationship between genetics and cardiovascular disease has been equally disappointing. After adjustment for traditional cardiovascular risk factors, a genetic risk score comprising 101 single nucleotide polymorphisms was not significantly associated with the incidence of total cardiovascular disease. (14) Similarly, research about cancer markers has been characterized by inflated expectations followed by disappointment when original results cannot be reproduced. (15)

Irreproducible results should not be surprising when the culture of laboratory medicine has not appreciated the difference between experimental research and observational epidemiology. (8) Overfitting can occur when a large number of variables are assessed for a small number of outcomes. When large quantities of potential predictors in discovery-based research can be analysed without a hypothesis, a large number of results or patterns can be derived that might be caused by chance and have no clear biological meaning. (15) The multivariable model might then just

7. Gregersen PK, Olsson LM. Recent advances in genetics of autoimmune disease. *Annu Rev Immunol* 2009;27:363-91.

8. Ransohoff DF. Discovery-based research and fishing. *Gastroenterology* 2003;125:290.

9. Dahl FA, Benth JS. Do split your epidemiological data. *Eur J Epidemiol* 2010;25:759-60.

10. Dahl FA, Grotle M, Benth JS, Natvig B. Data splitting as a countermeasure against hypothesis fishing: with a case study of predictors for low back pain. *Eur J Epidemiol* 2008;23:237-42.

11. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747-53.

12. McClellan J, King M-C. Genetic heterogeneity in human disease. *Cell* 2010;141:210-7.

13. Paynter NP, Chasman DI, Buring JE, Shiffman D, Cook NR, Ridker PM. Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. *Ann Intern Med* 2009;150:65-72.

14. Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, Ridker PM. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 2010;303:631-7.

15. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004;4:309-14.

describe random error instead of the underlying relationship. When independent validation to eliminate random error is carried out in less than 15% of all reports about microarray research (2003), (16) non-reproducible results and inflated expectations should not be surprising.

The missing heritability has been termed the "dark matter" of genome-wide association—dark matter in the sense that one is sure it exists and can detect its influence, but one simply cannot "see" it (yet). (11) My interpretation of the missing heritability in non-communicable diseases is that heritability is without consequence unless an infectious process is initiated. When an infection is introduced in a person, the genetic predisposition of the immune system may make the person more or less vulnerable to the infection and to harmful interactions between the infection and the immune system. After all, immunological resistance to infections and proper modulation of the immune response have obviously been important during evolution among humans as well as among other living creatures.

If we hold on to what we actually know about the causes of diseases and disregard the perpetual optimism of the geneticists, it is reasonable to presume that human diseases with still unknown aetiology are actually caused by microorganisms.

A STANDARD MODEL OF MEDICINE

According to the reasoning above, I will revive Louis Pasteur's "germ theory of disease" and propose a Standard Model of medicine:

> *The deterministic cause of all physical and psychiatric diseases is microorganisms, except for some already known genetically determined disorders, prions, physical injuries, harmful chemical substances, and radiation, as long as the body's metabolic needs are met. The expression of disease is the result of the interaction between the microorganisms, the immune system, and the tissues involved. Susceptibility to disease may be increased due to a genetic predisposition of the immune system, or due to manipulation of the immune system by infections.*

Generally, the chances of discovering something increase by inductive reasoning no matter what you are searching for. Just as searching for oil has proved to be more successful by examination and analysis before drilling than by random drilling, the search for the aetiology of diseases will probably prove to be more successful by examination and analysis in advance than by random searches in medical databases. A medically most important, but somewhat underappreciated, lesson learned from John Snow is to think clinically first and then test and verify the hypothesis by appropriate methods.

16. Ntzani EE, Ioannidis JPA. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 2003;362:1439-44.

11. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747-53.