THE NEW MEDICAL MODEL:
CHRONIC DISEASE and EVIDENCE-BASED MEDICINE

by

Jonathan Fuller

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Graduate Department: Institute of Medical Science
University of Toronto

The New Medical Model:
Chronic Disease and Evidence-Based Medicine

Jonathan Fuller

Doctor of Philosophy

Institute of Medical Science
University of Toronto

2016

Abstract

In the traditional medical model, physicians cure acute diseases using knowledge from the basic biomedical sciences. With the rise of chronic disease and evidence-based medicine (EBM) in the second half of the Twentieth Century, 'the new medical model' emerged. In the new medical model, physicians prevent or manage chronic diseases using the principles of EBM. Chronic diseases and EBM have created a host of daunting problems for modern medicine. In this thesis dissertation, my Central Aim is a philosophical analysis of chronic disease and of EBM, especially concerning treatment and prevention. I also maintain a Central Thesis: many practical problems in chronic disease care and in EBM are intimately connected to conceptual, metaphysical and epistemic problems. The practical problems include: reductionism; fractured care; multifactorialism; the growing burden of chronic diseases; treatment effect heterogeneity, treatment futility and treatment harm; cookbook medicine; the tyranny of aggregate outcomes; RCT worship; unrepresentative trials; and multimorbidity. The philosophical problems I explore are the following. In Chapter 2, I examine the nature of chronic diseases and advance a metaphysical account in which chronic diseases are bodily properties. In Chapter 3, I re-examine existing models of disease classification and argue for a new descriptive model (the 'constitutive model') as well as a

new guiding principle for disease prevention ('the monomechanism ideal'). In Chapter 4, I develop a general account of why mechanistic models often fail to predict the results of medical interventions. In Chapter 5, I reconstruct the standard model of prediction in medicine, the 'Risk Generalization-Particularization (Risk GP) Model'. In Chapter 6, I develop a theory of causal inference in comparative group studies that illuminates the roles of randomization, confounders and causes. Finally, in Chapter 7 I show why EBM's preferred approach to generalizing trial results, 'simple extrapolation', is a deeply problematic solution to the problem of extrapolation. Throughout the ongoing discussion, we indeed find that many practical problems in modern medicine are tangled up with philosophical problems. It will require both medical and philosophical wisdom to unravel them.

For my father,

a physician.

The best physician is also a philosopher.

Galen

Acknowledgments

I am indebted to my thesis advisory committee: Ross Upshur, Ayelet Kuper and Paul Thompson. I was fortunate to have a PhD supervisor in Ross, who encouraged and supported me as I roamed widely both intellectually and geographically during this PhD. Ross's profound and eclectic wisdom, as a physician and philosopher, were a source of tremendous insight and inspiration; his voluminous work on complex chronic disease care and evidence-based medicine inspired many questions explored in this dissertation. I was equally fortunate to have Ayelet as supervisor of my research fellowship in health professions education at the Wilson Centre, undertaken during this PhD. Ayelet provided endless academic and personal support. I benefited greatly from her wide-ranging expertise, and from her constant companionship at the Centre. I am also grateful for Paul's expert guidance, especially early in the PhD when I was a fledgling graduate student studying philosophy and he gave me the tools I needed to become a philosopher.

Much of this dissertation is a tribute to the work of Nancy Cartwright, who supervised me during my visit to the Philosophy Department at the University of California, San Diego in 2015. Nancy was and remains a dedicated intellectual coach and mentor whose work has shaped my thinking more than anyone else's research. I was also privileged to have David Papineau as a supervisor during my visit to the Philosophy Department at King's College London in 2013-14. David inexhaustibly challenged my philosophy and prompted me to think deeper.

I am grateful for my close friend and collaborator Luis Flores, who shares my obsessive drive for getting to the bottom of things, and who is the coauthor of Chapter 5 of this dissertation.

A version of Chapter 5 was previously published with the following citation: Fuller, Jonathan, and Luis J. Flores. 2015. "The Risk GP Model: The standard model of prediction in medicine." *Studies in History and Philosophy of Biological and Biomedical Sciences* 54:49-61. doi: http://dx.doi.org/10.1016/j.shpsc.2015.06.006.

I am thankful to the University of Toronto's MD-PhD Program and the Institute of Medical Science for bravely allowing me to do a philosophy of science PhD in their programs without

Contributions

I am responsible for all work in this dissertation. I express my thanks to Luis Flores, who is coauthor of Chapter 5. Luis helped me to conceptualize the chapter's research question and think through its arguments.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACE	angiotensin converting enzyme
AR	absolute risk
ARR	absolute risk reduction
CAD	coronary artery disease
CAST	Cardiac Arrhythmia Suppression Trial
CCD	causal classification of diseases
COPD	chronic obstructive pulmonary disease
CPG	clinical practice guideline
CVD	cardiovascular disease
DM	diabetes mellitus
DMD	dispositional metaphysics of disease
DN	deductive-nomological
DSM	Diagnostic and Statistical Manual of Mental Disorders
EBM	evidence-based medicine
EBP	evidence-based policy
ES	effect size
EV	external validity
FDA	Food and Drug Administration
G2i	group to individual
GERD	gastroesophageal reflux disease
HPV	human papillomavirus
IHD	ischemic heart disease
ITT	intention-to-treat
IV	internal validity
K(P)	know that P
LDL	low-density lipoprotein
LIP	low insulin production
LIR	low intracellular response
LRA	low receptor activation
MOD-G	Method of Difference Group
MTBC	*Mycobacterium tuberculosis* complex
NCD	noncommunicable disease
NHS	National Health Service
p(O)	probability of O
p($O|F$)	probability of O conditional on F
pheOH	phenylalanine hydroxylase
PKU	phenylketonuria
PUD	peptic ulcer disease
PVD	peripheral vascular disease
RALES	Randomized Aldactone Evaluation Study
RCT	randomized controlled trial
Risk GP	Risk Generalization-Particularization
Rns	lifetime rate of lung cancer among male non-smokers

RR   relative risk
RRR relative risk reduction
Rs   lifetime rate of lung cancer among male smokers
TB   tuberculosis
WHO World Health Organization

Chapter 1 - Introduction

1.1 The Medical Model

"The dominant model of disease today is biomedical" (Engel 1977, reprinted in 2012, 379). When the psychiatrist George Engel wrote these words in 1977 in the journal *Science*, scientific medicine was well established and the model of which he spoke – 'the biomedical model', 'biomedicine' or 'the medical model' – was deeply entrenched. Engel christened a tradition of criticizing 'biomedical' aspects of mainstream scientific medicine that has flourished not only in medicine, but also in the social sciences and humanities, including philosophy (e.g. Leder 1984; Foss 1989; Hewa & Heatherington 1995; Marcum 2008a; Marcum 2008b; Miles 2009; Kontos 2011).

Engel argued that the biomedical model was both the traditional model of disease and the dominant model of practice throughout all of medicine (save for psychiatry). As a model of disease, it conceptualized sickness as a perturbation from normal biological functioning resulting from biological causes, described in the language of the basic medical sciences: anatomy, physiology, pathology, molecular biology and so on. As a model of medical practice, it directed the physician to correct diseases and restore normal functioning using knowledge derived from these very sciences. Engel quotes the psychiatrist Arnold Ludwig, who summarized the model nicely: "According to Ludwig, the medical model premises "that sufficient deviation from normal represents *disease*, that disease is due to known or unknown natural causes, and that elimination of these causes will result in cure or improvement in individual patients" (Ludwig's italics)" (Engel 2012, 378).

This thesis dissertation is about modern medicine. Or rather, it is about two defining facets of contemporary medicine, *chronic disease* and *evidence-based medicine (EBM)*, and about how they have transformed the traditional medical model. For the purpose of this discussion, I will understand *the old medical model* or *the traditional medical model* as follows.[1] Disease is biological, acute and singular; diseased patients have one pathological

---

[1] Donald Gillies (2005) and William Goodwin (2015) characterize our theory of diseases or the set of medical practices dominant in a particular period as a kind of Kuhnian paradigm. Engel suggested that the biomedical model was devised by scientists, and "involved a shared

entity (the disease), and the task of the physician is to diagnose and cure that disease using knowledge from the basic biomedical sciences (as well as the teachings of their forebears and their own clinical experience). Authors from Joseph Levenstein and colleagues (1986) to Mary Tinetti and Terri Fried (2004) agree that our traditional model is a disease-centred one. Although this picture is a caricature of scientific medicine, it is truth-like enough to suit my present purposes.[2]

By the time Engel (1977) wrote his "challenge to biomedicine" in the second half of the Twentieth Century, medicine was changing, both in terms of the kinds of diseases physicians were combatting and the scientific principles upon which they were told to base their judgment. Medicine had entered – or was entering – the era of chronic disease and EBM.

1.2 Chronic Disease

For almost all of human history, acute conditions – infections, deficiencies, and injuries – preoccupied physicians, and the epidemics that brought death and despair to populations were infectious. Then during the Twentieth Century, the world went through an 'epidemiologic transition' (McKeown 2009); noncommunicable diseases (NCDs), including chronic diseases, have joined infectious diseases as the new epidemics and are now the leading killers worldwide (WHO 2014). Because of medicine's new ability to prevent death from acute disease, people are living longer, which puts them at risk for chronic diseases as they age (WHO 2011; 2015).

---

set of assumptions and rules of conduct" (2012, 380), which are indeed features of Kuhn's paradigms.

[2] Not all of the elements of the old medical model are unique to 'scientific' or 'allopathic' medicine. Like the medical model, the ancient Hippocratic physician was concerned with unhealthful deviations from natural equilibrium (Potter 1988), and physicians after Sydenham in the Seventeenth Century diagnosed specific diseases according to a standard disease taxonomy (Jutel 2011). What changed in the Twentieth Century after the birth of scientific medicine was that physicians now received an extensive scientific education during their training, and for the first time had a powerful therapeutic armamentarium with which they could cure many diseases, especially acute infectious diseases (Bliss 2011).

The World Health Organization (WHO) defines chronic diseases as "diseases of long duration and generally slow progression" (quoted in Goodman 2013). They include cardiovascular diseases like hypertension, respiratory conditions like chronic obstructive pulmonary disease (COPD), musculoskeletal disorders like osteoarthritis, and metabolic diseases like diabetes mellitus (DM). Chronic diseases are ubiquitous; the prevalence of chronic conditions in Ontario, Canada was over 50% in 2009 (Koné Pefoyo et al. 2015). Chronic diseases are impactful not only because they are prevalent and have a high mortality but also because of the heavy and persistent burden they impose on patients (IOM 2012). They are expensive for patients and health systems (Hoffman 1996 et al.; Wolff et al. 2002), and often crippling for socially and economically fragile countries (WHO 2014). Chronic diseases are common, consequential, costly, and – for patients and physicians alike – confounding.

Chronic diseases differ in many interesting and important ways from acute infectious diseases, which are the paradigm diseases for the old biomedical model. Because chronic diseases are long-lasting, patients must learn to cope or adapt, to live well with chronic diseases (IOM 2012) rather than attempt to live without them. Cure and restoration of prior biological functioning, the telos of the biomedical model, is no longer the primary aim of care. Instead, chronic diseases must be prevented or managed.

In the realm of prevention, while infectious diseases are often seen as having a simple, single etiology, chronic diseases are often said to be complex and 'multifactorial' in their etiology and to have particular risk factors (WHO 2005). Rather than biological causes, behaviours and social circumstances comprise many of the important risk factors for chronic diseases (WHO 2014). In treatment or management, because chronic diseases are persistent, so too must be our therapeutic efforts. A long duration of therapy and the difficult work of enacting complex treatment often imposes a significant treatment burden on the patient (May et al. 2009; Gallacher et al. 2011). Meanwhile, our therapies typically have modest effect sizes (Moynihan & Cassells 2005), more modest than we might have come to expect given dramatic successes in the treatment of infectious diseases in the first half of the Twentieth Century. Treatments for chronic diseases are also highly heterogeneous in their effects (Kravitz et al. 2004).

Finally, while patients with acute infectious diseases generally present with one infectious disease at a time, many patients with chronic disease have multimorbidity, the co-occurrence of several chronic diseases in the same patient (Akker et al. 1996). Not only does multimorbidity compound suffering for the patient (Gijsen et al. 2001; Fortin et al. 2004; Walker 2007), but it also challenges the biomedical model of diagnosing and treating a single disease (Welsby 1999). We will soon see how medicine has responded to the challenge, but first I will introduce another fundamental shift in the landscape of contemporary medicine, a shift in the very principles upon which medicine is 'based'.

1.3 Evidence-Based Medicine

'EBM' is the name for a powerful movement in medicine, as well as a model of clinical decision-making that this movement promotes and that has become orthodox in medical education and practice. EBM has also been called a paradigm shift in medicine (typically by its proponents (EBM Working Group 1992; Djulbegovic et al. 2009)). The most frequently quoted definition of EBM was provided by David Sackett and colleagues:

> Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research (1996, 71).

This definition does not reveal what kind of external clinical evidence EBM considers to be the 'best evidence'. For that, we must look to the EBM hierarchies of evidence. As Jeremy Howick (2011a) as well Robyn Bluhm and Kirstin Borgerson (2011) argue, the hierarchy of evidence is the central tenet of EBM. It posits that randomized controlled trials (RCTs) and especially systematic reviews of RCTs are a better grounds for clinical decision-making than so-called observational studies, and that both kinds of evidence are superior to evidence from either 'mechanistic reasoning' or clinical experience (EBM Working Group 1992; Guyatt et al. 2008; Howick 2011a).

The EBM movement grew out of clinical epidemiology (Bluhm & Borgerson 2011; Howick 2011a); both began at McMaster University in Canada - clinical epidemiology in the 1960s, and EBM in the 1990s. Evidence-based medicine is essentially the application of the statistical tools, concepts and theories of the clinical epidemiology tradition – plus the

4

hierarchy of evidence and few other 'rules of evidence' - to clinical reasoning. After the movement was announced to the world in 1992 by the EBM Working Group, it was quickly and enthusiastically embraced by the major medical journals and leading medical schools around the world. Yet as rapidly as it was welcomed, it was criticized by many authors (e.g. Tanenbaum 1993; Naylor 1995; Feinstein & Horwitz 1997; Tonelli 1998; Upshur 2002) and outright rejected by others (Charlton & Miles 1998). Bluhm and Borgerson (2011) provide a good summary of the critiques by physicians, philosophers and others. To this day EBM remains deeply rooted and deeply controversial in medicine.

The practice of EBM differs in several notable ways compared to the practice of medicine under the old biomedical model. While scientific reasoning under the old model relied upon knowledge and models borrowed from the basic biomedical sciences (the 'bio' in 'biomedicine'), reasoning in EBM relies on the theory and methods of a very different science, the science of clinical epidemiology. Mechanistic thinking has been supplanted by statistical and probabilistic reasoning, and scientific expertise has come to reflect the ability to critically appraise epidemiological studies in addition to the possession of an impressive store of biomedical knowledge (EBM Working Group 1992). The emphasis on epidemiological studies in EBM elevates the role of population-level reasoning. Doctors must first consider how a population of other patients fared in a diagnostic, prognostic or intervention study. They must then 'apply' the study results to their patient (Glasziou and Mant 2007; Guyatt et al. 2008).

EBM has persuaded physicians to be more reflective about their evidence, thus placing a greater emphasis on explicit clinical reasoning and a far lesser emphasis on tacit knowledge and intuition (EBM Working Group 1992; Malterud 1995; Shaughnessy et al. 1998). As a normative epistemology, EBM has also created standards for *good* clinical reasoning. Meanwhile, evidence-based clinical practice guidelines (CPGs) now set the standards for good clinical practice (Tinetti et al. 2004; Boyd et al. 2005; Fuller 2013a). Practice guidelines are typically disease-specific, which means that the use of guidelines links EBM to our disease-centered medical model. Guidelines are now especially influential in dictating chronic disease care (van Weel & Schellevis 2006; Upshur & Tracy 2008; Dawes 2010; Guthrie et al. 2012). It is within this charged field that the old biomedical model, the treatment of chronic diseases and the practice of EBM have collided, ejecting a new model.

1.4 The New Medical Model

At first glance, applying the traditional medical model to evidence-based chronic disease care appears uncomplicated. Chronic diseases are after all diseases, and the medical model's core imperative is to diagnosis and treat the disease. Further, the seemingly uncontroversial term 'evidence-based' suggests no departure from the approach of treating disease based on evidence from the basic medical sciences. However, recall that in the traditional medical model the aim of therapy is cure of acute disease. Chronic diseases are not acute and are rarely cured; instead, the aims of medical care are disease prevention and disease management. Moreover, in EBM clinical epidemiological evidence is preferred over mechanistic evidence or 'pathophysiologic rationale'. Thus, if the old medical model was to survive the epidemiologic transition and the evidence-based turn, it would have to adapt. And adapt it did.

In describing our dominant model of evidence-based chronic disease care, it is perhaps most accurate to say that the old medical model retained its core principle of diagnosing and treating diseases using scientific principles, while absorbing the aims of chronic disease treatment and the standards of EBM. I will call the orthodox model of diagnosing, preventing and managing chronic diseases according to the rules of evidence-based medicine *the new medical model*.

The model I just described was not what Engel had in mind when he announced "the need for a new medical model" in the title of his famous *Science* essay (1977); the new medical model is the old medical model in new scrubs. Nor is it quite right to claim that the new medical model has succeeded the old model. We have not closed the book on acute medical conditions (infections and injuries), and much of medical practice lies outside of the scope of EBM. It is better to say that the new medical model operates alongside the old medical model (and perhaps alongside still other models). Nonetheless, given the ubiquity and significance of chronic diseases and the wide and commanding influence of EBM, it is fair to say that the new medical model increasingly rules medicine, and is worthy of focussed scholarly attention.

The new medical model diverges from the old model because of its unique diseases and epistemic principles. Moreover, recall that many patients with at least one chronic

disease have multiple diseases. Applying a disease-centred model to the care of patients with multimorbidity results in multiple care plans, each tailored to an individual disease. Kevin Grumbach refers to the approach to acute infection that once defined medicine as the "reductionistic, mechanistic model of disease" (2003, 5), and worries about attempts to impose it on multiple chronic disease care. Ross Upshur and Shawn Tracy (2008) argue that the multiple guideline approach indeed defines our current model for patients with multimorbidity. They argue forcefully that "[t]he current model of clinical management needs to be rethought" (2008, 1656).

Chronic diseases, EBM and the way that the two have come together within the new medical model have precipitated serious problems for modern medicine. It is to these problems that I now turn.

1.5 Problems of Modern Medicine

Over the last several decades, many medical insiders and outsiders have commented on the problems of modern medicine, problems that beset the care of patients with chronic disease, that complicate the practice of EBM, and that challenge our current medical models. Here I will introduce ten of these problems in part to motivate my research. They are: reductionism; fractured care; multifactorialism; the growing burden of chronic diseases; treatment effect heterogeneity, treatment futility and treatment harm; cookbook medicine; the tyranny of aggregate outcomes; RCT worship; unrepresentative trials; and multimorbidity.

1.5.1 Reductionism

Reductionism is a problem for the old medical model as well the new. In fact, Engel labeled the traditional model "The Reductionistic Biomedical Model" (2012, 381). In the medical model according to Engel, diseases are described by the biomedical sciences, especially molecular biology, and they exclude the social, psychological and behavioural dimensions of illness. This reductionistic approach can be characterized as 'methodological reductionism' (De Vreese et al. 2010), or simply 'reduction' (Schaffner 2011). In brief, entities at higher levels of organization (e.g. the level of the whole person) are described in terms of their parts

at lower levels of organization (e.g. molecules) for the purpose of explanation, diagnosis or treatment. Many authors, including many physicians, have criticized our medical models for being 'reductionistic' (Welsby 1999; Grumbach 2003; Cox et al. 2007; Federoff et al. 2009; Beresford 2010; Miles 2009). There is often a worry that the patient's illness experience, values, goals and psychosocial circumstances are neglected or 'reduced away' by our approach to diagnosing and treating biological diseases. Reductionism is a special concern in chronic disease care as illness is often especially burdensome, values especially relevant in evaluating treatment decisions, and psychosocial circumstances particularly relevant in the prevention and management of disease.

1.5.2 Fractured Care

Modern medical care for complex patients is often uncoordinated or fractured (Grumbach 2003; Tinetti & Fried 2004; Starfield 2006; van Weel & Schellevis 2006; Campbell-Scherer 2010; Dawes 2010; Haggerty 2012). Fractured care is intricately linked to reductionism. In reductionism, we fractionate the patient into parts through disease diagnosis or reductive explanation. When care is then organized according to those parts - or "atomized", as Grumbach (2003, 6) puts it - it can easily become fractured, especially when different parts are managed in isolation, by different health providers, and in different settings. Fractured care can lead to confusion for patients and physicians, medical error, treatment interactions, health system inefficiencies, high costs and treatment burden for the patient. Meanwhile, the patient's goals may fall between the cracks. The potential for care fracturing is greatest when patients have multiple chronic diseases that are managed by multiple specialists, often on the basis of multiple disease-specific guidelines (Haggerty 2012).

1.5.3 Multifactorialism

Chronic diseases are often said to be multifactorial in their etiology. Multifactorialism poses several challenges. First, multifactorial etiology is often associated with complexity (Macmahon & Pugh 1970; Krieger 1994; WHO 2005). When there are multiple causes, there is no one cause upon which we can direct all of our prevention efforts. As an illustration, the

8

WHO (2014) prioritizes nine targets for NCD prevention, including insufficient physical activity, salt intake, tobacco use and high blood pressure. Because none of these causes are necessary for disease (they are merely 'risk factors'), preventing one or another will not invariably prevent the disease. Furthermore, because many causal risk factors have a behavioural or social component, preventing chronic disease will often require behaviour or social change, which societies and individuals often struggle to implement, perhaps partly because the behavioural and social determinants of health are less understood and appreciated compared to the biological determinants.[3]

1.5.4 The Growing Burden of Chronic Diseases

The worldwide prevalence of chronic diseases is rising (WHO 2014), despite our best efforts to curb the rising tide. We often attribute the growing burden of chronic diseases in the developed world to aging populations (Jadad et al. 2010; WHO 2011; 2015). Yet perhaps the growing burden also reflects either the peculiar nature of chronic diseases or a failure on the part of medicine to intervene effectively. Part of the problem may be that chronic diseases are generally incurable (Bernstein et al. 2003; McKenna & Collins 2010), and part of our failure may reflect the modest effectiveness of our preventive interventions, including pharmaceuticals and public health measures. The rising prevalence of chronic diseases threatens to overwhelm health systems with costs and high demand for care (Hoffman 1996 et al.; Wolff et al. 2002) lest we find new ways to avert the crisis.

1.5.5 Treatment Effect Heterogeneity, Treatment Futility and Treatment Harm

In modern medicine, the term 'effectiveness' must be tagged with an asterisk. 'Effective' means effective overall in a population, effective for some – but not for all. In fact, effect sizes for newer interventions are typically far less than 50% in absolute terms (Moynihan & Cassels 2005). When we look we often find different effect sizes in different patient subgroups across a range of treatments (Fine 1995; Rothwell 1995; Horwitz et al. 1996;

---

[3] The neglect of the behavioural, psychological and social in the medical model spurred Engel (1977; 1981) to propose a rival biopsychosocial model of disease.

Rothwell 2005; Greenfield et al. 2009; Hlatky 2009; Pignon 2009; Fournier et al. 2010). In short, modern medical interventions demonstrate *treatment effect heterogeneity* (Kravitz et al. 2004). Treatment effect heterogeneity is problematic because it becomes difficult to predict which patients will benefit from a treatment and which will not. At the same time, new drugs – on which our hopes and wishes for medical progress often ride – rarely make it to market (Hay et al. 2014), failing one test or another on the road from discovery to approval. Sometimes, chronic disease interventions that do make it to market are less effective than initially predicted, have unexpected side effects, or are outright harmful. For instance, the diuretic spironolactone initially caused unanticipated and sometimes deadly hyperkalemia in many patients taking ACE inhibitors after the publication of the Randomized Aldactone Evaluation Study (RALES; Juurlink et al. 2004). Modern medical interventions are sometimes *futile*, or worse: outright *harmful*. Yet it is unclear why modern interventions are commonly heterogeneous, futile or harmful in their effects.

1.5.6 Cookbook Medicine

Early in the EBM movement, Sackett and colleagues wrote: "Evidence based medicine is not "cookbook" medicine. Because it requires a bottom up approach that integrates the best external evidence with individual clinical expertise and patients' choice, it cannot result in slavish, cookbook approaches to individual patient care (1996, 72). Soon after, Alvan Feinstein and Ralph Horwitz objected, "Nevertheless, the products of EBM readily lend themselves to the establishment of guidelines and other "slavish cookbook approaches."" (1997, 534). Many others have worried about undue standardization of medical practice due to EBM or CPGs (Tanenbaum 1993; Feinstein & Horwitz 1997; Timmermans & Berg 2003; Mercuri & Gaffni 2011; Fuller et al. 2014; Greenhalgh et al. 2014; Hartzband & Groopman 2016). The major concern with cookbook medicine is that it ignores the relevant particularities of individuals - their unique health problems, circumstances, preferences and values - which compromises 'conscientious' and 'judicious' patient care (Sackett et al. 1996).

1.5.7 The Tyranny of Aggregate Outcomes

Another reason that our standard model of care can lose sight of the individual is that it is centrally concerned with aggregate outcomes in study groups (Tonelli 1998; Timmermans & Berg 2003; Kravitz et al. 2004; Greenhalgh et al. 2014; Tanenbaum 2014). In a population study any differences among individuals are averaged out. There is thus a vexing question of what aggregate outcomes can tell us about individual outcomes. The two kinds of outcome seem prima facie to be incommensurable. Indeed, a recent study argues that doctors often misunderstand the meaning and relevance of treatment effects (Johnston et al. 2016). As Sandra Tanenbaum argues, "EBM has remained mostly silent on the inferential leap from aggregate to individual that is required for actual clinical care" (2014, 934). Understanding the inference from groups to individuals is of urgent importance for a model that privileges epidemiological evidence.

1.5.8 RCT Worship

Commentators pointing to the privileging of RCTs in EBM often quote Sharon Straus and colleagues writing in the textbook *Evidence-Based Medicine*: "If the study wasn't randomized, we'd suggest that you stop reading it and go on to the next article in your search…Only if you can't find any randomized trials should you go back to it" (2005, 118). Despite recent more nuanced approaches to assessing the evidential strength of studies (Balshem et al. 2011), the RCT is still the star that sits atop the hierarchy of evidence. Systematic reviews and guidelines often include *only* RCTs when RCTs are available (Stegenga 2011; Campbell-Scherer 2012; Zuiderent-Jerak et al. 2012; Fuller 2013a). The concern is not only that the reliability of RCTs may be exaggerated, but that other kinds of evidence may be ignored, even in situations in which the other evidence might be epistemically stronger or more ethical. Controversy remains concerning the reliability of RCTs and the importance of randomization (Worrall 2002; La Caze et al. 2012; Senn 2013; Fleming & Ellenberg 2016; Upshur & Fuller 2016).

1.5.9 Unrepresentative Trials

Even while acknowledging the internal validity of RCTs, many commentators worry about their external validity, the extent to which the trial results predict the results of intervening elsewhere (Feinstein & Horwitz 1997; Black 1998; Rothwell 2005; Upshur 2005; Rawlins 2008; Campbell-Scherer 2010). As Richard Kravitz and colleagues put it, "Although randomized controlled trials (RCTs) usually are good experiments, they often are poor surveys" (2004, 667). Enrollment criteria in trials often exclude older patients, patients with multiple diseases, and patients taking multiple medications (Van Spall, 2007). Yet given the demographics of patients in hospital and community practice, target populations often include these very patients that trials exclude. The question of how to cross the gulf between study patients and real world patients is a troubling one (Upshur 2005). Unfortunately, I have found that guidelines extrapolate uncritically to excluded subpopulations (Fuller 2013a).

1.5.10 Multimorbidity

Multimorbidity strains the new medical model more than any other reality of modern medicine. Most people with chronic disease have multiple chronic diseases (Fortin et al. 2005; Salisbury et al. 2011; Barnett et al. 2012), which presents challenges both for patients and their physicians. Multimorbidity is a serious problem in its own right, and the nature of multimorbidity deserves further analysis.

Multimorbidity is also particularly problematical from the perspective of all of the issues raised in this section so far. Managing multimorbidity according to individual diseases is a sort of reductionism in itself: the reduction of complex pathology to component diseases. The reductionist strategy leads to the emergence of new problems. When multiple diseases are managed with multiple medications (polypharmacy), medication regimens can become intractably complex (Fuller & Upshur 2011). Those medications can also interact with one another (drug-drug interactions) and with comorbid diseases (drug-disease interactions) (Tinetti et al. 2004; Boyd et al. 2005; Starfield 2006; Caughey et al. 2011; Guthrie et al. 2012; Mangin et al. 2012). In a frequently cited paper, Cynthia Boyd and colleagues (2005)

apply disease-specific CPGs to a hypothetical 79 year-old patient with five common chronic diseases and identify twelve known treatment conflicts that could result.

Care is greatly at risk of becoming fractured for patients who have multimorbidity because diseases that we assume are discrete are often managed discretely. Along with the growing burden of chronic diseases, there is a growing burden of multimorbidity (Koné Pefoyo et al. 2015). To halt this growth, we might opt for aggressive intervention; unfortunately, we should expect a fair amount of treatment effect heterogeneity, treatment futility and treatment harm within the multimorbid population as comorbidities can plausibly modify treatment effects. The cookbook medicine approach is simply to apply aggregate results to individual patients, regardless of whether or not they have multimorbidity. Yet rather than standardized treatment, the heterogeneous population with multimorbidity requires heterogeneous care (Fuller et al. 2014).

On top of these concerns, patients with multimorbidity are among those individuals that our unrepresentative trials typically exclude. This practice should call into question the relevance of clinical research evidence for real world practice (Upshur & Tracy 2013), as well as the assumption that RCTs are far superior to other studies. Feinstein once cautioned, "By excluding patients with associated diseases from trials of therapy for a particular disease, statisticians can design the trial to contain a presumably 'homogeneous' population, but the results achieved in these 'purified' circumstances cannot be extrapolated to the heterogeneous world of clinical reality" (1970, 456).

1.6 Medicine's Philosophy…Meet the Philosophy of Medicine

1.6.1 The Microscope Problem

The multitude of problems I described in the previous section are problems *for* the new medical model, for physicians engaged in chronic disease care and the practice of EBM. Many of the problems are also *due to* the new medical model, attributable to its faulty concepts, theories, assumptions and inferences. A close examination of chronic disease in general and of EBM might help us to better understand the problems of modern medicine, and might even assist towards their resolution.

The ten problems I selected are largely scientific problems. They concern the science of medicine: the biomedical sciences, including clinical epidemiology, and their application in clinical medicine. Therefore, one might think that we could utilize a 'bioscientific approach' to study them, drawing on the methods of these sciences to gather data and test assumptions; we could put these problems under the scientific microscope, so to speak. Unfortunately, one cannot turn a microscope upon itself. If there is a flaw in a microscope lens - a scratch, let's say – then by peering down the microscope the observer can see through the defect, but they cannot look *at* it. To try and use the methods and assumptions of a medical model to study those methods and assumptions is like attempting to solve the problem of induction inductively (Hume 1777). As the history of philosophy has taught us, it cannot be done.

David Papineau (2014) writes: "The characteristic philosophical predicament is that we have all the data we could want, but still cannot see how to resolve our theoretical problems." The kinds of questions in which I am interested are the following. What could it mean to have multiple co-occurring chronic diseases? Why are chronic diseases 'multifactorial' while acute infectious diseases are not? Does excluding certain patients from clinical trials protect the soundness of our causal inference? How do we extrapolate from study results to target populations? We have some good empirical data on multimorbidity, on chronic disease risk factors, on exclusions from clinical trials, and even some data on the transportability of treatment effects. By no means do we have all the data we could want, nor is it the case that gathering further data would not help to further our understanding. But to answer the kinds of questions I am asking, we cannot rely – at least, not exclusively – on empirical science because the predicament in which we find ourselves is – at least, partly – a philosophical one.

In medicine, we often speak of a 'philosophy of care', which is usually particular to the patient and the circumstances. What are the present goals of the therapeutic partnership between this physician and this patient, what will be done to meet those goals, and just as importantly, what will not be done? The new medical model represents a broader philosophy of care; it embodies goals for clinical medicine (diagnose, prevent and manage chronic diseases), an epistemology (EBM), and perhaps even an ontology describing the entities that

comprise the clinical encounter (e.g. chronic diseases). One way to better understand chronic disease, EBM and our philosophy of care is through philosophical analysis.

1.6.2 The Philosophy of Medicine: an Emergent Field

The philosophy of medicine is a relatively new area of research. As recently as 1992, Arthur Caplan declared that as a field it did not yet exist. Since then, the philosophy of medicine has acquired an international biennial research meeting (the Philosophy of Medicine Roundtable) and an annual journal issue (in the *Journal of Evaluation in Clinical Practice*). The number of sessions and presentations in the philosophy of medicine at major philosophy of science meetings has grown, several philosophy of medicine monographs have been printed (e.g. Howick 2011a; Broadbent 2013; Solomon 2015), a handbook has surveyed many important topics in philosophy of medicine (Gifford 2011), and several textbooks and readers in the philosophy of medicine are forthcoming. Regardless of whether or not I can rightly call the philosophy of medicine a field, it is the area of research in which I will situate this dissertation.

The philosophy of medicine is commonly classified as a sub-discipline of the philosophy of science (Marcum 2008b; Gifford 2011), comparable to the philosophy of biology or the philosophy of physics. I believe that we should more broadly conceive of the philosophy of medicine as cutting across multiple philosophical disciplines within and beyond the philosophy of science, including general metaphysics, epistemology and ethics. Nevertheless, this dissertation will contribute to the philosophy of science literature within the philosophy of medicine. I will also draw selectively on literature outside of the philosophy of medicine (e.g. from the mechanisms literature in the philosophy of biology) and from outside of the philosophy of science (e.g. from the dispositions literature in general metaphysics).

Consistent with most of the philosophy of science work in the philosophy of medicine, I will adopt an analytic approach, *analyzing concepts* of disease, *formally representing* scientific models, and *scrutinizing the logic* of evidence-based practice. I will also endeavor to answer Nancy Cartwright's (2012) plea. Reflecting on evidence-based policy in her presidential address to the Philosophy of Science Association, Cartwright

urged, "We philosophers of science are faced then with a hard job. Here as elsewhere in the natural and social sciences, in policy, and in technology, we can help. But to do so we need to figure out how better to engage with scientific practice and not just with each other" (2012, 988). The philosophy of medicine is exemplary for its engagement with practitioners and practice. I will continue this tradition by grounding my arguments in the medical literature. To understand chronic diseases, I will turn to internal medicine textbooks (e.g. Longo 2012) and credible reports (e.g. WHO 2014); and to understand EBM I will look to EBM guides (e.g. Guyatt et al. 2008; Straus et al. 2011). I will use real cases and cite the primary medical literature whenever it is helpful to my arguments. Medicine is messy and fleshy, philosophy is neat and abstract. However, it is my intention and my hope to provide an insightful and accurate philosophical analysis of normative and actual medical practice.

In the next two sections I will briefly survey the literatures on the philosophy of chronic disease and the philosophy of EBM to identify gaps that impede our understanding of the new medical model.

1.6.3 The Philosophy of Chronic Disease

Very little has been written about chronic disease in the philosophy of medicine. There are more gaps to fill than solid ground to walk on. On the other hand, much more has been written about disease in general. The vast majority of this literature has explored our concept(s) of disease (e.g. Boorse 1977; Engelhardt 1986; Reznek 1987; Wakefield 1992; Nordenfelt 1995; Cooper 2002; Kingma 2007; Ereshefsky 2009; Boorse 2011; Lemoine 2013); in other words, what do we mean by the term 'disease'? There is also some literature on the metaphysics of disease, encompassing the question of whether or not diseases are real and the question of what they are (e.g. King 1954; Whitbeck 1977; Reznek 1987; Nordenfelt 1995; Lange 2007; Simon 2008; Scheuermann et al. 2009; Simon 2011; Smart 2014; Binney 2015; van Loo & Romeijn 2015). However, none of this literature has looked specifically at the concepts and metaphysics of chronic disease and of multimorbidity. A deeper understanding of the nature of chronic disease and multimorbidity may help us to better understand many of the problems of modern medicine.

A few authors in the philosophy of medicine have examined models of somatic disease classification (Whitbeck 1977; Thagard 1999; Carter 2003; Broadbent 2009; Clarke 2011; Broadbent 2014; Smart 2014), and a few authors have looked at risk factors or multifactorialism (McCormick 1988; Rizzi & Pedersen 1992; Stehbens 1992; Schwartz 2008; Broadbent 2013). There is much more to be said about how our models of disease influence disease prevention, especially for multifactorial chronic diseases.

Finally, there is a substantial mechanisms literature in the philosophy of science (e.g. Machamer et al. 2000; Glennan 2002; Bechtel & Abrahamsen 2005; Craver 2007; Bechtel 2008; Darden 2008; Steel 2008; Craver & Darden 2013; Illari 2013). There is also a mechanisms literature in the philosophy of medicine (e.g. Russo & Williamson 2007; Campaner 2010; Andersen 2012; Bluhm 2013; Clarke et al. 2014; Howick 2011b; Howick et al. 2013). Jeremy Howick (2011a; 2011b) has written on mechanistic reasoning involving medical interventions, but otherwise few authors have examined mechanistic prediction in medicine or the link between mechanisms and interventions. Further research is needed to help us understand why our mechanistic predictions often fail and why our chronic disease interventions often deliver unsatisfactory results.

1.6.4 The Philosophy of Evidence-Based Medicine

Compared to research on chronic disease, there is a hefty literature on the philosophy of EBM (for extensive bibliographies see: Bluhm & Borgerson 2011; Howick et al. 2015). Most of the philosophy of EBM literature is critical of EBM (e.g. Tonelli 1998; Worrall 2002; Bluhm 2005; Upshur 2005; Cartwright 2007; Borgerson 2009; Goldenberg 2009; Thompson 2010; Stegenga 2014; Tanenbaum 2014), but some authors defend EBM or various aspects of EBM (e.g. Djulbegovic et al. 2009; La Caze 2009; Howick 2011a).

Despite the fact that EBM is centrally concerned with reasoning from clinical research evidence, there is a surprising paucity of literature on predicting from population studies, especially when it comes to predicting for individual patients (but see Flores 2015). Then again, there is little work on prediction in the philosophy of science in general (Broadbent 2013; Fuller et al. 2015). Most of the authors in the philosophy of EBM literature come at the problem of predicting from group studies through the problem of extrapolation

17

(see below). There is also a small literature on epidemiological risks (Schwartz 2009; Djulbegovic et al. 2011; Broadbent 2013; Dawid 2015). More research is needed on the inference from aggregates to individuals and on the meaning and role of risk in medicine.

In contrast, there is a sizeable literature on probabilistic causality, interventionism and statistical causal inference in the philosophy of science (e.g. Suppes 1970; Eells 1991; Spirtes et al. 2000; Woodward 2003; Cartwright 2010; Steel 2011). There is also a substantial literature on randomization and EBM hierarchies of evidence (e.g. Urbach 1993; Papineau 1994; Worrall 2002; Worrall 2007; Borgerson 2009; La Caze 2009; Howick 2011a; La Caze 2013; Stegenga 2014). However, more work must be done to establish the role of randomization in real RCTs (as opposed to 'ideal RCTs' (Cartwright 2010)) within a robust theory of causal inference, and thus help us to evaluate the reliability of RCTs compared to other comparative group studies.

Finally, the problem of extrapolation or external validity is relevant to the question of what we can learn from comparative group studies, and has been studied in the philosophy of science (e.g. Cartwright 2007; Steel 2008; Steel 2010; Worrall 2010; Cartwright 2012; Howick et al. 2013; Clarke et al. 2014; Marcellesi 2015). Yet EBM's solution to the problem has received scant attention, except in articles by Howick and colleagues (2013) and by Jacob Stegenga (2015). I previously offered a short critique of the EBM approach (2013b), but a fuller analysis is needed given the immense significance of the problem of external validity for RCTs.

1.7 Aims and Thesis

1.7.1 Central Aim, Central Thesis

In this dissertation, my Central Aim will be the following:

A philosophical analysis of chronic disease and of evidence-based medicine, especially concerning treatment and prevention.

The rise of chronic disease and of evidence-based medicine are two of the most consequential developments in medicine over the past hundred years. They transformed the

old medical model of curing acute diseases based on the principles of the basic medical sciences into a new medical model in which chronic diseases are prevented and treated through the practice of EBM. Adapting the traditional medical model to the new realities of chronic disease and EBM has led to the exacerbation of old problems as well as the emergence of new problems for the modern doctor and the modern patient. I documented some of these problems in section 1.5. Many of the problems concern treatment or prevention in particular. For example, *multifactorialism* complicates our prevention efforts; *treatment effect heterogeneity, treatment futility and treatment harm* are problems for many of our therapeutics; most of our interventions are tested in *RCTs* that are *unrepresentative* of real world practice; and treatment interactions arise when we treat patients who have *multimorbidity* with multiple drugs.

I will undertake a philosophical analysis of chronic disease and EBM because many of the practical problems are rooted in philosophical problems. Thus, we can gain a deeper understanding of the practical problems by asking philosophical questions. These philosophical questions concern, for instance: the nature of chronic disease, our chronic disease concepts, the structure of various inferences in EBM, and the justification for various aspects of the EBM methodology. My dissertation will thus contribute to the philosophy of medicine literature on chronic disease and EBM. The literature has begun to address several interesting and relevant problems, but many important gaps in our understanding remain.

Rather than one or more ongoing arguments that are carried from one chapter to the next, each chapter of this dissertation will answer distinct research questions. I do however maintain the following Central Thesis:

Many practical problems in chronic disease care and in evidence-based medicine are intimately connected to conceptual, metaphysical and epistemic problems.

I have chosen the 'intimately connected to' relation to describe the link between the practical problems and the philosophical ones precisely because it is imprecise. There is no single term that captures all of the dependency relations between the practical troubles of modern medicine and its philosophical puzzles. Many practical problems *arise from* philosophical confusion, others *give rise to* philosophical questions, and others we would simply understand better through an analysis of certain philosophical problems. I will restrict my analysis to conceptual, metaphysical and epistemic problems of medicine and bracket a

discussion of ethical problems, not because I think that ethical problems are unrelated to practical problems or that they are unimportant but because historically the ethical problems of medicine have received far more attention. I do not promise to solve the problems of modern medicine from the philosopher's armchair, but in each of the core chapters of this dissertation I will suggest connections between some of the problems of section 1.5 and the arguments of that chapter, and in the Discussion we will revisit the ten problems and see what light my philosophical investigations have shed upon them. The first time one of these ten problems is mentioned in a core chapter, it will be italicized.

This dissertation is not a comprehensive treatment of every interesting philosophical problem concerning chronic disease and EBM (not nearly a comprehensive treatment). It is mostly concerned with the science of medicine – disease entities and evidence-based reasoning. Other important elements of medicine – the social context of medicine, informal clinical judgement, the patient's illness – will be saved for a future discussion. A discussion of various other scientific topics – basic science research, evidence-based diagnosis, decision analysis - will also have to wait.

The core chapters of the dissertation are divided into two parts, each consisting of three chapters. Part One will focus on chronic disease, while Part Two will focus on evidence-based medicine. Chapter 8, the Discussion, will make connections across the previous chapters, highlight my main conclusions, and point the way towards future research.

1.7.2 Part One: Chronic Disease

In Chapter 2, my main research question is: *what are chronic diseases?* In other words, what kind of a thing – object, property, process, or other type of entity – is a chronic disease? I will argue against the usefulness of concept analysis for answering the question. Instead, I will proceed inductively from examples of chronic diseases as they are described in the medical literature. I will argue that chronic diseases are properties, and distinguish the disease from related entities: etiology, pathogenesis and manifestations. I will defend my account against rival views. I will then investigate what makes chronic diseases chronic (long-lasting). Finally, I will explore the nature of multimorbidity, and make a new distinction between token multimorbidity and type multimorbidity.

In Chapter 3, I will ask: *what is our current model of disease classification, and what classificatory ideal should we seek in order to promote effective disease prevention?* I will reconstruct the monocausal model of disease classification that defines a disease type according to a specific cause. I will argue that while we might think that multifactorial chronic and noncommunicable diseases resist the monocausal model, they in fact do not. But I will reject the monocausal model as descriptive of modern disease classification and as an adequate guide to disease prevention. I will argue that a constitutive model describes modern classification for both chronic and infectious diseases. I will further argue that a monomechanism ideal should guide disease prevention, and in many cases it should also guide disease classification.

In Chapter 4, I will investigate the question: *why do mechanistic predictions miss in medicine, especially when predicting the results of intervention?* I will start by defining important concepts ('mechanistic prediction', 'intervention mechanism', 'intervention on a mechanism'). I will also explain four representational properties of mechanistic models: correctness, completeness, abstractness and scope. I will argue for a general account of mechanistic prediction failure in which failure to properly predict the results of medical intervention results from a problem with one of the first three representational properties, creating difficulties for the determination of the model's scope. I will show how my account explains various prediction pitfalls and examples of failed predictions documented in the philosophy of science literature.

1.7.3 Part Two: Evidence-Based Medicine

In Chapter 5, I will ask: *what is the standard model of prediction in medicine?* A brief historical survey will reveal that our standard model grew out of the empiricist tradition. I will call it the Risk Generalization-Particularization (Risk GP) Model because it involves two stages. In the first stage, a risk measure in a study is extrapolated or generalized to the target population. In the second stage, the risk measure is transformed or particularized to yield probabilistic information about an individual member of the population. I will analyze several assumptions on which the model depends, and explain why those assumptions are often problematic. Finally, I will argue that we should be prediction model pluralists, and

will suggest alternate models of prediction that we can use when Risk GP's assumptions are out of reach.

In Chapter 6, my question will be: *what roles do confounders and causes serve in comparative group study causal inference?* First, I will develop a premise called the balance assumption, which maintains that all confounding cases are balanced in a comparative group study. I will show how this assumption probably fails in an RCT, but also that is not the correct ideal for causal inference. I will propose a new theory of causal inference in which confounding causes are not directly relevant. This theory will distinguish between the ideal comparative group study and the required conditions for causal inference. My theory will help ground the idea that randomization prevents selection bias, while justifying the intuition that comparability is the foundation of comparative group study inference. I will close by applying my theory to argue that a common defense of exclusions from RCTs, the homogeneity rationale, is misguided.

Finally, in Chapter 7 I will explore the question: *why is simple extrapolation a poor solution to the problem of extrapolation?* I will analyze EBM's approach to extrapolating from clinical trials: simple extrapolation. The idea is to extrapolate the relative effect size, unless there is a compelling reason not to extrapolate. I will suggest that simple extrapolation rests upon a general generalizability assumption ('most often the relative effect size is generalizable'). I will first argue that this assumption lacks empirical, mathematical, and theoretical justification, and that simple extrapolation is plagued by two reference class problems. I will then argue that without the assumption simple extrapolation is akin to an argument from ignorance. Yet neither epistemic nor pragmatic considerations can justify the argument. Simple extrapolation is too simple, and might lead us to extrapolate too often.

PART ONE: CHRONIC DISEASE

Chapter 2 - What are Chronic Diseases?

2.1 On the Nature of Disease

What kind of a thing are chronic diseases? To use a philosopher's favourite comparison, are they objects like tables and chairs? Or are they properties like mass and rigidity? Or perhaps processes like acceleration and shattering? In a patient with a chronic inflammatory disease, is the disease the collection of inflammatory cells and their products? Or is it the redness and hotness these cells produce in their target organs? Or the activities of the cells, including the reddening and heating they generate? Or is the disease something else entirely? If diseases are constructed, then perhaps they are fictions?[4] And why are chronic diseases chronic (long-lasting), anyway? Is there something in the nature of chronic diseases that explains why they endure through time, causing persistent suffering and public health peril?

Questions about the nature of disease are often of particular interest to historians of medicine and historians of science, perhaps because changes in the dominant conceptions of disease represent perceptible shifts in the intellectual and scientific landscape of medicine. Historians often distinguish two historically important and distinct ideas about disease: the ontological and physiological conceptions (Rather 1959; Temkin 1961; King 1982; Hoffman 2001; Duffin 2010). Although understood differently by different authors (Hofmann 2001), on one reading the ontological view conceives of diseases as separate from patients, while

---

[4] A note on terminology: philosophers often use 'disease' or 'disease entity' to refer to a particular diagnostic category or taxon (Whitbeck 1977; Severinsen 2001; Hucklenbroich 2014), or to the textbook model or knowledge of a particular condition (King 1982; Simon 2008). Wondering about the nature of diagnostic categories might mean wondering whether they are natural kinds or artificial constructs. Meanwhile, wondering about the nature of textbook models of disease might mean wondering about whether they are mathematic models or mechanism schema, or about whether they are ideas, or Platonic objects, or something else. Unless otherwise noted, throughout this chapter I will use the term 'disease' to refer to instances of disease. Wondering about the nature of instances of disease involves asking what the (presumably) concrete disease entity is like (e.g. the arthritis in an arthritic patient). Jeremy Simon (2011) usefully distinguishes disease types (i.e. disease categories) from disease tokens (i.e. instances of disease). In the next chapter, we will turn our attention to models of disease classification, which describe how disease types are defined.

the physiological view conceives of diseases as continuous with patients (Duffin 2010).[5] The contrast is most stark between ontological theories of diseases as foreign *bodies* like miasmas (bad airs) or germs, and physiological theories of diseases as disturbed physiological *states* or *processes*. But an ontological theory in the broader sense could also hold that diseases are foreign states or processes, like the process by which a bacterium infects the body and replicates within it; and a physiological theory could maintain that diseases are diseased parts of the body, like the tumour that consists of the body's own cells.

Throughout history, several influential medical theories or systems of disease have proposed a particular disease ontology. According to Hippocratic humoural theory, diseases were states or properties, specifically humoural imbalances (Duffin 2010). Later, the disease taxonomist Thomas Sydenham viewed diseases as constellations of clinical features, as bundles of signs, symptoms and clinical characteristics (Rather 1959). For instance, according to his famous characterization of gout, "The patient goes to bed and sleeps in good health. About two o'clock in the morning he is awakened by a severe pain in the great toe, more rarely in the heel, ankle or instep…Then follow chills and shivers and a little fever. The pain, which was at first moderate, becomes more intense" (Sydenham 1848).

As Jeremy Simon has observed (2010), ontological questions about diseases have received scant attention from philosophers. A recent exception is Simon himself (2010; 2008; 2011). Simon directs more attention to the question of whether diseases are real, which he compares to realist/antirealist debates about theoretical entities in the philosophy of science (2011). In this paper, I am mainly interested in the more Aristotelian question of what kind of a thing a disease is; to what category of being – object, property, process, or other - does a disease belong?[6] Answering this question might locate the metaphysical roots

---

[5] Jacalyn Duffin (2010) quotes Logan Clendening as writing in 1925: "Surgery does the ideal thing – it separates the patient from his disease. It puts the patient back to bed and the disease in a bottle." Clendening is clearly supporting the ontological view. In contrast, L. J. Rather (1959) quotes the great pathologist Rudolph Virchow as arguing in 1847 for a more physiological conception: "we recognized that diseases were neither self-subsistent, circumscribed, autonomic organisms, nor entities which have forced their way into the body, nor parasites rooted in it, but that they represent only the course of physiological phenomena under altered conditions".

[6] Aristotle believed that there are only two basic categories: 'substances' (objects) and 'accidents' (properties) (Graham 1991). He made no room for fundamental processes or change in his *Categories*.

of biomedical *reductionism*, particularly if chronic diseases involve parts of the body rather than the patient as a whole.

Chronic diseases provide an interesting entry point into the nature of modern disease for reasons that we have already touched on. Unlike our paradigmatic acute diseases, chronic diseases are generally not infections (though some are). They are usually multifactorial in their etiology. They are often progressive in one sense, but always static in another sense (they remain for a long duration of time). They are incurable, despite the many curative miracles of modern medical science. Finally, they are emblematic of modern medicine; to know chronic disease is to know a lot about contemporary disease and contemporary medicine.

In section 2.2, I will argue that to begin to know what chronic diseases are, we need a method that differs markedly from the standard approach to disease in the philosophy of medicine (i.e. concept analysis). In section 2.3, I will apply this method to chronic diseases. Despite what some doctors and medical commentators might have thought, chronic diseases are not fictional entities, objects, processes, or bundles of signs and symptoms. Chronic diseases are bodily *properties*, typically dispositional in nature, but sometimes categorical. I will also clarify the nature of related pathological entities: etiology, pathogenesis, signs and symptoms. In section 2.4 I will ask whether there is something in the nature of chronic diseases that explains why they are chronic. Finally in section 2.5, I will investigate the nature of *multimorbidity*, the phenomenon that magnifies many problems of modern medicine.

2.2 Top-Down vs. Bottom-Up Approaches to Disease

2.2.1 Top-Down

We might expect that the philosophy of medicine has already delivered a consensus answer to our title question. For a good while, philosophers of medicine have been diligently working away at the questions: 'what is health?', and 'what is disease?' In fact, these questions occupy so much of the philosophy of medicine literature that they perform a boundary-setting function for the field, just as the question 'what is knowledge?' helps to

mark out the scope of epistemology and the question 'what is the good?' helps to map the territory of ethics.

Philosophers of medicine have generally understood the question 'what is disease?' as a request for the meaning of the term 'disease'; they have interpreted it as asking for a description of our concept of disease, typically a concept they suppose is shared by all. Thus, the standard philosophical approach to disease is concept analysis. As Maël Lemoine (2013) describes it, concept analysis in the philosophy of medicine involves conjecturing a definition of a term like 'health' or 'disease', usually a reductive definition consisting of necessary and sufficient conditions. The concept analyst then tests their own definition as well as rival definitions against actual or hypothetical cases, each of which is either indisputably a case of disease or indisputably *not* a case of disease. When the goal is a widely held concept of health or disease, the concept analyst usually judges whether each test case is an instance of disease or not an instance of disease by consulting their intuitions. The approach to disease in the philosophy of medicine is thus very similar to the traditional approach to knowledge in epistemology. In epistemology, test cases often take the form of thought experiments like the well-known 'Gettier problem' cases (Nagel 2014), a procedure also used in the philosophy of medicine (Boorse 2011).

In order to determine what kind of thing chronic diseases are, we might then adopt the following procedure. We could first determine our everyday reductive concept of disease, consulting our intuitions as needed. Because a reductive concept includes necessary conditions, we could then scan the list of necessary conditions until we found the one describing the kind of thing – process, property or other - that diseases (necessarily) are, including chronic diseases.

Unfortunately, we should worry that analyzing our concept of disease will fail to deliver an answer. Several concerns motivate my skepticism. First, as Lawrie Reznek argues (1987), it might be that our everyday concept of disease is not reductive, cannot be decomposed into a neat list of necessary and sufficient conditions. Of course, we might be able to rationally revise our everyday disease concept by describing a reductive concept that matches our everyday concept closely enough. On the other hand, we might worry that our everyday concept is just too fuzzy for even this procedure; it would be somewhat like searching for a reductive definition for 'art' or 'game'. The concern that our concept of

disease cannot be analyzed in this manner amounts to a general skepticism about the program of disease concept analysis.

Perhaps the concern is well motivated. We do not have a clear, reductive consensus definition of disease or of chronic disease in healthcare. Richard Goodman and colleagues (2013) list ten definitions of 'chronic disease' or 'chronic condition' from authorities that include the World Health Organization and the US Department of Health and Human Services. All of these definitions differ to varying degrees and none are reductive; most of them define a chronic disease or condition as a disease or condition that satisfies one or more criteria like the criterion of existing for a long period of time. Crucially, none of the definitions tell us what kind of a thing a disease is.

Admittedly, these definitions should probably be viewed as pragmatic working concepts rather than rigorous formal definitions. Maybe the more rigorous concept analysts in the philosophy of medicine have given us what I am looking for? Or maybe not. Despite the fact that the concept of disease debate has been noisily bubbling away since the mid-Twentieth Century, it has not yet distilled agreement, but instead has provided a mix of highly divergent results[7]. Lack of agreement does not entail that the quest is futile, especially in a philosophical debate. Yet one highly plausible explanation for the lack of agreement is that our everyday concept of disease is vague enough (and perhaps also ambiguous enough) to allow for multiple quite different reconstructions, each sufficiently successful.

On the other hand, we do not have to be a general skeptic about disease concept analysis to doubt that the project will deliver a satisfying answer to the question of what sort of entity a disease is. For one thing, if the pure 'normativist' is right that our disease concept is strictly evaluative, then our disease concept will not tell us whether diseases are more like tables or rigidity or shattering. Granted, few if any current participants in the disease concept debate would subscribe to such a strict form of normativism, expunging all descriptive content from our disease concept. But some philosophers can be considered normativists nonetheless because they argue that diseases are bad things without placing constraints on the physical characteristics of those things. For instance, Rachel Cooper defines a disease as a *condition* that "is a bad thing to have, that is such that we consider the afflicted person to

---

[7] See for overviews of the literature Ereshefsky (2009) and Boorse (2011).

have been unlucky, and that can potentially be medically treated" (2002, 271). In medical parlance, 'condition' is a broad term that does not commit us to any specific view on the nature of the entities to which it is applied. For Cooper, if unusual shyness is harmful and can be medically treated, it is a disease (social anxiety disorder); it matters not whether social anxiety disorder is understood as a personality trait (attribute) or a cluster of behaviours (processes).

Even if the 'naturalist' is right that our disease concept is purely biological and not at all evaluative, or if the 'hybridist' is right that our concept is part biological, part evaluative, our concept might still cover entities from multiple categories of being. The fundamental nature of chronic diseases like diabetes and stroke would then be a synthetic fact about the things in the world we call 'diseases' rather a necessary component of our disease concept. Indeed, many naturalist and hybridist accounts currently on offer impose no strong ontological restrictions on diseases. For instance, Jerome Wakefield (1992), a leading hybridist, defines a disorder as a *condition* that (a) causes harm (his normative condition) and (b) results from dysfunction in an internal mechanism (his natural or biological condition).

Christopher Boorse, a leading naturalist, defines a disease as "a type of internal *state* which impairs health, i.e., reduces one or more functional abilities below typical efficiency" (1977, 562; my emphasis). While Boorse *does* take a stance on what type of thing a disease is – a 'state' – he does not defend why a disease is an internal state versus an internal process or some other type of internal thing. Boorse and Wakefield – and most naturalists and hybridists – direct their attention to the nature of dysfunction rather than the fundamental nature of diseases or disorders. In general, the work done by the disease concepts proposed in the philosophy of medicine literature is to distinguish normal (good and/or functional) conditions from abnormal (harmful and/or dysfunctional) conditions, without serious regard to what those conditions really are.

It seems right to me that our concept of disease (if we work with a unitary concept at all) is agnostic as to the nature of disease entities. As evidence, consider the following three conditions, each belonging to a different category of being. The first condition is a statistically and functionally abnormal heart rhythm leading to occasional fainting and an increased risk of death from cardiac arrest. The second condition is the statistically and functionally abnormal conductivity of the heart leading to the deadly heart rhythm I

described. The third condition is the (statistically and functionally) abnormal mass of cells interrupting electrical transmission in the heart, leading to the impaired conductivity and the abnormal heart rhythm. When I consider each of these three conditions on its own, my intuition tells me that it is a disease.[8] Most of the contributors to the concept of disease literature should agree, as each condition has all of the hallmarks of various accounts of disease: the sufferer is unlucky, the condition is potentially medically treatable, the condition is harmful, the condition is statistically abnormal, and the condition is functionally abnormal.

Yet the first condition (the rhythm) is a process, the second (the conductivity) is a property, and the third (the mass of cells) is an object. If the first condition is a disease, our disease concept cannot require that a disease is a property or object; if the second condition is a disease, our disease concept cannot require that a disease is a process or object; and if the third condition is a disease, then – you guessed it - our disease concept cannot require that a disease is a process or property. It certainly seems as though our everyday disease concept permits multiple types of entities. Thus, even if concept analysis succeeds (or has already succeeded) in reconstructing our concept of disease, we have good reason to doubt that analyzing that concept will reveal what chronic diseases are.

2.2.2 Bottom-Up

We need a different approach. The concept analytic approach is a top-down, deductive approach; it identifies overarching conditions that apply to all diseases. If a disease is necessarily a cluster of clinical features (in Sydenham's tradition), then all diseases are clusters of clinical features. In contrast, I will employ a bottom-up, inductive approach. I will analyze particular types of disease such as hypertension and stroke and make inductive generalizations about the nature of diseases.[9] Lemoine (2014) similarly characterizes

---

[8] It might seem redundant and unusual to call all three conditions 'diseases' at once, but that is merely because it is unusual to consider the cause of a disease as a disease, and it is unusual to consider a manifestation of a disease as a disease (though in both cases it would not be unprecedented). If the conductivity is a disease, it seems odd to label its proximal etiology, the mass of cells, as a disease, as well as its proximal effect, the abnormal heart rhythm.

[9] This approach agrees with Simon's contention that we do not need to settle the disease concept question before taking on the disease ontology question; we can answer the latter

traditional concept analysis as a deductive exercise, and argues for an inductive approach whereby we search for unified properties of disease as suggested by science. We must resort to an inductive approach to the metaphysics of disease because no general metaphysical understanding of diseases currently exists in medicine or medical science.

I will focus on the nature of chronic diseases rather than all diseases, sacrificing generality for sound induction. The class of all diseases is heterogeneous. We might expect greater metaphysical diversity within it, which would limit the generalizations we can make about diseases. In comparison, the class of chronic diseases is less heterogeneous in many respects: it consists of entities that persist over a long period of time, and that are often described as multifactorial in their etiology. Further, we can list a tableful of chronic disease kinds (see Table 1) that make up the bulk of what we are typically referring to when we talk about chronic disease. Making sound generalizations about chronic diseases seems tenable. Given the great prevalence and consequence of chronic disease for society, it is also of particular relevance.

In order to throw light upon the nature of chronic conditions, we could turn to several different sources, including popular culture and patient narratives. From these sources we might come to understand chronic conditions as cultural constructs or chronic conditions-as-lived ('chronic illness'), respectively. Although this pursuit is worthy, it is not my pursuit at the moment. I am presently interested in the referents of chronic disease diagnoses ('type I diabetes', 'type II diabetes', and so on), chronic diseases as scientific theoretical entities. I will thus rely on descriptions of chronic diseases in the scientific medical literature.

Which metaphysical categories should we contemplate as candidates in answering the question, 'what kind of a thing is a chronic disease?' Simon lists seven 'realist' possibilities: (1) a concrete entity "wholly separable from the patient"; (2); a "bundle of signs and symptoms"; (3) a "bundle of signs and symptoms plus a (specific) cause"; (4) a "physical state of the body in question"; (5) a "physical state of the body in question plus a (specific) cause"; (6) a "bodily process"; and (7) a "bodily process plus a (specific) cause" (2011, 69).

Simon considers as the first possibility an entity that is separable from the patient. I earlier remarked that a separable entity could be an object, but it could also conceivably be a

---

question by turning to diagnostic entities like cystic fibrosis and asking, "to what do these diagnosis names refer and what is the nature of these referents?" (2011, 65).

process (e.g. bacterial replication), though perhaps not a "bodily process". Instead of an entity "wholly separable from the patient", I will thus explore whether chronic diseases are objects. I will consider it a distinct question as to whether the disease entity – be it an object or process – is separate from the patient (the ontological perspective) or continuous with the patient (the physiological perspective).

Corresponding to possibilities 4 and 5 in Simon's list, I will ask whether a chronic disease is a property (or properties), though I will assume the term 'property' to be roughly interchangeable with the term 'state'. Metaphysicians usually recognize two different kinds of properties: dispositional[10] (or realizable) properties and manifest (or categorical) properties. Dispositional properties are said to realize/manifest certain processes or properties on certain occasions when certain conditions are met. For instance, *solubility* is the disposition of a solute to dissolve when the solute encounters an appropriate solvent, while *fragility* is the disposition of an object to break due to stress or impact. Dispositional properties are often present when their characteristic manifestation is absent, in which case we can say that the dispositional property is present but not realized. A solute is soluble (in some solvent) whether or not it is dissolved, and a china plate is fragile regardless of whether it has broken.

In comparison, manifest properties include our more passive and prototypical examples of properties like shape and size. Manifest properties are always fully realized whenever they are present. Whenever an object has circularity, its circular shape is manifest. Likewise, a person with tallness always has great height.

In addition to Simon's seven 'realist' possibilities, we should consider an 'antirealist' possibility: that chronic diseases are fictions. To say that chronic disease tokens are fictions is to say that there is some sense in which they are not real. Perhaps they are not real because they are not mind-independent (they exist only as ideas), or because they only subsist, or because they are nothing (our chronic disease diagnoses do not refer to anything). As Simon wisely argues (2011), disease *token* realism/antirealism must be distinguished from disease *type* realism/antirealism. It is one thing to ask whether disease types or categories are real

---

[10] Authors sometimes contrast dispositions with other kinds of realizable properties such as tendencies and propensities (Jansen 2007), but I will use 'disposition' as a broad term that encompasses all kinds of realizable properties.

(i.e. natural kinds) or not real (i.e. artificial kinds), and quite another thing to ask whether disease tokens are real (i.e. mind-independent physical entities) or not real (i.e. fictions).

Adapting and enlarging Simon's list, I will consider nine potential answers to the question 'what are chronic diseases?': objects; bundles of signs and symptoms; properties (whether dispositional or categorical); processes; objects and their specific cause, bundles of signs and symptoms and their specific cause; properties and their specific cause; processes and their specific cause; or fictions. While objects, properties and processes are the three categories most commonly considered potentially fundamental in metaphysics, I will not weigh in on which of these nine categories are truly fundamental. Some of these categories – such as 'bundles of signs and symptoms' - are never considered fundamental by metaphysicians. I list these nine answers because all possibilities have been vigorously defended at some point in history by physicians or medical commentators.

To summarize, if we want to know what kind of entity chronic diseases are, we should not turn to concept analysis, the dominant approach to disease in the philosophy of medicine. Rather, we need a new method: a bottom-up, inductive approach to diseases. This method requires making inductive generalizations from a sample of diseases (in this case, chronic diseases) using the medical scientific literature. Although I have argued that the category of being to which chronic diseases belong is not an analytic truth, we will now see that there is a surprising amount of consistency among chronic diseases in this respect. In sections 2.3.1 and 2.3.2, I will present my positive account of the nature of chronic diseases. Then in section 2.3.3, I will defend this account against the eight alternatives under consideration.

2.3 What Kind of Thing are Chronic Diseases?

2.3.1 Chronic Diseases as Bodily Properties

In choosing a sample of chronic diseases, I strove for representativeness. I wanted a list of diagnoses that are unanimously considered chronic disease diagnoses and that account for most of the diagnosed chronic disease in the world so that I could make inferences about chronic diseases generally. I adapted Table 1 from a list of twenty chronic conditions

compiled by Goodman and colleagues (2013) for the US Department of Health and Human Services.

Table 2-1. Chronic Diseases: Type of Entity, Pathogenesis, and Characteristic Manifestation

Chronic Disease	Description	Type of Entity	Pathogenesis	Characteristic Manifestation
Type I Diabetes Mellitus	Glucose intolerance: the inability to physiologically regulate blood glucose, resulting in episodes of sustained hyperglycemia	Dispositional property	Autoimmune destruction of beta cells causing insulin insufficiency	Hyperglycemia
Type II Diabetes Mellitus	Glucose intolerance	Dispositional property	Unclear mechanisms causing insulin resistance (low responsiveness) and insulin insufficiency	Hyperglycemia
Chronic Obstructive Pulmonary Disease (COPD)	A state characterized by airflow limitation that is not fully reversible	Dispositional property	Immune-mediated destructive mechanisms resulting in small airway obstruction and/or decreased lung elasticity	Airflow limitation
Gastroesophageal reflux disease (GERD)	Tendency towards reflux of gastric contents into the esophagus, causing symptoms and/or injury	Dispositional property	Multiple mechanisms leading to failure of the esophagogastric junction to prevent excessive reflux of gastric contents	Reflux of gastric contents into the esophagus
Hypertension	A state giving rise to high blood pressure	Dispositional property	Variable and often unclear mechanisms involving sodium balance, endocrine and autonomic regulation, cardiac output, and renal function	High blood pressure

Coronary Artery Disease (CAD) or Ischemic Heart Disease (IHD)	Tendency towards episodes of myocardial ischemia: insufficient blood supply to heart muscle	Dispositional property	Atherosclerosis occurring in a coronary artery	Myocardial ischemia
Asthma	Airway hyperresponsiveness: tendency towards a physiological state giving rise to airflow obstruction; *or* a constellation of symptoms (wheezing, shortness of breath, coughing) due to variable airflow limitation	Dispositional property; *or* Syndrome	Poorly understood immune mechanisms leading to chronic inflammation of the respiratory mucosa and airway hyperresponsiveness	Airflow limitation; *or* Wheezing, dyspnea, coughing
(Left Ventricular) Heart Failure	Left ventricular dysfunction; *or* a constellation of symptoms (shortness of breath and fatigue) and signs (edema and rales) due to left ventricular dysfunction	Dispositional property; *or* Syndrome	Multiple variable and partly unknown mechanisms, often involving an index event (e.g. heart attack) and compensatory changes (left ventricular remodelling)	Reduced left ventricular blood ejection (or increased filling pressure); *or* shortness of breath, fatigue, edema, rales
Osteoporosis	Decreased bone density or bone strength	Manifest property or dispositional property	Disordered bone remodeling due to various nutritional, endocrine, medication-related and/or lifestyle-related variables	Low bone density, bone fractures
Osteoarthritis	Structural abnormalities of a joint, especially lack of cartilage	Manifest properties	Stress to the joint accompanied by failure of supportive joint structures to prevent injury	Structural joint abnormalities

HIV Disease	Being infected with the human immunodeficiency virus (HIV)	Manifest property	Sexual or blood-born transmission of virus, internalization into CD4+ T Cells, incorporation of HIV genes into T Cell genomic DNA	Infection with HIV
Chronic Viral Hepatitis	Being infected with hepatitis B or C virus	Manifest property	Virus transmitted through blood (hepatitis C) or through blood/bodily fluids (hepatitis B), virus-specific mechanism of hepatocyte infection	Infection with hepatitis B/C virus

I excluded from Table 1 several diagnoses included by Goodman and colleagues that are mental disorders according to the *Diagnostic and Statistical Manual of Mental Disorders* (DSM): 'autism spectrum disorder', 'depression', 'schizophrenia', and 'substance abuse disorders'. Though these disorders are indisputably chronic *conditions*, it is not uncontroversial that DSM disorders are *diseases*; they are more commonly referred to as 'mental disorders' or 'mental illnesses' than as 'diseases'. This terminological difference may be more than merely conventional; mental disorders are defined by phenomenological and behavioural criteria, while so-called 'somatic diseases' generally are not. Because I am after undisputed cases of chronic disease, I will put chronic mental disorders aside for the time being.[11]

In order to reconstruct descriptions of diseases and the pathogenesis of diseases (Table 1), I relied mainly on *Harrison's Principles of Internal Medicine* (Longo 2012), the gold standard as far as medical textbooks go. From the description of a chronic disease I

---

[11] I left out a few other diagnoses included by Goodman and colleagues because the diagnoses were too nonspecific, accounting for multiple types of chronic disease; examples include 'cardiac arrhythmias' and 'chronic kidney disease'. The diagnosis 'cancer', also listed by Goodman and colleagues, is especially nonspecific, and without further qualification could include many acute cancers.

determined the kind of entity described. I also inferred what I call the 'characteristic manifestation' of the disease, which I will explain below.

With the exception of atherosclerosis (discussed in section 3.2), it turns out that all chronic diseases in Table 1 are *properties*. Properties depend on an object, a bearer, for their existence, but are distinguishable from their bearer. Such is the case with these diseases, which depend the patient's body or parts of the patient's body, but are not identical with that body or those parts.

To illustrate, *Harrison's* defines osteoarthritis as "joint failure, a disease in which all structures of the joint have undergone pathologic change, often in concert. The pathologic sine qua non of disease is hyaline articular cartilage loss" (Felson 2012). The disease consists in the pathologic structure of the joint, particularly its cartilage deficiency. Osteoarthritis does *not* consist in the joint itself (an object), which existed in an osteoarthritis-free state before the disease came into existence. Nor does osteoarthritis consist in the "pathologic change" or pathogenic process resulting in the structural abnormalities, which is something that the joint structures have already "undergone". Moreover, osteoarthritis can be asymptomatic and undetectable on physical exam, so it not a bundle of clinical signs and symptoms. Instead, osteoarthritis consists in the osteoarthritic state or properties of the diseased joint.

Here's another example: "Chronic obstructive pulmonary disease (COPD) is defined as a disease state characterized by airflow limitation that is not fully reversible" (Reilly Jr. et al. 2012). COPD is not an object obstructing airflow, nor a pathogenic process resulting in airflow obstruction, nor any symptoms or signs of airflow obstruction, but a "state" giving rise to obstructed airflow. Likewise, type I and type II diabetes are an intolerance to glucose, asthma is airway hyperresponsiveness, osteoporosis is bone porousness or weakness, and HIV disease and chronic viral hepatitis are 'infectedness' of the body's cells with a particular virus.

So chronic diseases are properties. But what kind of properties do these diseases exemplify? Osteoarthritis, HIV disease and chronic viral hepatitis appear to be manifest or categorical properties; they are present just when their characteristic manifestation is present. By *characteristic manifestation* I mean the property or process that defines the disease. For a disease that is a manifest property, the characteristic manifestation is the disease itself. Thus,

the disease is always present when its characteristic manifestation is present, and absent whenever its characteristic manifestation is absent. Osteoarthritis is defined by structural abnormalities, especially cartilage deficiency; HIV disease is defined by infection with human immunodeficiency virus (HIV); and chronic viral hepatitis is defined by infection with the hepatitis B or hepatitis C virus. An individual has one of these three diseases if and only if the corresponding characteristic manifestation is present. This suggests that each of these diseases is identical with its characteristic manifest property.

In comparison, COPD, diabetes, gastroesophageal reflux disease (GERD), asthma and heart failure seem to be dispositional properties. For a disease that is a dispositional property, the characteristic manifestation is the process or property towards which the disease is disposed. Thus, it is not the case that the characteristic manifestation is always and only present when the disease is present; the disease can be absent when its characteristic manifestation is present, and present when its characteristic manifestation is absent. The characteristic or defining manifestation of COPD is airflow obstruction, but COPD is sometimes absent when airflow obstruction is present; for instance, obstruction is sometimes present in asthmatics who do not have COPD. This suggests that COPD is a *disposition towards* airflow obstruction. On the other hand, diabetes is often present when its characteristic manifestation (hyperglycemia) is absent due to effective therapeutic control of blood sugar levels. Diabetes is a disposition towards hyperglycemia rather than the hyperglycemic state itself; while GERD is a disposition towards reflux of stomach contents into the esophagus.

Osteoporosis is an interesting case because it seems to be either a disposition *or* a manifest property, depending on how it is defined. According to *Harrison's*: "*Osteoporosis* is defined as a reduction in the *strength* of bone that leads to an increased risk of fractures…The WHO operationally defines osteoporosis as a bone *density* that falls 2.5 standard deviations (SD) below the mean for young healthy adults of the same sex" (Lindsay & Cosman 2012; emphasis on 'strength' and 'density' is mine). We can call the first definition in this excerpt the 'theoretical definition', and the WHO's definition the 'operational definition' of osteoporosis. On the theoretical definition, osteoporosis is decreased bone strength. Strength in this context can be understood as the tendency to resist fracturing under physical stress (hence the increased risk of fractures when strength is

38

reduced); it is a dispositional property. On the operational definition, osteoporosis is decreased bone density, which seems refer to the porousness of the bone, a manifest property.

There is some confusion – perhaps even disagreement – concerning the definitions of asthma and heart failure. Asthma is often defined as a diseased or disordered (hyperresponsive) state of the airways giving rise to variable airflow limitation (Bateman et al. 2008); but sometimes it is instead defined as a clinical syndrome, a constellation of specific symptoms (wheezing, dyspnea, coughing) resulting from this physiological state (Barnes 2012). Similarly, (left ventricular) heart failure is typically defined as a syndrome (dyspnea, fatigue, edema, rales) resulting from left ventricular dysfunction (Arnold et al. 2006; Mann & Chakinala 2012; McMurray et al. 2012; Moayedi & Kobulnik 2015); yet it is sometimes understood as left ventricular dysfunction itself. Nicholas Binney (2015) quotes Niels Gadsbøll et al. (1989) as distinguishing two distinct conceptual entities: heart failure as "clinical syndrome", and heart failure as "disease state" (cardiac dysfunction).

Whenever there is genuine conceptual ambiguity concerning a particular disease category, it is reasonable to follow Gadsbøll et al. in recognizing distinct concepts, which will often represent distinct things in the world. Based on this strategy, there are chronic diseases as well as clinical syndromes that go under the names of 'asthma' and 'heart failure'. The diseases – like many of the examples encountered so far – are dispositional properties. Heart failure is a disposition towards reduced left ventricular blood ejection. Asthma is a disposition of the airways (realized by triggers such as allergens) towards a state that is disposed towards airflow limitation (an asthma attack); in other words, asthma is a disposition towards a disposition.

Acknowledging conceptual ambiguity and ontological plurality may help resolve disagreements around the diagnosis and management of certain diseases. One physician might insist that echocardiographic evidence of cardiac dysfunction is enough to establish the diagnosis of heart failure, while another might object that the diagnosis additionally requires the detection of signs and symptoms of cardiac dysfunction on clinical exam. Their disagreement may originate in differing opinions on the reliability of echocardiogram for detecting cardiac dysfunction. Or their disagreement may fundamentally concern the definition and nature of heart failure, whether heart failure is cardiac dysfunction or a

syndrome resulting from cardiac dysfunction. The two different ontologies may even promote two different treatment goals: the optimization of cardiac function versus the treatment of cardiac signs and symptoms. Metaphysical incursions into the territory of disease can help to clear up the occasional clinical confusions that are ontological in origin.

So far we have answered the main question we started with. Chronic diseases are properties. In the previous section I argued that the problem of whether chronic diseases are separate from the body or continuous with the body is a distinct question. However, it is one that we can readily answer. After all, chronic diseases are *bodily* properties; that is, properties of the body or its parts. A surgeon could not remove a chronic disease from the patient and place it in a jar intact without removing a part of the patient. Removing osteoarthritis requires removing the osteoarthritic joint (as in joint replacement), while removing HIV disease would require removing all of patient's infected cells. Therefore, the physiological conception has it: diseases are *not* separate from the patient.

We can see from the descriptions in Table 1 that chronic diseases are generally properties of *parts* of the body; COPD and asthma are properties of the *airways*, GERD is a property of the *upper gastrointestinal system*, coronary artery disease is a property of the *coronary arteries*, heart failure is a property of the *heart*, and osteoporosis and osteoarthritis are properties of *bones*. Moreover, dispositional chronic diseases are generally dispositions towards properties of body parts or processes involving body parts; COPD and asthma are dispositions towards *airway* resistance, and GERD is a disposition towards reflux of *stomach contents*. The fact that chronic diseases are properties of parts provides a metaphysical basis for reductionism in medicine, which will be further explored in the Discussion (Chapter 8).

While I have argued that some chronic diseases are dispositions, many metaphysicians believe that many or all dispositional properties have a 'causal base', and some believe that all causal bases are categorical properties (e.g. Armstrong et al. 1996). A disposition depends on its base in some way, though in what way is matter of controversy. Some authors believe that a disposition is identical with its causal base (Armstrong et al. 1996; Mackie 1977; Mumford 1998); we can pick out the property through a disposition ascription, but that property is actually a manifest property if the base is a manifest property. Other authors argue that a disposition is not identical with its base (Prior et al. 1982; Jackson 1998), in which case it is never reducible to a manifest property.

Whether or not we accept an identity thesis, dispositional chronic diseases do seem to have manifest bases. COPD depends on anatomical changes in the lungs, diabetes depends on abnormalities in insulin-secreting and/or insulin-responsive cells, GERD depends on an abnormal esophagogastric junction, and heart failure depends on structural or muscular heart defects. Perhaps dispositional properties like COPD and type I diabetes just *are* manifest properties like anatomical lung changes or low insulin secretory cell mass that are picked out by a disposition ascription. Or perhaps these dispositional properties are metaphysically distinct from their underlying categorical bases.

It seems that we can rule out one possibility: that each of medicine's *types* of dispositional chronic diseases is identical to one of medicines *types* of categorical bases. Consider COPD, which includes three subtypes: emphysema, chronic bronchitis and small airways disease. Each of these subtypes of COPD depends on a different type of lung pathology (Reilly Jr. et al. 2012). Emphysema depends on destruction and enlargement of lung alveoli, while small airways disease depends on narrowed small bronchioles. We can consider these two pathologies as two different types of categorical bases. COPD can depend on either of these two bases because it is multiply realized by emphysema and by small airways disease. Thus, COPD cannot be 'type reducible' to one of these categorical bases.[12] Similarly, diabetes has many subtypes, including type I and type II. Type I and type II diabetes have different physiological bases, the latter's base involving dysfunction in insulin-responsive cells in addition to insulin-secreting cells. So diabetes is not identical to one of these two types of bases.

The examples above serve to rule out the possibility of type identity of each dispositional chronic disease with a manifest base. But as Stephen Mumford argues (1998), though the multiple realizability of dispositions may preclude type identity, it leaves open the possibility of token identity. It could be that nevertheless each *instance* of COPD is identical to an instance of some manifest base; one instance of COPD might be identical to an instance of

---

[12] I have not ruled out the possibility that COPD (D) is 'type reducible' to some disjunction of recognized manifest properties (M_i). That is to say, $D = M_1 \lor M_2 \lor \ldots$ (where M_1 is 'enlarged alveoli' and M_2 is 'narrowed small bronchioles'). However, even if we could perform such a reduction, this disjunction of manifest properties is not a recognized medical category. Medicine does not have an independent categorical label for this type. Thus, COPD as a type is not identical with any medical category of categorical bases.

alveolar enlargement, while another instance of COPD might be identical to an instance of narrowed small bronchioles. Whether token identity of dispositional chronic diseases with their bases is true is a further metaphysical problem that I cannot settle here.

Before moving on, there is another question that I would like to raise. I have argued that chronic diseases are bodily properties - some manifest, some dispositional. This possibility was among the eight 'realist' candidate answers I set out to explore. But what kind of realism does this answer commit me to? It is not committed to a kind of realism that says that chronic diseases exist independently from the patient. On the contrary, chronic diseases are *bodily* properties that are part of the patient - but they are things nonetheless. Further, my account is not committed to 'type realism' about chronic disease categories; it is agnostic as to whether categories like 'type I diabetes' or 'stroke' are natural kinds or artificial, constructed kinds. Nor is it committed to scientific realism about pathophysiological models of chronic disease. A scientific realist believes that our scientific models and descriptions of chronic diseases, their pathogenesis and their manifestations are accurate or approximately accurate. I am claiming only that chronic diseases are properties. My account is neutral as to whether medical science has represented and described the *correct* properties. When a textbook describes HIV disease as infection with HIV, we can wonder whether the virus – a theoretical entity that is unobservable with the unaided eye – really exists. The scientific realist might assert the existence of HIV, and of the property 'infectedness with HIV'. My account makes no such assertions as I have provided no philosophical defense of scientific realism about chronic disease descriptions.

Yet in claiming that chronic diseases are properties, my account is committed to the assumption that for patients diagnosed as having a chronic disease, there is *some* enduring property or properties giving rise to their persistent signs, symptoms and laboratory findings - the observable heralds of disease. The assumption is that there are mind-independent properties that are causally responsible for a patient's chronic ills, and that these properties are exactly what scientific descriptions of chronic disease aim to describe. On this account, chronic disease properties really do exist, though perhaps they are not quite like the entities posited by our current science.

2.3.2 Pathogenesis and Manifestations of Disease

A complete account of the nature of disease should shine light on related entities like pathogenesis, etiology, manifestations, signs and symptoms. Having argued that chronic diseases are bodily properties, I will now explain just what these other pathological entities are.

Let's start with pathogenesis. As the name suggests, pathogenesis describes the genesis of a disease. Returning to Table 1, the pathogenesis of type I diabetes is the autoimmune *destruction* of insulin-producing beta cells of the pancreas, resulting in a lack of insulin secretion. The low beta cell mass or lack of insulin secretion is the physiological base of type I diabetes, upon which the disease itself – a disposition towards hyperglycemia – rests. How about HIV disease, a manifest disease? Its pathogenesis involves the *transmission* of HIV through blood or bodily fluids, and the *internalization* of the virus into CD4+ T cells. At this point the patient is infected with HIV, meaning that they now have HIV disease. The pathogenesis of COPD consists in immune *mechanisms* often triggered by smoking, while the pathogenesis of osteoporosis is disordered bone *remodeling*. Destructions, transmissions, mechanisms and remodelings are all *processes*, events or changes. Thus, pathogenesis is whatever process generates the disease; in the case of chronic disease, the process that produces some enduring pathological state.

Here we run into the curious case of coronary artery disease (CAD) and atherosclerosis. CAD is often cited as one of the leading causes of death in developed countries (Hansson 2005). It is a disposition towards myocardial ischemia or insufficient blood flow to cardiac muscle, and its base is the narrowed state of a coronary artery, narrowed due to atherosclerotic plaque that has built up in the arterial wall. The pathogenesis of CAD is an inflammatory process known as atherosclerosis, which is the buildup of this atherosclerotic plaque.[13] Curiously, atherosclerosis itself is also usually considered a disease (Hansson 2005; Ross 1999; Libby 2012).

---

[13] Annemarie Mol (2002) argues that atherosclerosis is more than just a process, it is also pain in the leg brought on by walking (intermittent claudication), an atherosclerotic artery section viewed under the microscope, and falling blood pressure over the area of restricted blood flow, among other things. For Mol, all diseases are multiple in this manner. However, Mol is explicitly interested in how diseases are enacted by various healthcare practices rather

This convention is unusual for a few reasons. Not only is atherosclerosis a disease that generates other diseases, but the sole clinical relevance of atherosclerosis is that it generates these other diseases. Atherosclerosis is the pathogenesis of CAD and of peripheral vascular disease; it is also part of the pathogenesis in most instances of stroke. Any harmful effects we can attribute to atherosclerosis we can equally attribute to the diseases that atherosclerosis produces. Atherosclerosis has no other harmful effects that would warrant calling it a disease in its own right. Because this process takes place over decades, if atherosclerosis is a disease we should probably consider it a chronic disease. But then as a process it breaks the mold for chronic diseases, which are typically properties. There seems no good reason to consider atherosclerosis a disease – except perhaps political and financial reasons. Evidence suggests that atherosclerosis is widespread, and commonly present even in asymptomatic individuals (Libby 2012). Labeling atherosclerosis as an 'epidemic disease' (Estol 2011) perhaps galvanizes public health efforts to control it, while creating huge markets for drugs designed to treat it.

Related to pathogenesis is a disease's etiology. Sometimes the etiology of a disease is considered synonymous with its pathogenesis. Other times 'etiology' refers to a particular privileged cause of a disease, as when referring to '*the* etiologic agent'. The privileging of a particular cause of a disease is the trademark of what Alex Broadbent (2009; 2013) calls the "monocausal model". In the next chapter, I will argue that the monocausal model is a poor fit for chronic diseases; they do not have a single privileged etiology.

A second family of relevant pathological entities are the manifestations of a disease. We already encountered one kind of manifestation that I named the 'characteristic manifestation' of a disease. Diseases often have manifestations beyond their characteristic ones, which are generally effects of the disease. Among the manifestations of a disease are its signs and symptoms. Signs are manifestations of a disease that are detectable to an observer on physical examination, while symptoms are phenomenological manifestations of a disease that are experienced by the patient.

---

than the unique referents of disease diagnoses or the nature of diseases, which is our focus. According to my disease ontology, 'atherosclerosis' refers to arterial wall thickening; meanwhile, intermittent claudication and falling blood pressure are manifestations of atherosclerosis, and the microscopic image is a representation of the atherosclerotic artery.

Caroline Whitbeck (1977) argues against the view that the manifestations of a disease are correctly conceived as effects of the disease. In her view, "the signs and symptoms which are identifying marks of a disease are not properly understood as effects of the disease process but states of affairs that occur within it" (634). Whitbeck argues that signs and symptoms are parts of the disease. A whole cannot cause its parts, and thus a disease cannot cause its signs and symptoms. However, Whitbeck's argument relies on a different disease ontology than the one I have proposed. As illustrated by the above quote, Whitbeck views diseases as processes – importantly, processes of which signs and symptoms are a part. I will argue against this picture in section 2.3.3, at least as an image of *chronic* diseases.

For now, I note that the diseases in Table 1 do not contain either signs or symptoms as parts. It might sometimes be the case that a manifestation of a disease is part of the disease (e.g. when the manifestation is a characteristic manifestation of a categorical disease). However, it is also clear that many disease manifestations, especially signs and symptoms, are *not* part of the disease. GERD is a disposition towards acid reflux; the acid reflux and the resulting sensation of heartburn (manifestations) are effects of GERD rather than parts of it.[14]

In summary, the pathogenesis of a chronic disease is the process generating the disease property, while the signs and symptoms of a disease – the observable manifestations - are generally effects of the disease. Figure 2-1 illustrates these ideas through the example of

---

[14] At first glance, hypertension might look like a counterexample to my generalization. Hypertension is considered a disease, and hypertension is synonymous with high blood pressure. Yet high blood pressure is a sign (or close to a sign – it is really a theoretical state inferred on the basis of measurements). Then isn't this (pseudo)sign a part of the disease, as the disease per se? The distinction between a disease and its manifestations helps to disambiguate hypertension. It is not generally recognized that there are two senses of 'hypertension': as high blood pressure, and as a disease manifesting in high blood pressure. Yet we need this distinction if we are to make any sense of talk of 'controlled hypertension' (Chobanian 2003; Wolf-Maier 2004). Hypertension is controlled when a patient's blood pressure is normalized, usually through drug therapy. If hypertension *just is* high blood pressure, then we should describe patients with controlled hypertension as having been cured of hypertension (at least, temporarily). But these patients often retain their chronic diagnosis, and that is because hypertension also refers to a disposition towards high blood pressure that endures even when its characteristic manifestation – high blood pressure – is removed. Applying the disease-manifestation distinction to this case helps untie a conceptual knot that we otherwise might trip over.

type I diabetes. As we have already seen, this explanation of the nature of disease is not the only account on offer. In the next section, I defend my view against an octet of rivals.
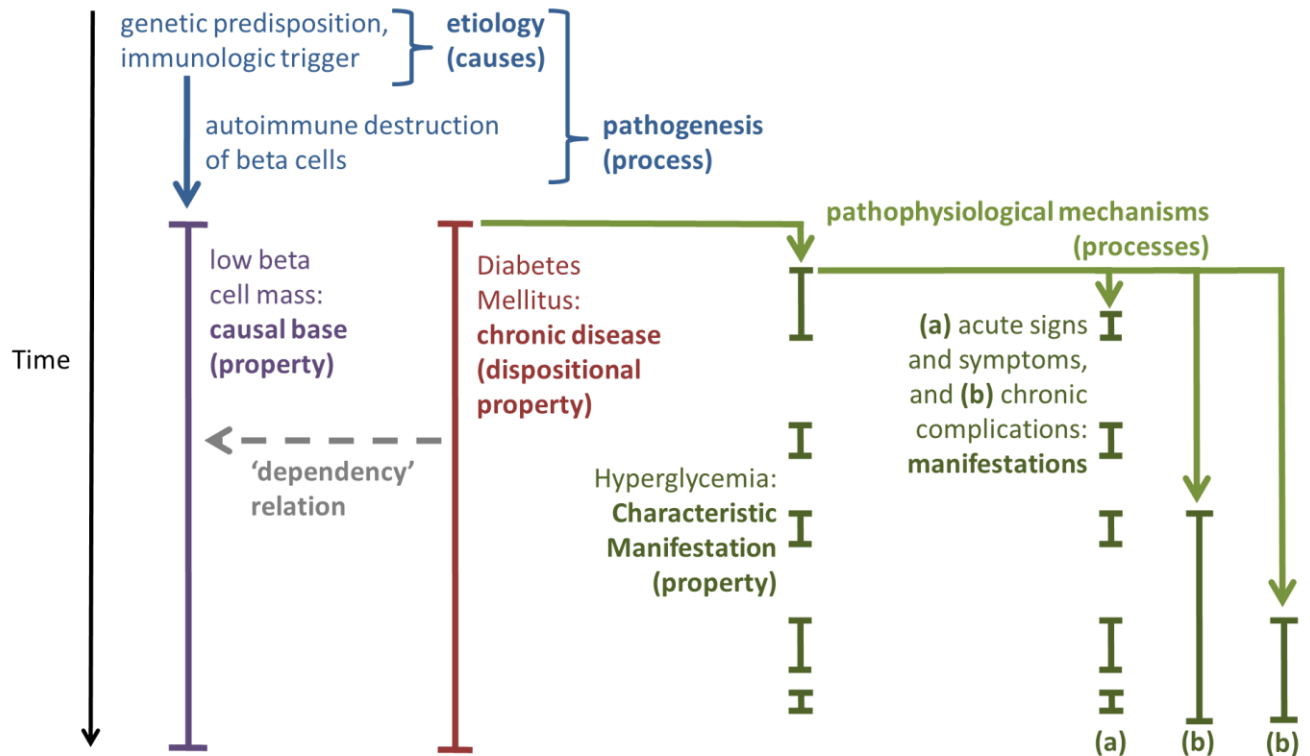
Figure 2-1. The metaphysics of dispositional chronic diseases: type I diabetes mellitus.

2.3.3 Rival ontologies

Earlier, I mentioned eight historically important alternative answers to the question, 'what kind of entity are chronic diseases?': fictions; objects; bundles of signs and symptoms; processes; and any of the previous three entities – or a property - *plus* a specific cause. The difficulty of answering the question is due in part to the fact that pathological entities of these kinds usually coexist with the disease; there are signs and symptoms, diseased body parts, mechanisms of disease, and etiological factors. The challenge is determine which entity at the pathological potluck is the actual disease. I will consider each of the alternatives in turn.

First, there is the fictionalist possibility. One can imagine several versions of this antirealist position. Perhaps our disease diagnoses do not really refer. Perhaps when

physicians diagnose a patient as 'having arthritis' or 'having hypertension', and when they talk about treating arthritis or hypertension, they are being metaphorical rather than literal. Naming the disease could merely be a way of classifying patients who are similar in certain respects. Diagnosing a patient 'with' a particular disease might really mean classifying a person as a particular type of patient, and treating the disease might be medical speak for treating that person as a member of that type.

However, once we have granted that there are certain respects in which a particular class of patients are similar (in virtue of which we class them together), it is more reasonable to believe that disease talk that appears to be about some of these respects *really is* about these respects of similarity. If osteoarthritic patients really do have certain joint similarities, it is strange to think that when a doctor tells an osteoarthritic patient that they 'have arthritis in their joint' what they really mean is that the patient is of the osteoarthritis type.

Another version of fictionalism might maintain that diseases are ideas: when textbooks and physicians describe a disease, they are describing something abstract, not something concrete. This interpretation might explain why two patients with the same disease diagnosis can have unique physiologies and clinical presentations, yet a textbook description of the disease depicts a single entity; the disease entity being described is an abstraction from the unique cases. On the surface, Simon appears to propose such a view when he claims, "Ontologically, a disease is an abstract entity" (2008, 362).

Yet when an abstraction abstracts away the features that make individual cases unique, it leaves shared features represented. To think that ideal textbook entities are anything other than representations is to think that they only *appear* to represent shared patient features. They appear this way because textbooks present them as representations of patient pathology. It is most reasonable to take this practice at face value. After all, there is real patient pathology to be represented. We do not have to believe with the scientific realist that these representations are accurate, but we should believe that they are representations.

Once we accept that abstract textbook entities represent concrete patient pathology, it is a mistake to think that the abstract entity is the disease token. The disease is what is being represented (the patient pathology), not the representation. As evidence for this claim, diseases cause clinical manifestations: signs, symptoms and other findings. Concrete pathology can certainly cause clinical manifestations, but abstract representations cannot.

The abstract representation is best conceived as a *model of* the disease. On closer reading, Simon seems to support this interpretation. Simon argues that medical textbooks include a model - the "abstract entity" he mentions - and that this model "will represent a (constructively) real disease" (2008, 363)[15]. He notes that the abstracted nature of a model can account for variations among patients with the same diagnosis, while preserving the intuition that the model attaches to something in reality.

Returning to our list of alternate ontologies, among the seven 'realist' alternatives are the four 'cause plus' possibilities. These alternatives hold that a disease is an object or syndrome or process or property – *plus* a particular etiologic factor. We can dispense with these four possibilities for two reasons. First, as I already mentioned most chronic disease types are not defined based on a single privileged cause, but are viewed as multifactorial in their etiology. But second, even chronic infectious diseases like HIV disease and hepatitis B do not consist partly in their infectious etiologic agent. HIV disease is not the property of being infected with HIV *plus HIV*. If HIV disease were 'HIV infection plus HIV', it would make no sense to say that HIV disease is caused by HIV because it would amount to saying that 'HIV infection plus HIV' is caused by HIV. Insofar as it *is* in fact coherent to say that HIV disease is caused by HIV, diseases cannot be some entity *plus* its specific cause.

Another alternative to consider is whether chronic diseases are objects. This conception of diseases was more prominent in earlier times when diseases were sometimes thought to be afflictions that invaded the body - bad airs or germs. Diseases are now conceived as *conditions of* the body. When a person has a disease, there is something wrong with the body itself. With the advance of osteoporosis, bones transition to diseased bones, and with heart failure a heart transitions to a diseased heart. The disease that afflicts the person with osteoporosis or heart failure is not an object, it is not the diseased bone or the diseased heart itself.

The view that diseases are bundles of signs and symptoms was also historically popular, beginning with Sydenham. It has also been defended in more recent times. Lester King writes: "When most people talk of disease, they have in mind some particular condition like diabetes or pneumonia, or peptic ulcer. In this more limited sense, the term refers to a

---

[15] Simon argues that the disease is 'constructively real' because his account adapts Ronald Giere's constructive realism (1988; 1989).

pattern of factors which somehow hang together and recur, more or less the same, in successive individuals. Thus, pain in the right lower quadrant of the abdomen, with nausea, vomiting, a fever, and a high white count, spell out the features of acute appendicitis" (1954, 197; my emphasis). L. J. Rather claims that the "clinical entity…is a discrete, orderly *pattern* of events" (my emphasis); these events are "phenomena observed by clinicians" (1959, 366). King and Rather seem to be talking about disease *types*. After all, patterns are ideal entities that we abstract from similar cases. The disease tokens that these patterns represent are "clusters" (King 1954, 198) of observable clinical features, or syndromes.

Rather uses an example to argue for this ontology. As we have seen, Sydenham's conception of gout – like his conception of all diseases – was based on the clinical picture: gout was essentially pain in the great toe or ankle that followed a characteristic clinical course. Rather argues that the clinical entity gout has remained intact since Sydenham's day, despite the advance of physiology and biochemistry. For Rather, the resilience of this disease type despite the changing scientific understanding of the body is evidence that diseases are Sydenham's congeries of clinical events.

Yet though there is still a diagnostic category labelled 'gout' today, it is not the case that our conception of gout has remained unchanged. Sydenham would classify two presentations of ankle pain that fit his description as two cases of gout, but today we would classify these two cases differently if they had a different biochemical basis. We would classify one case as gout if there was a buildup of urate crystals in the joint, and the other case as pseudogout if there was a buildup of calcium pyrophosphate crystals (Schumaker & Chen 2012). That is because we no longer consider gout to be a particular clinical picture, but now consider it to be a property of a joint that has been infiltrated by urate crystals.

Today, constellations of signs and symptoms are called 'syndromes' rather than being called 'diseases'. The two categories of syndrome and disease are generally considered mutually exclusive; thus, chronic *diseases* are not clusters of signs and symptoms (syndromes). The fact that many important chronic maladies are syndromes explains why organizations like the US Department of Health and Human Services refer to the class of chronic medical maladies as 'chronic conditions' rather than 'chronic diseases' (Goodman et al. 2013; WHO 2005; Nasmith et al. 2010).

The final alternative that I will consider is whether diseases are processes, another view that has many supporters. Neil Williams (2007) thinks that a disease "is itself a *process*: a process of deviant cellular interaction outside of typical homeostasis" (556); on the other hand, "disorders" like broken legs, hernias and cataracts "are states of the body that fall outside the standard range" (557). Reznek (1987) contrasts diseases – as processes – with static states like defects, deformities and disabilities. Similarly, Lennart Nordenfelt (1995) distinguishes impairments (states) from diseases (processes).[16] So does Whitbeck: "The distinction is commonly made between diseases and impairments on the basis of the contrast between static conditions and processes. Static abnormal conditions, like hare lip (cleft lip) say, contrast with diseases in that diseases follow a characteristic course. Thus chronic disease is defined as one "of slow progress and long continuance" not as a permanent or enduring state" (1977, 625). Whitbeck concludes that diseases are complex processes rather than states.

However, the fact that diseases in some sense "follow a characteristic course" does not entail that diseases *are* that course. As properties, chronic diseases are produced by a pathogenic process. That pathogenic process may continue after the disease property arises. Because disease properties are usually described qualitatively, a chronic disease persists even as the pathogenic process manifests changes. So in type II diabetes the disposition towards hyperglycemia endures even as insulin resistance and insulin insufficiency worsen. What changes is the severity of the disease. For instance, the disease eventually manifests in hyperglycemia even with certain therapies that previously could control the hyperglycemia. Further, as Marc Lange notes (2007), the disease property itself can produce changes in the body, and these changes are part of the course of the disease. Hyperglycemia produces microvascular and macrovascular changes in the body, yet the disposition towards hyperglycemia endures throughout this progression. A property view of chronic diseases is fully consistent with the slow progression of chronic disease pathology.

---

[16] Nordenfelt (1995) adopts a procedure similar to my own, generalizing from a sample of six diseases described in a medical textbook (Beeson and McDermott 1975). The only chronic disease on Nordenfelt's list is diabetes mellitus, which he characterizes as a biochemical process or state.

Reznek argues that we can determine whether diseases are processes or whether they are states "by seeing whether it makes sense to ask how long the condition takes, or when it began or ended, or whether it has become better or worse" (1987, 72). According to Reznek, it is intelligible to ask these questions of processes but not of states. Since diseases like malaria take time, have a beginning and end, and can get better or worse, diseases are processes rather than states. Yet contra Reznek, states or properties *do* typically last over a period of time; the sun has possessed the properties of being large and hot for billions of years. Further, these properties had a beginning (they came into being – along with the sun - sometime between the Big Bang and the moment at which this sentence is being read) and will have an end (hopefully in the distant future). Finally, as I just argued, there is even a sense in which properties can 'progress'. When it eventually runs out of fuel, the sun will expand into a red giant (becoming even larger) and then contract into a white dwarf (becoming smaller, but large nonetheless). Thus, a property view has no trouble accounting for several temporal features of diseases.

In contrast, a process view has trouble accounting for important characteristics of several chronic diseases. In general, there are two notable processes taking place in patients with chronic disease, either of which might plausibly be the disease. The first is the progression, course, or history of the disease pathology. If type II diabetes were identical to its course, the disease would then consist in the following sequence: development of insulin resistance, followed by development of insulin insufficiency (both of which worsen over time), with the progression of complications like retinopathy and nephropathy (which also multiply and worsen over time). However, some chronic diseases have virtually no progression past a certain point; progress ends yet the disease endures. GERD is a good example. In GERD patients whose heartburn symptoms are well controlled, the disease course is generally non-progressive; symptom severity might fluctuate from day to day but typically does not worsen from year to year. Atrial fibrillation, another chronic disease, similarly does not progress once it develops. If the progress of a disease ends but the disease continues, then the disease must not be the progress.

The second notable process in patients with chronic disease is the pathophysiology in which the disease manifests. In type I diabetes, the characteristic pathophysiological process consists in the mechanisms of hyperglycemia, involving blood glucose, insulin receptors and

insulin-responsive glucose channels as key players. In a patient whose hyperglycemia is poorly controlled, the hyperglycemic process is indeed occurring. Yet in a type I diabetic who is receiving optimal insulin therapy, the disease process has been temporarily suspended. In its place is a new process, a therapeutic process involving insulin and producing normoglycemia. Though the pathophysiological process is gone (at least for now), the disease is not cured. The patient is still possessed of type I diabetes. Again, we should conclude that the disease is not the process, the disease is something that remains throughout treatment. Lange (2007) makes a similar argument with respect to the (chronic) genetic disease phenylketonuria (PKU). Only upon consumption of the amino acid phenylalanine is PKU's characteristic pathophysiologic process - producing excess phenylalanine in the blood – manifest. When the patient is not consuming phenylalanine, there is no pathological disease process, yet the patient has PKU.

Benjamin Smart offers a "dispositional metaphysics of disease (DMD)" in which "diseases are causal processes best seen as simultaneously acting sequences of mutually manifesting dispositions" (2014, 252). Rather than dispositions, for Smart diseases are the processes in which dispositions manifest themselves. He uses the example of epilepsy (a chronic disease). "Epileptics are disposed to have seizures after excessive alcohol consumption; that is to say, alcohol is a stimulus condition for seizures which manifests when consumed by epileptics (in the absence of interfering factors)" (2014, 261). On Smart's DMD, we can consider epilepsy to be the process in which the body's disposition towards seizures and the alcohol's disposition to invoke seizures mutually manifest in seizures. However, like the hyperglycemic process in type I diabetes, epileptic seizures are only realized some of the time. The rest of the time, the patient still has epilepsy. Rather than the causal process in which the bodily disposition towards seizures manifests itself, epilepsy is the dispositional property itself.

Lange also proposes a sort of dispositional account of disease ontology: "diseases as natural kinds of incapacities" (2007, 278). For Lange, an incapacity is a *lack* of a particular disposition; for instance, "PKU is the incapacity to make enough active pheOH" (phenylalanine hydroxylase; 2007, 276), the enzyme that converts phenylalanine to tyrosine. It is questionable whether an incapacity, as the absence of something, is an entity. If we permit the existence of absences as entities in themselves, then we populate the universe with

an infinite number of things – the absence of unicorns, the absence of temperatures below zero degrees Kelvin, and so on. To avoid this metaphysical baggage, we can often construe an incapacity as a capacity. We can redefine PKU as 'the capacity or disposition towards insufficient active pheOH'.

Even if we accept the existence of incapacities, Lange's account over-generalizes.[17] Some diseases are described structurally or constitutionally rather than in dispositional terms. Osteoarthritis and (operationally defined) osteoporosis are structural properties of bone, while HIV disease and chronic viral hepatitis are constitutional properties of cells. It may be that categorical diseases are identical to some disposition (e.g. low bone density is identical to a disposition towards bone fractures), but at least on a descriptive level not all diseases are dispositional. Many authors in the philosophical literature use examples to argue for their accounts of disease ontology. We can criticize their analysis of these particular examples (as I have often done). But furthermore, without establishing that the diseases they choose are representative of all diseases, their inductions are hasty.

We have concluded our tour of rival answers to the question, 'what kind of thing are chronic diseases?' Having ruled out the plausible alternatives, we must again conclude that chronic diseases are properties or attributes. We are now in a good position to explore another aspect of the nature of chronic diseases: their chronicity.

2.4 Why are Chronic Diseases Chronic?

Much of the public health significance of chronic diseases can be attributed to their chronic time course. With a long duration of disease comes a vast accumulation of suffering and financial costs (Hoffman 1996; IOM 2012). Might something in the nature of chronic diseases explain why they are chronic? As we will see, chronicity corresponds to a brute fact about most cases of chronic disease: the persistence of a defining disease property; yet the

---

[17] Similarly, Richard Scheuermann and colleagues (2009) propose that all diseases are dispositions. Although their proposal suggests a consistent ontology for diseases that could provide a revisionary, standardized disease terminology (their objective), it is not descriptively accurate (my objective).

proper explanation of why chronic diseases persist has much more to do with the state of medicine than with the disease state itself.

First, a couple of clarifications are in order. Typically when public health authorities mention chronic diseases they are referring to disease *types*. A particular case of gout or malaria might also be a chronic problem, but - as categories of disease – gout and malaria are not usually considered chronic diseases. Instead, hypertension and osteoarthritis are chronic diseases because the great majority of instances of those diseases are long-lasting. How long is 'long lasting'? Some definitions include a numeric cut-off (commonly, 'longer than a year'), but others do not (Goodman et al. 2013). The WHO's definition merely states that chronic diseases have a "long duration".

Despite a lack of consensus on precisely how long chronic diseases endure, we can still wonder what exactly is doing the enduring. It is not necessarily the pathogenesis of the disease. After all, the process of viral transmission and infection happens relatively quickly in HIV disease and in chronic viral hepatitis. Nor is it necessarily the manifestations of the disease, which can be transient and infrequent, especially if the disease is well managed. Indeed, it is the disease itself that is long-lasting. Chronicity refers to the persistence of the bodily property that instantiates the chronic disease type. Yet there is nothing special about properties compared to other kinds of entities that would explain why chronic diseases endure for long durations. In fact, many bodily properties like 'being asleep' or 'being infected with chicken pox' are short-lived. In order to understand why chronic diseases are chronic, we must explain why *chronic disease* properties in particular tend to persist.

The short answer is that chronic diseases are generally not fatal in the short-term, and are generally incurable. Why are chronic diseases generally not fatal in the short-term? In some cases it is because the manifestations of the disease are typically not lethal (as with acid reflux in GERD). In other cases it is because the potentially deadly manifestations of the disease are well controlled (as with therapeutic insulin preventing hyperglycemic shock in type I diabetes). In these latter instances the disease is not fatal in the short-term because of the existence of effective medical treatment.

Why are chronic diseases generally incurable? Many chronic diseases are incurable because they are associated with irreversible anatomical lesions, as with the destroyed pancreatic islets of Langerhans in type I diabetes, the pathologically remodelled heart in

heart failure, and the destroyed lung alveoli in emphysema. Yet there are such things as *reversible* anatomical lesions. Most bone fractures will heal if we cast them. Irreversible anatomical lesions are irreversible because of a lack of reversing medical treatment. Other chronic diseases like type II diabetes and essential hypertension are incurable because we do not understand enough about the disease in order to intervene and cure.

There is very little in the nature of chronic diseases that explains why they are chronic. Many potentially deadly chronic diseases are not fatal due to the existence of life-prolonging medical treatment, and many chronic diseases are incurable because we lack knowledge of how to cure. Even the 'irreversibility' of certain chronic diseases is not a fact about the disease or its physical basis, but again about the state of medicine's knowledge and its therapeutic arsenal. It does not seem physically (or biologically) impossible that we might be able to therapeutically reverse these lesions at some point in the future. The promise of 'regenerative medicine' (Mason & Dunnill 2008) is to reverse the destruction of pancreatic islets, heart tissue, lung alveoli and other bodily structures by replacing lost islet cells, cardiac cells, alveolar cells, and other cell types through stem cell therapy. While regenerative medicine is little more than a research program at the moment, it does not seem biologically implausible that it should succeed. Thus, these lesions of chronic disease are not irreversible (and incurable) in principle but are merely irreversible (and incurable) at present. To say that certain chronic diseases are irreversible or incurable at present is to make a statement about our current medicine; it is simply to say that currently physicians cannot cure them.[18] Thus, once more it is not the nature of the disease state but the present state of medicine that explains why chronic diseases are chronic.

Which diseases are chronic diseases (and which diseases are not) is historically contingent, dependent partly on our current medical science. HIV disease was once an acute disease. Given its fairly rapid progression and dangerous immune-compromising consequences, it was a terminal condition with few long-term survivors. However, with the development of antiretroviral therapy and improved means of preventing and treating opportunistic infections, patients are surviving longer (Fauci & Lane 2012). Because HIV

---

[18] Of course, the fact that these chronic diseases are curable in principle is perfectly consistent with science ultimately never delivering cures.

disease is now not usually fatal in the short-term (especially in countries with greater access to therapy) and there is still no cure, it has become a chronic disease.

Along similar lines, before the discovery of insulin in the 1920s, type I diabetes carried a dismal prognosis; children lived for a few years at best following diagnosis (Gale 2002). Since the advent of insulin therapy, survival is generally measured in decades rather than months or years. If type I diabetes was not a chronic disease at the beginning of the Twentieth Century, it is certainly one now at the beginning of the Twenty-First Century.

Conversely, peptic ulcer disease (PUD) was probably once a chronic disease, but is no longer. Before the 1980s, most people with PUD were thought to be disposed towards ulcers of the stomach or duodenum due to abnormal gastric acid and gastric mucosal physiology. The goal of therapy was to reduce acid secretion, and acid-suppressing drugs were the mainstay of treatment (Valle 2012; Thagard 1999). However, ulcers would often recur in PUD patients. If the underlying disease was a disposition towards ulcers, then acid suppression might manage the disease, but usually would not cure it. Then in the 1980s, it was discovered that a bacterium – *H. pylori* – was an important cause of peptic ulcer (Thagard 1999); in fact, it is responsible for the majority of PUD (Valle 2012). Most cases of PUD are infectious diseases in which *H. pylori* causes ulceration. Because the underlying disposition towards ulcers depends crucially on the colonization of the gastrointestinal tract by the bacterium, the disease can be cured by eradicating the bacterium using antibiotics. This is exactly what is now done for *H. pylori*-dependent PUD. As a result, PUD is no longer generally considered a chronic disease.

In the case of hepatitis C, the last few years have witnessed the introduction of protease and polymerase inhibitors, ground-breaking new treatments that are highly effective at curing the disease. Yet the cost of the new treatments is prohibitive for most patients in most countries (Brennan & Shrank 2014), and thus in practice hepatitis C has not yet become an acute disease. This example goes to show that it is not only the state of medical science and medical therapy but also the financial and political context of healthcare that determines which diseases are acute and which are chronic at a given place and time.

Notwithstanding a handful of success stories of discovery and cure, as I noted in the Introduction there is a *growing burden of chronic diseases*. The prevalence of chronic diseases in the world may be higher now than in all of human history. Why is there so much

chronic disease? The usual explanation is that human beings are living longer, and since chronic diseases are often caused by cumulative lifestyle exposures and age-related deterioration, people are accumulating chronic diseases (Jadad et al. 2010; WHO 2011; 2015). This explanation interprets the question in a particular contrastive manner: why is there so much chronic disease now compared to any earlier period in the history of medicine?

There is another way to understand the question, as asking why there is so much chronic disease in the world when there could conceivably be much less (even in a world that is aging like ours)? Again I think the default answer would be to point to the current ubiquity of lifestyle factors – poor diet, sedentary lifestyle, smoking - that variously play important roles in the pathogenesis of most chronic diseases. This explanation emphasizes lifestyle components of the *multifactorial etiology* of chronic diseases. A much different explanation would instead focus on the *chronicity* of chronic diseases: the global burden of chronic disease is so high because so many instances of modern disease are instances of pathology that we can manage somewhat but cannot cure[19]. The current chronic disease epidemic reflects the state of medical science and scientific medicine in addition to social determinants of health and disease. We find ourselves in an era of modern medicine (and public health) characterized by a remarkable ability to prolong life, but a lamentable inability to prevent and cure chronic diseases.

---

[19] Another reason why there is currently a preponderance of chronic disease tokens in the world is that - beginning in the second half of the Twentieth Century - medicine has expanded its list of chronic disease types, particularly to include types of states that predispose people to harmful outcomes. Peter Schwartz (2008) calls these new diseases "risk-based diseases". Once considered to be merely predispositions towards cardiac risk factors, the states that dispose people to high blood pressure and to high blood cholesterol are now considered diseases per se ('hypertension' and 'hypercholesterolemia', respectively). Low bone density, at first simply a risk factor for bone fractures, is now the disease osteoporosis; and obesity, initially only a risk factor for type II diabetes and CAD, is now itself a disease. Because predispositions like hypertension, hypercholesterolemia, osteoporosis and obesity have proven difficult to cure, these new disease types are types of chronic disease. I am grateful to Ayelet Kuper for suggesting this complementary explanation for the growing burden of chronic disease.

2.5 What is Multimorbidity?

In the Introduction, I remarked that multimorbidity is a particularly vexing problem for the new medical model. Perhaps as a result of a lack of philosophical reflection on the nature of multimorbidity, there are many conflicting ways of measuring it (Diederichs et al. 2011). Different multimorbidity indices measure different measurands, which suggests that - as in the case of chronic disease - medicine lacks a general metaphysical account of what multimorbidity is. In this section, I will propose such an account.

In part, a metaphysical understanding of multimorbidity eludes medicine because medicine lacks a consensus definition of multimorbidity and of related terms like comorbidity and polypathology (Akker 1996). Alvin Feinstein first defined 'comorbidity' as follows: "the term *co-morbidity* will refer to any distinct additional clinical entity that has existed or that may occur during the clinical course of a patient with the index disease under study" (1970, 455). Marjan van den Akker and colleagues later defined 'multimorbidity' as: "the co-occurrence of multiple chronic or acute diseases and medical conditions within one person" (1996, 69). Akker and colleagues distinguished multimorbidity from Feinstein's concept of comorbidity: while comorbidity refers to an additional 'clinical entity' that exists concurrently with one that the investigator or clinician has designated as the 'index disease', multimorbidity refers to the co-occurrence of the two (or more) entities. Akker and colleagues found inconsistent usage of the term 'comorbidity' in the medical literature, but noted that "[a]ll definitions assume the co-occurrence of *states* or situations related to health." (1996, 66; my emphasis).

Since Akker's review, 'multimorbidity' has similarly come to be used in different ways (Fortin et al. 2005; Smith et al. 2008; Jadad et al. 2010; Diederichs et al. 2011; Salisbury et al. 2011; Barnett et al. 2012; Mercer et al. 2012), with various authors disagreeing on whether the term is restricted to diseases (or whether it refers to medical conditions generally), and whether it is restricted to *chronic* diseases/conditions (or whether it also includes acute diseases/conditions). Akker's definition is among the least restrictive; it is inclusive of chronic and acute diseases as well as other medical conditions. At the less inclusive end of the spectrum, Claudia Diederichs and colleagues define multimorbidity as ""the coexistence of two or more chronic diseases" in the same individual" (2011, 301). In

order to apply my account of chronic diseases, I will adopt Diederichs and colleagues' definition of multimorbidity, although most of my arguments will apply equally well to multimorbidity as defined in more inclusive ways. For us, *multimorbidity* is the co-occurrence, the temporal overlap, of multiple distinct chronic diseases in the same patient.

In medicine, 'disease' is often used to refer to an instance of disease, or disease token. Other times it is used to refer to a type of disease. Both possibilities are left open by the definitions of multimorbidity we have considered. I will refer to the co-occurrence of multiple chronic disease tokens as *token multimorbidity*, and the co-occurrence of multiple chronic disease types as *type multimorbidity*. Since chronic diseases are properties, ontologically speaking token multimorbidity is the co-occurrence of multiple distinct (chronic disease) properties of the same person. Meanwhile, type multimorbidity is the co-occurrence of multiple distinct types of (chronic disease) property. In both cases 'distinctness' is taken to mean 'non-identity' – multimorbidity is the co-occurrence of multiple *non-identical* diseases.

In token multimorbidity, the multiple disease tokens are distinct in that they are not numerically identical, they are not one and the same thing. We can conclude that token multimorbidity is present if we count more than one chronic disease property, and measure token multimorbidity by counting properties. If a patient has bone porousness and joint deterioration in the left knee, they have two chronic disease tokens and have token multimorbidity. We know that the bone porousness and joint deterioration are not numerically identical because they are qualitatively discernable; for instance, one is a property of a bone, while the other is a property of a joint. Two disease tokens can be numerically distinct without being spatially distinct. Bones and joints overlap in space. Similarly, a patient might have a disposition towards myocardial ischemia and a distinct disposition towards poor ventricular blood ejection, which are both properties of the heart.

In type multimorbidity, the multiple disease types are distinct in that they are not conceptually identical, they are not defined the same way. Osteoporosis is often defined as a statistically low bone density, while osteoarthritis is defined as certain structural joint abnormalities (especially cartilage loss); and while CAD is defined as a tendency towards episodes of myocardial ischemia, heart failure is often defined as left ventricular dysfunction. We can conclude that type multimorbidity is present if we count more than one chronic

disease diagnosis, and measure type multimorbidity by counting diagnoses. If a patient meets the diagnostic criteria for osteoporosis and osteoarthritis, they have two types of chronic disease and have type multimorbidity.

Which variety of multimorbidity – type or token multimorbidity – creates problems for the patient and physician within the new medical model? And to which variety are authors in the medical literature typically referring when they define or research 'multimorbidity'? In answer to the first question, either variety of multimorbidity can serve as the basis for *fractured care*, a prominent contemporary problem that I noted in the Introduction. The care of patients with multimorbidity can become fractured if we partition care into pieces according to numerically distinct chronic disease tokens, *or* if we divide care into parts according to conceptually distinct chronic disease types.

In answer to the second question of whether medical authors are typically referring to type multimorbidity or token multimorbidity, authors defining comorbidity or multimorbidity usually stipulate that the multiple diseases exist 'in' or 'within' the same individual, ostensibly ascribing to the diseases a definite physical location. Further, authors often describe the potential for physical or causal interaction among diseases (e.g. Feinstein 1970; Akker et al. 1996). Thus, it seems that they are more likely referring to disease properties and token multimorbidity than disease concepts and type multimorbidity. On the other hand, multimorbidity is often measured by counting distinct diagnoses (Diederichs et al. 2011), which, I remarked, is a direct measure of *type* multimorbidity. Therefore, it may be that various authors presuppose that type multimorbidity and token morbidity map onto each other, at least in practice. If this is the case, perhaps we need not worry about distinguishing type multimorbidity from token multimorbidity; and perhaps we can measure token multimorbidity by counting disease diagnoses.

To be more precise, medical researchers assume that each disease token belongs to one and only one disease type,[20] and that each disease diagnosis a patient receives applies to one and only one disease token; thus, the number of disease tokens in a patient equals the number of disease diagnoses that apply to the patient. If the first part of this presupposition is

---

[20] Peter Hucklenbroich (2014) calls the assumption that each disease token belongs to one disease type the 'principle of completeness', and the assumption that each disease token belongs to *only* one disease type the 'principle of unambiguousness'.

true, then medicine's nosology resembles botanical and zoological classification as each instance of a plant or animal belongs to just one species.[21]

However, the assumption that each disease token belongs to one and only one disease type does not follow from the conceptual distinctness of disease types. Our disease types might not be collectively exhaustive nor mutually exclusive. First, there might be some disease tokens that cannot be classified according to any of medicine's recognized categories (thus, not every disease token belongs to one disease type, and medicine's disease categories are not collectively exhaustive). The high prevalence of 'medically unexplained physical symptoms' reported in several studies of general practice (Ring et al. 2005) might suggest that a substantial number of disease tokens are presently 'typeless'. If not all disease tokens belong to a disease type, then counting the number of disease diagnoses that apply to a patient might *under*estimate the number of disease tokens, and a patient with token multimorbidity might not have type multimorbidity.

Second, some disease tokens might belong to more than one disease type (thus, medicine's disease categories are not mutually exclusive). This logical possibility is left open particularly because many chronic diseases are understood dispositionally. Two dispositional diseases might depend on the same categorical base; or a categorical disease might serve as the base for a dispositional disease. For example, hyperaldosteronism (a dispositional disease) might give rise to hypertension (another dispositional disease), while certain cardiomyopathies in adults or congenital heart defects in newborns (categorical diseases) might give rise heart failure (a dispositional disease). Recall that philosophers disagree on whether disposition tokens are identical to their categorical bases. If token identity holds true, then it is metaphysically possible that two dispositional diseases like hyperaldosteronism and hypertension are identical to the same categorical base (and thus to each other), and that a dispositional disease like heart failure is identical to a categorical disease like a cardiomyopathy or a congenital heart defect. If some disease tokens belong to more than one disease type, then counting the number of disease diagnoses that apply to a

---

[21] Sydenham, who was perhaps the father of nosology, modeled disease classification after botanical classification (Jutel 2011). Just as there were plant species, there were species of disease ('species morbi').

61

patient might *over*estimate the number of disease tokens, and a patient with type multimorbidity might not have token multimorbidity.

The assumption that each disease diagnosis that a patient receives applies to one and only one disease token also does not follow from the conceptual distinctness of disease types. Most diseases are properties of unique parts of the body like the airways or the heart, but some diseases are properties of non-unique body parts. A patient might have joint deterioration of the left knee *and* joint deterioration of the right knee. These two properties are not numerically identical because they are discernable as properties of different parts of the body. Thus, we should say that the patient has two distinct disease tokens. Yet they only have one disease type: osteoarthritis. Once again, counting disease diagnoses might *under*estimate the number of disease tokens, and a patient with token multimorbidity might not have type multimorbidity.

In summary, type multimorbidity might generally be a reliable guide to token multimorbidity, but in some cases the number of disease types might differ from the number of disease tokens. Counting disease diagnoses might underestimate the number of disease tokens (if some tokens do not belong to recognized types and/or if some types apply to multiple tokens), and token multimorbidity might occur without type multimorbidity. In other cases, counting diagnoses might overestimate the number of disease tokens (if some tokens belong to more than one type), and type multimorbidity might occur without token multimorbidity.

We have just seen that multiple diseases can 'converge' on one another in the sense that two diagnoses can converge on the same disease token (token identity), and two disease tokens can converge on the same diagnosis (type identity). There is a distinct sense in which diseases can 'converge' but while remaining distinct: they might converge in terms of their causal relations. One disease might cause a second disease, as when CAD causes heart failure via a heart attack. Or two diseases might share a similar pathogenesis, as when atherosclerosis causes both CAD and peripheral vascular disease.[22] Finally, two diseases might share similar manifestations, as when heart failure and COPD both cause shortness of breath. Therefore, in multimorbidity when multiple diseases are distinct from one another

---

[22] John Piette and Eve Kerr (2006) call comorbidities that have a common pathogenesis or management plan 'concordant'.

their pathogenesis and manifestations might not be distinct. To further complicate matters, the treatments that doctors prescribe for patients with multimorbidity might causally affect multiple diseases (drug-disease interactions), and might have the same adverse manifestations as one of the diseases; in other words, there could be disease-treatment convergence.

The problem of causal convergence among diseases and treatments gives rise to what Ross Upshur and Shawn Tracy (2008) call 'confluent morbidity'. The name conjures the image of a culture plate in which bacterial colonies have lost their discreteness and the plate is overgrown with bacteria. As Upshur and Tracy argue, causal interactions among diseases and treatments in multimorbidity challenge the model of viewing diseases as discrete units that physicians can manage in isolation. At the same time, convergence offers promising opportunities for research and care management. Convergence in terms of token or type identity potentially reduces the number of diseases in need of managing; instead of managing two dispositions, we can manage their mutual categorical base. Meanwhile causal convergence of diseases potentially provides fewer treatment targets; instead of managing two diseases, we can manage their common pathogenesis or manifestations. Convergence might thus offer a partial fix for fractured care and treatment interactions by reducing the number of treatment plans. I will further address fractured care and convergence in Chapter 8.

To reiterate, multimorbidity is the co-occurrence of multiple chronic diseases – either disease tokens, disease types, or both. Token multimorbidity consists of multiple numerically distinct disease properties, while type multimorbidity consists of multiple conceptually distinct disease types. Although token and type multimorbidity might align with one another in general, in certain situations they might come unhinged, creating problems for measurement. Diseases that appear distinct might not always be distinct due to token or type identity. Even when disease tokens are distinct they might not be discrete in that they might causally converge. The phenomenon of convergence creates challenges as well as opportunities in the care of patients with multimorbidity.

2.6 Conclusion

Our conception of the nature of disease has evolved through time. In order to uncover the nature of our current diseases, we could utilize the standard approach in the philosophy of medicine and analyze our concept of disease. But we should worry that this procedure will fail to reveal what kind of thing diseases are. Instead, a bottom-up, inductive approach is needed. By focussing on descriptions of representative chronic diseases in the medical literature, we can make generalizations about chronic diseases, a subset of particular contemporary relevance. We find that chronic diseases are bodily states or properties. Typically they are dispositional, but sometimes they are categorical. Meanwhile, the pathogenesis of the disease is the process generating the disease property, and the clinical manifestations – signs and symptoms – are generally effects of the disease. Several alternate disease ontologies were historically popular (and some still have adherents). Despite their initial plausibility, the alternatives all fail to account for various aspects of chronic diseases or the way we understand them. To explain why chronic diseases are chronic, we cannot appeal to their intrinsic nature. Instead, it is the state of modern medicine that explains why chronic diseases endure, and why they have engulfed modern healthcare. Finally, multimorbidity is the co-occurrence of multiple distinct chronic diseases, either disease tokens/properties (token multimorbidity) or disease types/concepts (type multimorbidity). A metaphysical understanding of chronic diseases provides deeper insight into three problems of modern medicine: reductionism, multimorbidity management, and fractured care.

Chapter 3 - Causation and Disease Classification: Models of Disease Re-examined

3.1 The Causes of Disease

The public health victories of the turn of the Twentieth Century – vaccines, antibiotics, improved sanitation and public hygiene in western countries - are elevated as some of the crowning achievements of scientific medicine. They are often traced to a major theoretical and conceptual advance in the Nineteenth Century: the germ theory of disease. Robert Koch contributed his share of scientific achievements (including the discovery of the cause of several infectious diseases), and was one of the main authors of the germ theory. He summarized a strong version of the theory at a meeting of the Botanical Institute of Breslau in 1876: "each disease is caused by one particular microbe – and by one alone. Only an anthrax microbe causes anthrax; only a typhoid microbe can cause typhoid fever" (quoted in Evans 1993, 20). Though it was soon learned that several diseases like cretinism and beriberi were due to a deficiency rather than an infectious agent, the discovery of deficiencies in *particular* factors preserved and even bolstered the idea - already gaining momentum at the end of the Nineteenth Century - that each type of disease had one specific, universal cause (Carter 2003).

As the story goes, understanding the causes of disease allowed for revolutions in their treatment and prevention. In 1882, Koch presented convincing evidence to the Berlin Physiological Society that the tubercle bacillus he had isolated was the cause of tuberculosis (TB). One of the many esteemed attendees of the meeting was Paul Ehrlich, who later opined: "That evening was engraved in my memory as the most majestic event I have ever participated in" (quoted in Evans 1993, 21-22). Ehrlich went on to lead a research laboratory that synthesized the world's first effective and specific antibiotic: arsphenamine (Salvarsan), a treatment for syphilis.

Contrast Koch's simple statement about disease etiology with this one from the WHO at the beginning of the Twenty-First Century: "Clearly, the determinants of chronic conditions are complex: multifactorial" (WHO 2005, 15). While most of the Nineteenth and

early Twentieth Century epidemics were acute and infectious, many contemporary epidemics are chronic and/or noncommunicable. As the above statement from the WHO attests, the causes of each these diseases are not one but many.

Alex Broadbent (2009; 2013) calls the tradition of understanding diseases descendent from Koch the "monocausal disease model" to emphasize the privileging of one particular cause. In contrast, the model of disease popular among epidemiologists and public health authorities beginning in the second half of the Twentieth Century is a "multifactorial model" that recognizes the contribution of multiple causal risk factors to the development of each type of disease. The monocausal model is as much a model of definition as it is a model of discovery. Not only are scientists to discover specific causes of specific types of disease, they are to define specific diseases as diseases produced by specific causes. Rather than disease tokens, in this chapter I am mostly concerned with disease types or taxa.

The model of nosology or disease taxonomy that the monocausal ideal is thought to have supplanted is one in which types of disease were defined constitutively, in terms of the components that comprised them. In the early Nineteenth Century, these components were typically symptoms, but by the middle of the Nineteenth Century these components often included pathological anatomical lesions (Carter 2003). For the simple reason that different causes can give rise to the same symptoms and lesions, the causes of these diseases were not singular and universal but multiple and variant. The long and eclectic lists of etiological factors in early Nineteenth Century medicine exasperated the famous physician Jacob Henle (Koch's one-time research supervisor), who called for the discovery and reporting of causes of disease that were "universal, necessary and sufficient" (Carter 2003, 25).

According to Codell Carter (2003), Koch's "etiological standpoint", the idea that unique diseases have unique causes, launched a prolific "etiological research program" in medicine concerned with understanding diseases in terms of causes that fulfilled at least the first two of Henle's desiderata: universality and necessity. A new medical taxonomic principle was born that came to characterize modern western medicine: disease types were no longer to be defined constitutively in terms of symptoms or anatomy, but etiologically in terms of their singular causes.[23]

---

[23] Paul Thagard calls the change from one organizing taxonomic principle to another principle "tree switching" (1999, 150).

When the etiology of a disease is singular, its prevention seems simple: prevent the cause to prevent the disease. Since anthrax is always caused by the bacterium *Bacillus anthracis*, we can avert anthrax in all cases by preventing the bacterium from causing it. In comparison, as we heard from the WHO, diseases with multifactorial etiology are complex. Consequently, avoiding a particular causal risk factor for the disease may not prevent the disease. A sedentary lifestyle and high blood pressure are both causes of cardiovascular disease, but avoiding one of them will not invariably avoid cardiovascular disease, which often occurs in the absence of one or both of these risk factors. A contemporary list of risk factors for a disease – often eclectic like the Nineteenth Century lists to which Henle so objected – suggests no obvious preventive strategy. *Multifactorialism* is thus viewed as a unique challenge for modern medicine.

This attractive line of reasoning concludes that preventing chronic and noncommunicable diseases is a complex matter because they defy Koch's monocausal ideal, which revolutionized the treatment of infectious diseases a century ago and continues to guide research and therapy for infectious diseases today. Unfortunately, this line of reasoning is mistaken. In what follows, I argue that modern diseases are not quite so defiant of a monocausal nosology. But I reject the monocausal model anyway, both as descriptive of infectious diseases and as a guide to prevention. Rather, I will argue that medicine operates under a constitutional model. Because the constitutive model is an equally problematic guide, we must supplement it with a new guiding principle for prevention: the monomechanism ideal. The failure to curb the modern rise in incidence of chronic and noncommunicable diseases might result in part from failure to attain this ideal.

3.2 The Monocausal Ideal, the Multifactorial Disappointment

3.2.1 The Monocausal Ideal

Carter describes Koch's etiological standpoint as follows: "The etiological standpoint can be characterized by the belief that diseases are best controlled and understood by means of causes and, in particular, by causes that are *natural* (that is, they depend on forces of nature as opposed to the wilful transgression of moral or social norms), *universal* (that is, the same

cause is common to every instance of a given disease), and *necessary* (that is, a disease does not occur in the absence of its cause)" (2003, 1).

The first criterion, the requirement that the cause of the disease is natural, immediately suggests a strategy for *discovering* a disease's etiology: empirical research, especially research in the natural sciences. Meanwhile, the third criterion, the criterion that the cause of the disease is necessary, suggests a principle for *defining* a disease category: define the disease according to the cause that was discovered. (The third criterion implies the second criterion of universality; if a certain cause is necessary for the disease, then that cause will always occur whenever the disease occurs.) The principle of demarcating disease categories according to necessary causes is an influential ideal that we will examine in this section and the next.

It is an ideal that Carter and many other authors believe guides etiologic research and faithfully describes our paradigmatic infectious diseases. Epidemiologist Mervyn Susser argues that Nineteenth Century discoveries by Pasteur and Koch "led to the redefinition and reclassification of many disease entities [disease types] by the criterion of cause…By current definition, tuberculosis is caused by the tubercle bacillus" (1973, 23). Similarly, epidemiologist Kenneth Rothman claims: "Necessary causes are often identifiable as part of the definition of the effect. For example,…infection with the tubercle bacillus is a necessary cause for tuberculosis" (1976, 588). And philosopher Caroline Whitbeck notes that after the success of the germ theory in the Nineteenth Century "the name of the disease came to reflect the type of entity thought to cause it, the so-called etiologic agent, and etiology soon came to be definitive (i.e., to be regarded as essential) for those diseases for which it was known" (1977, 622). Carter agrees that on this model, "[c]auses are made universal and necessary by adopting suitable disease characterizations" (2003, 110).

More recently, Alex Broadbent (2009, 2013) has referred to this principle of disease classification as the "monocausal model". He too emphasizes that "[t]he special status that the monocausal model offers to certain causes is not an empirical status, but a conceptual one. Certain causes *define* the disease in question" (2013, 156). According to Broadbent's reconstruction, the monocausal model places two requirements on the defining cause: a necessity requirement ("putative cause C is a cause of every case of disease D") and a circumstantial sufficiency requirement ("given certain circumstances, which are not

sufficient to cause D, every occurrence of C causes a case of D") (2003, 150). However, I take it to be a feature of *all* causes that they are capable of completing a sufficient causal condition given certain circumstances.[24] If there are no circumstances, hypothetical or actual, in which a putative cause brings about the effect, then we have no business calling it a cause of that effect. Because all causes are circumstantially sufficient for their effect, Broadbent's second criterion is redundant; it is implied by the first criterion, which states that C is a (necessary) cause of D. A circumstantial sufficiency requirement accords no special status on the defining cause of the disease. Thus, I will only include a necessity requirement in my reconstruction of the monocausal model.[25]

> We can represent the monocausal model as follows:

> > *a* is case of disease *D only if* an *E* caused *a*.

In a case of infectious disease or poisoning, *E* refers to a specific etiologic agent (a specific germ or a specific toxin, respectively); in a disease of deficiency, it instead refers to the *absence* of a specific agent like a specific hormone or a specific nutrient. The two key features of *E* – and thus of the monocausal model - are *specificity* and *necessity*. *E* is specific because it refers to one particular kind of causal agent; it cannot refer to a disjunction of several kinds of etiologic agents, or else the disease would not be *mono*causal. *E* is necessary because *D* only occurs if *E* caused it.

> Presenting the monocausal model in the above form draws attention to its role as a model for *defining* particular disease types/taxa such as anthrax or typhoid fever. Applied to

---

[24] J.L. Mackie (1965) calls these sufficient causal conditions "*minimally* sufficient conditions" to emphasize that they contain no idle parts; were any cause missing then the remaining causes would no longer be sufficient. In Chapter 6, I call these sufficient conditions "complex causes", and characterize them in such a way to allow for the possibility of indeterministic causation.

[25] Broadbent (2009) offers a different formulation of the second requirement: "given certain circumstances, a *C*-event is not a cause of any ¬*D* event (i.e. other diseases or good health)" (303). This restriction helps to limit the number of causes satisfying the requirements of the monocausal model – ideally, to one - by ruling out any necessary causes of D that are also causes of other ¬D states. Unfortunately, it appears that infectious diseases do not adhere to such a requirement. Assuming that each type of infectious disease has as its defining cause a specific germ, the defining cause of an infectious disease is a cause of other non-disease states (for instance, it is a cause of an immune reaction in individuals with immunity). Therefore, in order to construct a model that plausibly represents infectious diseases I will also omit the (2009) version of Broadbent's second requirement.

the example of anthrax, an instance of infection (*a*) is a case of anthrax (*D*) *only if* the germ *B. anthracis* (*E*) caused the infection. As a necessary cause, *E* is a cause every instance of *D*. Although this necessity arises because *D* is defined in terms of *E*, we cannot define *D* in terms of just any factor. The factor we choose must be a cause of *D*.[26] Whether or not a particular factor is a cause of *D* is an empirical matter, to be settled through empirical research rather than by stipulation.

Although the condition that *E* caused *a* is necessary for *a* to be a case of *D*, it is not sufficient.[27] *B. anthracis* can cause many things - an immune response in those who have been vaccinated against the bacterium, the death of livestock, public hysteria. These occurrences are not thereby cases of anthrax. As a model for defining diseases, the monocausal model is incomplete, yet the constraint that it places on disease classification is mighty nonetheless. The idea that the process of defining diseases is not fully constrained by one overarching principle reflects contemporary practice. Whatever ideal model guides our disease definitions, it seems to leave room for discussion and debate (for instance, at medical consensus conferences).

3.2.2 The Multifactorial Disappointment

Many chronic and noncommunicable diseases resist the monocausal model in that they are not defined according to a privileged cause. Rather, they are recognized as diseases of multifactorial etiology. As a slogan, multifactorial diseases are those that have multiple causes, typically including genetic and environmental factors. Each case of the disease is caused by the interaction of multiple causes, and unique constellations of causes produce

---

[26] One who holds that diseases form natural kinds might further maintain that for our disease classifications to be natural we must choose the *right* causes. Whether or not diseases form natural kinds – and if so, how we go about defining diseases accordingly – is a further issue for another paper.

[27] As Broadbent (2009) notes, in the absence of any further requirement beyond necessity it is possible that more than one cause will satisfy the criteria of the monocausal model. He provides the example of oxygen. Oxygen is needed for the survival of the patient and is thus necessary for the patient's having (any) disease *D*, yet no infectious disease is defined as a disease caused by oxygen. However, as a model for defining diseases, the monocausal model need not limit the number of potential causes that could in principle be used to define the disease; rather, it need only define the disease in terms of one of these potential causes.

unique instances of the disease. High blood pressure and poor diet might interact to produce one case of stroke, while smoking and sedentary lifestyle might interact to produce another case.

So far, little distinguishes monocausal from multifactorial diseases. Monocausal diseases are also produced through causal complexes that may vary from case to case. For example, many individuals who are exposed to the tubercle bacillus develop latent TB infection but not active tuberculosis disease. Progression from latent infection to active disease is in part caused by endogenous factors that can vary even among active TB patients (Raviglione & O'Brien 2012). In fact, the only feature that distinguishes all so-called 'multifactorial diseases' from all monocausal diseases is that 'multifactorial diseases' lack a defining necessary cause. All that we can say about 'multifactorial diseases' at this point is that they are not monocausal.[28]

The fact that many chronic and noncommunicable diseases are not monocausal might seem unfortunate from a therapeutic perspective. It is tempting to award some credit to the monocausal model for historical medical marvels like vaccines and antibiotics, which target the offending pathogen. Rothman argues that discovering necessary causes can be useful: "Whereas many different component causes have been identified for several types of cancer, the hope exists for identification of a final common pathway representing a necessary cause for cancer of all types" (1976, 588). Kraupl Taylor makes the even stronger claim that "the final hope and aim of medical science is the establishment of monogenic [monocausal] disease entities" (1979, 21).

If the monocausal model is so useful, why don't scientists classify a greater number of NCDs monocausally? After all, they have discovered modifiable causal risk factors for most of them. A causal risk factor is part of a complex cause of an NCD, just as a germ is part of a complex cause of an infectious disease. Smoking is considered a 'risk factor' for cardiovascular disease (CVD) because exposure to cigarette smoke confers a certain risk or probability of disease. But exposure to *Plasmodium* or *V. cholerae* also imparts a specific

---

[28] James McCormick writes that the term 'multifactorial', when applied to etiology, is a "tautology", and that what distinguishes multifactorial from infectious diseases is that multifactorial diseases "do not have a single necessary cause" (1988, 104). Broadbent (2009) refers to this characterization of multifactorial diseases as "bare multifactorialism".

risk of disease; and whether we are dealing with smoking and CVD or *Plasmodium* and malaria the probability of disease given exposure is less than one. So while one may argue that individually causal risk factors merely contribute to disease and are not sufficient for it, the same is true of infectious agents, and prima facie that does not stop medical scientists from applying the monocausal model to infectious diseases.

A more tempting explanation for why scientists do not classify a greater number of chronic and noncommunicable diseases monocausally is that there simply are no modifiable necessary causes for most of them. Carter (2003) suggests that some diseases - perhaps cancers - may not in principle be amenable to classification by a universal cause. Whitbeck (1977) argues that for multifactorial diseases like cancers there is no indispensable environmental cause, and thus our present model of classifying diseases according to unique etiologic agents is inadequate. However, recall that the necessity of a cause under the monocausal model is produced through stipulating that a certain type of disease is – *by definition* – caused by a certain type of etiologic factor. There are no necessary, nontrivial environmental causes of NCDs because noncommunicable diseases are not classified according to environmental causes. It is not that we don't define NCDs monocausally because they have no necessary cause; rather, NCDs have no necessary cause because we don't define them monocausally.

As Broadbent (2013) argues, if we like we can simply redefine non-monocausal diseases to make them fit the monocausal model. Redefinition may be required even for an infection. Before Koch isolated the tubercle bacillus, tuberculosis was understood symptomatically and pathologically in terms of 'tubers' in the lungs. So defined, 'tuberculosis' had several infectious causes, and could occur in the absence of Koch's tubercle bacillus. It was only after Koch redefined tuberculosis in terms of the tubercle bacillus that the disease came to have a necessary cause (Carter 2003).

The example of tuberculosis illustrates that non-monocausal diseases can be made to fit the monocausal model if we are willing to shift the borders of disease categories. In redefining tuberculosis as an infection caused by the tubercle bacillus, certain states previously considered to be cases of tuberculosis were excluded from the newly reconstituted tuberculosis category. We could similarly redefine coronary artery disease (CAD) as a coronary artery stenosis caused by smoking ('CAD*'). Cigarette smoke would then become

a necessary cause of CAD*, and preventing smoking would invariably prevent CAD*. Preventing smoking would do nothing to prevent a coronary artery stenosis that falls outside of the reconstituted CAD* category, but then again preventing exposure to the tubercle bacillus will do nothing to prevent symptomatically and pathologically similar infections that fall outside of the reconstituted tuberculosis category.[29]

Redefining NCDs according to the criterion of cause may not be as tidy as redefining infectious diseases. In principle (and, prima facie, in practice), infections can be sorted into a set of causally defined categories that are generally mutually exclusive and collectively exhaustive. Because *Plasmodium* and *V. cholerae* rarely causally interact, a case of 'infection caused by *Plasmodium*' ('malaria') would rarely be a case of 'infection caused by *V. cholerae*' ('cholera'); for all intents and purposes these categories are mutually *exclusive*. Moreover, because the etiologic agent has been identified for most infections, few infections would be left unclassified; the set of causally defined infectious disease categories is practically *exhaustive*. On the other hand, there is a substantial amount of causal interaction among the causal risk factors we have identified for chronic and noncommunicable disease states. Consequently, a case of 'coronary artery stenosis caused by smoking' might also be a case of 'coronary artery stenosis caused by high LDL-cholesterol'. Furthermore, for many NCDs there exist a number of patients that lack all known causal risk factors yet have the disease (D'Agostino et al. 2008). Classifying patients according to CVD risk factors alone would leave a good number of cases of CVD undiagnosed.

Notwithstanding this taxonomic untidiness, chronic and noncommunicable diseases are not quite as resistant to the monocausal ideal as we might have thought. In order to reap

---

[29] There is a serious problem with redefining NCDs monocausally that does not apply to infectious diseases. By preventing a coronary artery stenosis caused by cigarette smoke, one would prevent a case of CAD*. However, the patient might go on to have coronary artery stenosis anyway because coronary artery stenosis is often causally overdetermined by causal risk factors. In other words, in the absence of smoking another risk factor would cause the stenosis. Because the patient quit smoking, the patient's overdetermined coronary artery stenosis is not a case of CAD*, but it is still a case of coronary artery stenosis. As a result, preventing smoking would indeed prevent all cases of CAD*, but it would prevent fewer cases of coronary artery stenosis. Infectious diseases do not encounter a similar problem. Infections are not causally overdetermined by several infectious agents; thus, preventing exposure to *B. anthracis* would prevent all cases of monocausally defined anthrax, and it would prevent an equal number of infections in general.

the apparent benefits of monocausal classification that have been realized for infections, scientists could redefine NCDs according to singular specific causes. But they don't. In the next section I will argue that – contrary to popular opinion - the monocausal model is not the ideal that guides the classification of many infections; and it is rarely a reliable guide to intervention.

3.3 The Monocausal Model Re-examined

3.3.1 The Monocausal Model as a Descriptive Pattern

Carter calls the etiological standpoint "a defining characteristic of modern western thinking about disease" (2003, 1). According to the received view, it came to define infectious diseases beginning in the late Nineteenth Century (Susser 1973; Whitbeck 1977; Stehbens 1992; Evans 1993; Carter 2003; Broadbent 2014). But if ever the monocausal model (as characterized in section 3.2.1) was descriptive of infectious nosology, it is no longer. As the following problem cases illustrate, the monocausal principle fails to classify or correctly classify many important infections, as well as several deficiencies.

The first set of problem cases consists of types of infection that are not universally caused by a single unique pathogen, and thus are not classified according to the monocausal model because they violate the requirement of specificity. Certain infectious conditions or syndromes can be caused by a great many different germs. The common cold is caused by over 100 different viruses, and pneumonia is caused by several different types of microbe (Longo 2012). While the monocausal model is usually described as a model of *diseases* (and as I argued in Chapter 2 syndromes are not diseases), we might have thought that its guidance extended to infections generally, given its purported utility for infectious diseases. Yet some of the most common infections around do not adhere to its structure.

Even tuberculosis, an infectious disease that supposedly exemplifies the monocausal model, may not be caused by a universal agent. Koch understood tuberculosis as a disease caused by the 'tubercle bacillus'. We now recognize TB as a diseased caused by the "*M. tuberculosis* complex (MTBC)" (ATS 2000; Raviglione & O'Brien 2012; Horsburgh 2015). *Mycobacterium tuberculosis* is the species most commonly responsible for tuberculosis,

especially in North America, but MTBC also includes other species of *Mycobacterium*, including *M. africanum* and *M. bovis*. It would be too nonspecific to say that tuberculosis is caused by *Mycobacterium* because certain mycobacteria cause other diseases instead; for instance, *Mycobacterium leprae* causes leprosy. Rather, MTBC consists of a subset of mycobacteria. But how can we appropriately define this subset? If we define MTBC as the mycobacterium that causes tuberculosis, then defining tuberculosis as a disease caused by MTBC would be circular. If instead we defined MTBC as a disjunction of *M. tuberculosis*, *M. africanum*, and so on, then it is not clear that MTBC refers to a *specific* etiologic agent.[30] We could just as soon define 'stroke risk factor' as the disjunction of hypertension, smoking, atrial fibrillation, and so on, and then 'stroke risk factor' would almost be a necessary cause of stroke; stroke would nearly conform to the monocausal model after all. Tuberculosis is supposed to be our paradigm monocausal disease, yet it is not clear that its universal cause refers to a specific agent. Thus, at least as a global health entity, TB may not be monocausal.

The monocausal model also runs into trouble in classifying opportunistic infections. In these cases, it *does* classify the infection according to a specific cause, but it does so incorrectly. Consider opportunistic infections in patients with HIV disease. Infections are opportunistic when they result from a compromised immune system. In patients with HIV disease, the immune system is compromised because HIV infects T Cells, which normally coordinate the immune response. One relatively common opportunistic infection in patients with HIV disease is an active respiratory infection with *M. tuberculosis*.

It seems reasonable to define 'HIV disease' according to the monocausal model: as an infection caused by HIV. Doing so would correctly classify the T Cell infection as 'HIV disease'. But what about the respiratory infection? As the respiratory infection was caused by *M. tuberculosis*, we could classify it as tuberculosis. However, the respiratory infection was also caused by HIV, which weakened the immune system and facilitated the opportunistic infection. Thus, we should also define the respiratory infection as 'HIV disease' according to

---

[30] It is not totally clear what constitutes a *specific* agent. Influenza can be caused by the influenza A virus, the influenza B virus, or the influenza C virus. Is 'the flu virus' then a specific cause? One plausible test of specificity is that we must be able to usefully define the agent without resorting to a logical disjunction for that agent to count as specific. For bacterial infections, another plausible test is that the agent must refer to a particular species of bacterium. At present, MTBC consists of multiple mycobacterial species.

our monocausal definition. Because infections with *M. tuberculosis* are unequivocally *not* cases of HIV disease, HIV disease must not be defined as an infection caused by HIV.[31] Whenever one infection weakens the immune system, leading to a second infection, applying the monocausal model risks classifying the second infection incorrectly as the same disease as the first infection. Unfortunately, HIV infection can cause a wide range of opportunistic infections, so this problem for the monocausal model is not limited to *M. tuberculosis* infections.

A related problem is described by Benjamin Smart (2014). Human papillomavirus (HPV) types 6 or 11 can cause rare infections like laryngeal papillomatosis and recurrent respiratory papillomatosis (RRP). The virus more commonly causes anogenital warts. If we define either laryngeal papillomatosis or RRP as an infection caused by HPV (type 6 or 11), then we risk misclassifying many cases of anogenital warts as one of these two rarer conditions.

Finally, the monocausal model is often thought to apply fairly well to certain deficiencies (Carter 2003; Broadbent 2009). For these disorders, we might consider the specific cause to be the lack of a particular agent – a hormone or a nutrient. The cause of hypothyroidism might be a lack of thyroid hormone, while the cause of hypocalcemia might be a lack of calcium. However, these deficiencies are better understood as the disorder itself rather than its cause (Longo 2012). On this understanding, the causes of hypothyroidism (thyroid hormone deficiency) and of hypocalcemia (calcium deficiency) are several. For instance, thyroid hormone deficiency is sometimes caused by a lack of dietary iodine, and sometimes by autoimmune destruction of the thyroid gland. Thus, at least some deficiencies are not monocausal after all.

---

[31] We could instead define HIV disease as a *T Cell* infection caused by HIV, which would not classify the opportunistic respiratory infection as HIV disease. Unfortunately, this strategy will not resolve other problem cases. A serious untreated case of pulmonary tuberculosis can also make the patient susceptible to an additional respiratory infection that they would not have otherwise contracted. Even if we define pulmonary tuberculosis as an active *respiratory* infection caused by the tubercle bacillus, we would incorrectly classify the additional respiratory infection as tuberculosis.

3.3.2 The Monocausal Model as a Guide to Intervention

We have just seen that as a descriptive pattern the monocausal model is a loose fit for many infections and deficiencies. I will now argue that as a normative ideal it is also a poor guide to intervention, including rational disease prevention. Perhaps it is not so unfortunate that infectious and non-infectious diseases alike elude the model.

Whitbeck (1977) and Carter (2003) claim that classifying a disease according to a particular etiologic agent serves medicine's interest in preventing disease. How so? "If we wish to prevent something, and we do generally wish to prevent diseases, we need only prevent or eliminate something which is a necessary condition for it," Whitbeck argues (1977, 629). Preventing a necessary cause of a disease is a logically sound procedure for preventing that disease. Without its necessary antecedent condition, the disease cannot be.

But logical necessity is not always a *strategically* sound guide to action. If your single male friend complained to you that they were tired of being a bachelor, it would be somewhat unhelpful to suggest that they stop being an unmarried man, even though being an unmarried man is a necessary condition for being a bachelor. Similarly, globally eradicating the universal infectious cause of a monocausal disease will necessarily prevent all cases of the disease, but doing so is easier said than done.

Knowing that a particular infectious agent is a necessary cause of a disease is often a poor recipe for prevention, somewhat like a recipe for eggs benedict that merely listed eggs as a necessary ingredient. Perhaps the tubercle bacillus is transmitted only through physical contact with infected individuals, in which case one need only avoid physical contact. In reality, tuberculosis typically results from inhalation of the microorganism in aerosolized fluid droplets produced by infected patients. Knowing this detail about *how* the etiologic agent causes tuberculosis, the mechanism of transmission, suggests a better strategy: we can prevent the spread of TB by limiting exposure of susceptible individuals to active TB patients and by providing TB patients or their susceptible contacts with masks. Similarly, knowing that the *Plasmodium* parasite causes malaria or that the bacterium *V. cholerae* causes cholera does not imply a specific course of action. Only knowing that *Plasmodium* is usually transmitted in the bites of mosquitos (the disease vector) suggests the use of mosquito nets or insecticide; only knowing that *V. cholerae* is usually transmitted in drinking

water suggests the practice of sterilizing the water. Indeed, before cholera's mode of transmission was understood, authorities tried quarantining cholera sufferers to keep them away from the vulnerable, but failed spectacularly to prevent the spread of the disease (Ackerknecht 1982, 175).

On the side of chronic and noncommunicable diseases, simply knowing that cigarette smoke causes cardiovascular disease, lung cancer and chronic respiratory diseases is no recipe for prevention. Even if we redefined these diseases as conditions caused by smoking (like we attempted with CAD* in section 3.2.2), making cigarette smoke a necessary cause suggests only that if we threw out all of the cigarettes and destroyed all of the tobacco plants then we could evade these conditions. Such an approach would not win support from cigarette companies or lawmakers. We need a different approach; we first need to know *how* people come to start smoking in the first place, and how this behaviour is maintained through addiction and other mediating forces. Whether the target is an infectious disease or a chronic, noncommunicable one, a necessarily etiologic factor is not necessarily a guide to prevention.

We began our story with the hypothesis that Koch's etiological standpoint, the monocausal disease model, facilitated revolutions in the prevention and treatment of infectious diseases around the turn of the Twentieth Century. Referring to the etiological standpoint, Carter argues: "Of the numerous changes that have occurred in medical thinking over the last two centuries, none have been more consequential" (2003, 1). Yet as we will now see, a monocausal conceptual framework of the kind I have described could not have suggested many of the medical advances with which we might want to credit it.

Much of the impressive gain in life expectancy in western countries that was won over the first half of the Twentieth Century can be attributed to public health initiatives: water filtration and disinfection, food sanitation, milk pasteurization, sanitary sewage disposal, the cleaning up of factories and slums, relieving overcrowding in community housing, isolating the infectious from the healthy, and educating the public about personal hygiene (Ackerknecht 1982; Duffin 2010). Many of these interventions began in Britain in the middle of the 1800s, before Koch's etiological standpoint. While the discoveries of microbiology late in the century may have accelerated the sanitarian movement in western countries, it is difficult to credit the monocausal model in particular. These public health improvements do not discriminate among specific species of microorganism; rather, they are

each successful for preventing the transmission of many different types of germs and diseases. The idea that specific diseases are caused by specific germs is not directly relevant.

The concept of disease-pathogen specificity *is* seemingly relevant to the creation of vaccines, which exploded at the end of the 1800s with the development of vaccines for anthrax and rabies in the 1880s, and for diphtheria, typhoid fever, cholera and plague in the 1890s (Ackerknecht, 1982). Although these were not the first vaccines (a smallpox vaccine was discovered one hundred years earlier, largely through observation and analogy), they required the isolation of specific germs for specific maladies. If (i) long-term immunity can be generated against particular pathogens through inoculation and (ii) particular pathogens are necessary for particular diseases (the etiological standpoint), it follows that long-term immunity can be generated against particular diseases through inoculation. Note however that the etiological standpoint alone does not suggest this conclusion. Vaccine development also requires an understanding of the principles of immunity and inoculation. In fact, most vaccine research can ignore the concept of disease-pathogen specificity and focus on the identification and production of antigens that will trigger immunity to a specific pathogen. Rather than a recipe for prevention, the etiologic standpoint is better conceived as a useful organizing principle for vaccine research, the beginning of a potentially fruitful disease-centered research program.

What about antibiotics, magic bullets for infectious diseases? The monocausal model implies that without its defining bacterium, the disease will not occur. I will first remark that strictly speaking this suggests a strategy of getting rid the bacterium rather than one of killing it with an antibiotic. We need the knowledge that only the *live* bacterium is pathogenic before the strategy of killing the bacterium suggests itself.[32] Even with this knowledge, rational prevention demands that we know *how* to kill the bacterium as only certain antibiotics will kill a specific species of bacterium.

How were some of the earliest Nineteenth Century antibiotics discovered? The first effective and specific antibiotic, arsphenamine (1910), was named Salvarsan-606 because it

---

[32] Some bacteria, including the species that cause tetanus and botulism, bring about the disease by producing a toxin. It is not only the live bacterium that is pathogenic but also the toxin. The concept that *Clostridium tetani* is the cause of tetanus does not suggest antitoxin therapy as an intervention.

was the 606th compound that Ehrlich's team tried against the syphilis bacterium; and in 1928 when Alexander Fleming famously (re)discovered that a compound produced by the *Penicillium* mould kills certain bacteria, it was an accidental finding (Duffin 2010). The discoveries of Salvarsan-606 and penicillin were not guided by a rational recommendation of the monocausal disease model but were the result of trial-and-error and of a chance finding, respectively. One might argue that the monocausal model allowed physicians to *apply* the discovery of an antibiotic like arsphenamine to the control of an infectious disease like syphilis. But the monocausal model only offers the insight that if we stop the bacterium from causing the disease we will prevent the disease. How do we stop the bacterium? By washing our hands with antibiotic? We need to know that syphilis *consists in* bodily infection with the syphilis bacterium rather than the fact that is *caused by* the syphilis bacterium before we can think of giving patients the antibiotic to ingest.

The fundamental idea that killing bacteria will cure certain contagious diseases and might prevent further cases of disease may recommend the general principle of antibiotic treatment. But the monocausal model was not the theoretical advance which revealed that bacteria cause a range of diseases. It is some version of the germ theory or bacterial theory that suggests such a general hypothesis. Thus, we cannot credit the monocausal model with the universal rationale behind antibiotics either.[33]

The monocausal disease model is a suboptimal guide to therapy and prevention because a universal etiologic factor does not necessarily recommend a practical interventional strategy. For instance, the monocausal model does not directly suggest many of the medical breakthroughs in infection control that occurred in the late Nineteenth and early Twentieth centuries. It could hypothetically serve as a useful organizing principle for

---

[33] At the outset, I referred to the etiological standpoint as a strong version of the germ theory: the theoretical *and conceptual* claim that "each disease is caused by one particular microbe" (Evans 1993, 20). A weaker version of the germ theory might consist in an interconnected set of chiefly *theoretical* statements about the infectious pathogenesis and nature of diseases, statements that do not serve as a model for defining diseases. These statements might include the claim that (some or all) diseases consist in an infection with a microscopic germ, and the claim that these diseases are caused by the transmission of the germ from one host to another. I would speculate that Koch's great contribution to innovations like turn-of-the-Century improvements in sanitation and the development of antibiotics was actually the weaker version of germ theory.

research (e.g. research into vaccines). Yet we saw in section 3.2.2 that the monocausal model fails to describe chronic and noncommunicable diseases, and in 3.3.1 that it fails to describe many important infections. So it is not in fact a universal organizing principle.

What model *does* describe modern disease taxonomy?

3.4 The Constitutive Model and the Monomechanism Ideal

3.4.1 The Constitutive Model

Let's start by discussing the classification of chronic and noncommunicable diseases. Recall that for their stubborn defiance of any kind of single cause model, chronic and noncommunicable diseases are often labeled as 'multifactorial'. In the early-mid 1800s, diseases were similarly recognized as having many diverse causes, none of which were universally responsible for the disease. I explained that the reason for this multifactorialism was that diseases were defined constitutively, in terms of their component symptoms or pathological anatomical structures, and that several unique causal complexes can give rise to the same symptoms and anatomy. How can we explain modern day multifactorialism?

Like Nineteenth Century maladies, chronic and noncommunicable diseases are classified according to their constitution - what the disease is. What distinguishes one type of chronic disease from another is not a specific etiologic agent but the nature of the disorder. Two instances of chronic disease can have a similar set of etiologic factors. Sedentary lifestyle and poor diet might have caused Professor Plum's coronary artery disease *and* his peripheral vascular disease (PVD). What distinguishes his CAD from his PVD is that his CAD consists of stenosis in a coronary artery supplying the heart, while his PVD consists of stenosis in a peripheral artery supplying the leg.

As another example, recall that patients with diabetes mellitus (DM) are grouped together by virtue of having a disorder of glucose metabolism that disposes them to hyperglycemia. Diabetes mellitus *is* this disposition. As a typical medical textbook notes, diabetes is paradigmatically multifactorial ("[s]everal distinct types of DM are caused by a complex interaction of genetics and environmental factors" (Powers 2012)). The exact causal

factors implicated determine whether the patient is a type I diabetic, a type II diabetic, or a diabetic of another subtype, and are incompletely understood.

Similarly, recall that COPD is often defined as "a disease state characterized by airflow limitation that is not fully reversible" (Reilly Jr. et al. 2012). The etiology of COPD is variable; in some but not all cases, cigarette smoke is an important etiologic factor. Meanwhile, osteoporosis can be defined as "a bone density that falls 2.5 standard deviations (SD) below the mean for young healthy adults of the same sex" (Lindsay & Cosman 2012). A diagnosis of osteoporosis automatically applies whenever this criterion is satisfied. These examples illustrate a typical contemporary pattern: chronic and noncommunicable conditions are classified according to what the condition is (e.g. a state giving rise to airflow limitation, a low bone density, and so on). I will call this commonsensical taxonomic principle the *constitutive disease model*.

As I argued in Chapter 2, chronic diseases are bodily properties. Some of these properties are manifest (e.g. low bone density), while others are dispositional (e.g. a disposition towards hyperglycemia).[34] What distinguishes one type of chronic disease from another – osteoporosis from DM - is the particular property that is picked out.

For diseases like DM, COPD and osteoporosis, the constitutive criterion is both necessary and sufficient. We can thus represent the constitutive model as follows:

a is a case of disease *D* if and only if *a* is a *C*.

Applied to the example of osteoporosis, a bodily state (*a*) is a case of osteoporosis (*D*) if and only if *a* is a bone density that falls 2.5 SD below the demographic-adjusted mean (*C*). The core characteristic of *C* – and thus of the constitutive model - is *necessity*. *C* need not refer to a specific entity, but - whatever *C* includes - *D* only ever occurs when *C* occurs. When the constitutive principle is both necessary and sufficient, it might appear to constrain all taxonomic decisions. However, there is still an important choice of which property or properties to designate as *C*; whether, for instance, *C* is to represent a bone density that falls 2.5 SD or 2 SD below the mean, or whether it is to represent some other determinant of bone strength.

---

[34] Because dispositions are defined in terms of their manifestations (e.g. as 'a disposition towards hyperglycemia'), dispositional chronic diseases are not distinguished according to a specific cause (as the monocausal model would have it), but according to a specific effect.

Because chronic and noncommunicable diseases are defined constitutively, it should come as no surprise that they are multifactorial in their etiology. I earlier remarked that there is an important sense in which infectious diseases are also multifactorial. Although each infectious disease typically has a universal cause, different causal complexes can cause the same type of infectious disease. Perhaps infectious conditions are multifactorial for precisely the same reason that NCDs are multifactorial: they are defined constitutively. To explore this possibility, let us see if the constitutive model fares better than the monocausal model in classifying the problem cases from section 3.3.1.

The first collection of problem cases were infections like the common cold and pneumonia that are not caused by the same pathogen in all instances, and thus must not be defined in terms of any one pathogen. Instead, symptomatic viral infections of the upper respiratory tract are classified as 'the common cold', and symptomatic infections of the alveoli of the lungs are classified as 'pneumonia'. The common cold and pneumonia are defined according to what they consist in: a symptomatic infection of a particular part of the airway.[35] Similarly, as tuberculosis consists in an active infection with any one of the MTBC bacteria (*M. tuberculosis*, *M. africanum*, and so on), the simplest way to group these infections together as instances of 'tuberculosis' is to define the disease constitutively as an active infection with any one of these mycobacteria. In none of these examples does the definition refer to a specific pathogen; but that is not a problem because - unlike the monocausal model – specificity is not a constraint on the constitutive model.

As long as we are relaxing the requirement of having a specific cause, one might think that we can rescue the monocausal model by defining a particular infectious disease as an infection caused by a prescribed *set* of pathogens. Unfortunately, as the problem of the opportunistic infection demonstrated, this solution will not address all of the monocausal model's shortcomings. Pathogens like HIV can weaken the immune system, causing a second, opportunistic infection. We cannot define HIV disease as an infection caused by HIV because then the opportunistic infection might satisfy the definition for HIV disease, even if

---

[35] Rather than an infection giving rise to certain symptoms, 'the common cold' and 'pneumonia' can refer to the symptoms (syndrome) caused by the infection. In this case, the common cold and pneumonia are still defined constitutively but as a particular cluster of signs and symptoms.

it is really an instance of some other type of infectious disease. If we instead define infectious diseases constitutively, we do not run into this problem. An opportunistic infection with the influenza virus that was caused by HIV is *not* HIV disease because HIV disease is defined as an infection with HIV.

Despite the hypothetical possibility of misclassification, it does seem that at least sometimes infectious diseases are defined monocausally. For instance, *Harrison's Principles of Internal Medicine* (Longo 2012) defines influenza as "an acute respiratory illness caused by infection with influenza viruses", and cholera as "an acute diarrheal disease…caused by *V. cholerae*". Yet as the problem of the opportunistic infection makes visible, monocausal classification risks misclassification. On a given occasion, the flu virus or cholera bacterium could be part of the complex causal history of any number of diseases. Thus, perhaps monocausal descriptions are best viewed as rough characterizations rather than as formal definitions. The practice of loosely defining infectious diseases monocausally might result from conceptual conflation of the cause of an infectious disease with the infectious disease's constitution (to be discussed momentarily).

Finally, recall that deficiencies like hypothyroidism and hypocalcemia are sometimes thought to be monocausal, caused by the lack of a specific factor such as thyroid hormone or calcium. Instead, the deficiency *is* the lack of the specific factor. While hypothyroidism is indeed defined in terms of thyroid hormone deficiency, it is defined constitutively *as* thyroid hormone deficiency rather than as a disorder *caused by* thyroid hormone deficiency. The constitutive model is more consistent with current disease classification than the monocausal model, even for infections and deficiencies.[36]

The constitutive model is the standard model of classification for diseases of all sorts – infectious, noncommunicable, acute and chronic. Although infectious diseases often *do*

---

[36] We might wonder whether the monocausal model as I have described it was ever descriptive of infections and deficiencies, or whether the constitutive model ruled even around the turn of the Twentieth Century. If indeed the constitutive model continued to operate throughout the microbiological revolution, then the dramatic shift in disease classification was not a change from an earlier constitutive model to a monocausal model, but from the constitutive classification of diseases in terms of the observable (symptoms and anatomy) to the constitutive classification of diseases in terms of the theoretical and unobservable (microbial infections and molecular deficiencies). Some of Carter's (2003) remarks suggest that he might actually agree with my proposal.

have a nontrivial universal cause (a particular pathogen), this universality does not arise because we define infectious diseases monocausally. It arises because we define the disease as an infection with the specific pathogen, and it is a contingent fact about infections that they always arise through transmission of the pathogen from without (I say 'contingent' because in another possible world infections might arise through spontaneous generation of the pathogen in vivo). When we say that the cause of anthrax is the bacterium *B. anthracis*, we most plausibly mean that the cause of anthrax is transmission of *B. anthracis* to the host. But under the constitutive model it is not *analytically* true that anthrax, defined as infection with *B. anthracis*, always results from transmission of the bacterium from the environment. The statement that anthrax is always caused by transmission of the bacterium from the environment is a synthetic fact about the etiology of anthrax, an instantiation of a more general law that was discovered by the Nineteenth Century germ theorists: infections arise through the transmission of a microscopic germ from environment to host (Louis Pasteur called this idea the "hypothesis of the dissemination of germs" (Carter 2003, 66)).

In the previous section, I noted that the monocausal model might sometimes be a useful organizing principle for infectious disease research, including vaccine development, owing to its disease-pathogen specificity. Fortunately, the constitutive model is often applied in such a way that it also generates disease-pathogen specificity; namely, by defining an infectious disease as an infection with a specific pathogen. *Un*fortunately, the constitutive model fares no better than the monocausal model as a guide to intervention. If a model that merely lists a universal etiologic agent is a poor recipe for prevention, then a model that is completely silent on a disease's etiology is no recipe at all. In the territory of prevention, we are even more lost under the guidance of the constitutive model than under the monocausal model. In order to make sense of rational disease prevention, we need a separate principle.

3.4.2 The Monomechanism Ideal

We began our tour through models of disease classification with a tempting thought, the intuition that developing successful strategies for the prevention of chronic and noncommunicable diseases is often trickier than coming up with interventions to prevent infectious diseases because while infectious diseases are defined monocausally, chronic and

noncommunicable diseases are not. I have since argued that both infectious diseases and chronic/noncommunicable diseases are actually defined according to a constitutive model. Because the constitutive model is silent on disease etiology it is no guide to rational disease prevention.

It is a synthetic fact that many infectious diseases *do* have a universal infectious cause, but I earlier argued that merely knowing about a universal etiologic agent is also typically a poor guide to disease prevention. On the other hand, acknowledging that a disease is multifactorial in its etiology is in one sense trivial (because all diseases are multifactorial), and it is equally unhelpful from a preventive standpoint. In order to avert malaria or cholera, lung cancer or COPD, it is not enough to identify an etiologic agent – a microbe or a cigarette – or to point out that multiple causal risk factors interact to produce the disease. We need to know *how* various etiologic factors interact to cause the disease; in other words, the *pathogenesis* or *etiologic mechanism*.

There has been a recent flurry of intellectual interest in mechanisms in the philosophy of science, and there are many definitions of a mechanism to go around (e.g. Machamer et al. 2000; Glennan 2002; Bechtel and Abrahamsen 2005). Phyllis Illari and Jon Williamson propose a definition meant to capture the essential elements of other popular accounts: "A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon" (2012, 123). Often the first stage in the discovery of a mechanism is the identification of a phenomenon in need of explaining, in this case a type of disease. As Stuart Glennan (1996) argues, a mechanism is functionally individuated, meaning that we identify a particular mechanism as 'the mechanism that produces phenomenon *D*'. The constitutive model has a crucial role to play in this regard because it defines the disease phenomenon D in need of etiologic explanation.

'The pathogenesis of pulmonary tuberculosis' consists in the entities and activities that (organized in such a way) produce pulmonary tuberculosis in every instance. There might of course be etiologic differences between two individual cases of tuberculosis. The pathogenesis of TB, as a type of disease, refers to those entities and activities that are common to all cases of TB. For pulmonary tuberculosis, the entities would consist in the tubercle bacillus, mucus droplets, and a susceptible individual's alveoli, among other components. The activities would consist in an infected individual sneezing or coughing, the

aerosolizing of fluid droplets containing the tubercle bacillus, the inhalation of these fluid droplets, their deposition in the alveoli, and the activation of the bacteria, among other events. Finally, the mechanism's organization would consist of the intimate spatial relationship between the infected individual and the susceptible individual, the temporal order of the pathogenic events, and so on.

When an etiologic mechanism contains a specific pathogen, this agent is but one entity in the mechanism. In order to prevent an infectious disease, it is often beneficial to intervene on other entities, or on the activities or organization of the mechanism. To prevent the spread of pulmonary tuberculosis, we can kill the bug in infected individuals (including those with latent infection), which requires intervening on the mechanisms of bacterial growth and survival. Or we can intervene on the aerosolizing of fluid droplets by providing masks to patients with active tuberculosis or instructing them to cover their mouth when they sneeze or cough. Or we can intervene on the spatial organization of the mechanism by preventing susceptible individuals from coming in close contact with active TB patients. Knowing the etiologic mechanism provides several potential targets for intervention. Intervening on any one of these targets is not guaranteed to produce the intended outcome, but knowing an etiologic mechanism can be a very good guide to prevention. In the next Chapter, I provide an account of the main reasons why mechanistic predictions fail.

Knowledge of a particular mechanism is embodied in a 'mechanistic model', 'mechanical model' or 'model of a mechanism' (Glennan 2002; Bechtel and Abrahamson 2005; Darden 2008). There is more than one way to represent a particular mechanism, even within the same representational modality. One diagram might represent the pathogenesis of thromboembolic ischemic stroke fairly abstractly as follows: blood clot obstructing intracranial artery → thromboembolic ischemic stroke. Knowing this mechanism of ischemic stroke suggests that we might be able to mitigate the disease by busting up the blood clot. Another diagram might represent the upstream causal risk factors that contributed to the eventual formation of the blood clot (suggesting that we might prevent certain cases of stroke by avoiding certain risk factors), while a third diagram might go into much molecular detail about clotting factors and cellular mechanisms of brain tissue death (indicating that we can prevent the formation of blood clots by inhibiting certain clotting factors). Our representational choices influence the representational scope of the mechanistic model. The

fairly abstract representation above could serve as a general model for thromboembolic ischemic stroke because its elements are shared by all cases of the disease. If we include certain causal risk factors in the mechanistic model, we restrict the scope of the model to those cases of thromboembolic stroke in which those risk factors are indeed a factor.

We have already seen that as a guide to intervention the monocausal ideal is less than ideal. Instead, rational disease prevention ought to be guided by a *monomechanism ideal*. Rather than knowledge of one universal etiologic factor, the monomechanism ideal consists of sufficient knowledge of one universal etiologic mechanism for the disease type. Restricting the number of mechanisms to one is justified because if there are multiple mechanisms responsible for the disease type, then rationally intervening to prevent the disease might successfully interrupt only some of the mechanisms responsible for only some of the instances of that disease. The monomechanism ideal is a requirement on our knowledge of the mechanism; in other words, on the mechanistic model. However, enacting the ideal requires cooperation between our mechanistic model and the worldly mechanism it represents. The ideal requires that we adequately represent the etiologic mechanism using only one mechanistic model, but it also necessitates that there *really is* only one etiologic mechanism to represent.

Notice that the monomechanism ideal demands that we have *sufficient* knowledge of the mechanism, an *adequate* mechanistic model. We might have one mechanistic model representing one etiologic mechanism, but it will still serve as a poor guide to rational prevention if it represents the mechanism in a suboptimal way. Criteria for the ideal mechanistic model are bound to vary somewhat according to context, but we can envision several universal desiderata. These desiderata would include at least the following: the model must have sufficient scope (the *one mechanism requirement*), it must be accurate enough (the *correspondence requirement*), it must be complete enough (the *completeness requirement*), and it must include the right levels of organization (the *right levels requirement*).

When we struggle to design a preventative intervention for a chronic or noncommunicable disease, it is not because the disease is multifactorial or because it eludes the monocausal ideal. Rather, it may be that we have failed to obtain the monomechanism ideal. A mechanistic model falls short of the ideal when it fails to satisfy one or more of the above representational requirements. We might even be able to trace some of the difficulties

in modern disease prevention to mechanistic models that resist particular requirements of the monomechanism ideal.

The first and central requirement of the ideal, the one mechanism requirement, demands that we have one mechanistic model representing one etiologic mechanism for one particular type of disease. For many infectious diseases, this one-to-one-to-one mapping falls neatly into place as a consequence of how the disease is defined under the constitutive model. When the disease is defined as an infection with a specific pathogen it is thereby associated with a pathogen-specific etiologic mechanism, which we can (and do) represent with a textbook model. Even infections like the common cold that are not defined based on a specific pathogen do tend to share a similar sequence of etiologic events. In order to represent a single etiologic mechanism, we abstract away the details of the particular type of virus. Once we have a singular mechanistic model, we might then be able to propose a preventative intervention for the disease.

Oftentimes when we use the constitutive model to construct a chronic or noncommunicable disease type, a unitary mechanism will not line up dutifully behind it. For example, diabetes mellitus can be defined as a disposition towards hyperglycemia. Several distinct mechanisms can generate this disposition. In some cases it is produced by the autoimmune destruction of pancreatic beta cells, leading to insulin insufficiency. In other cases it is produced by a largely unknown mechanism leading to insulin resistance and insulin insufficiency. It may not be feasible to represent these two very different etiologic mechanisms using one model. Instead, we represent the pathogenesis of DM using two models, and the monomechanism ideal eludes DM.

When we define a chronic or noncommunicable disease constitutively, its pathogenesis might consist of multiple mechanisms (epidemiologists Rothman and Greenland (2005) refer to the existence of multiple disease mechanisms as 'multicausality'). If the worldly mechanisms do not cooperate, then the one mechanism requirement is out of reach. Thankfully, we can force the mechanisms to segregate into groups that correspond to disease categories by defining disease categories according to etiologic mechanisms, an approach resembling the monocausal model. For diabetes mellitus, we do in fact adopt this procedure. According to *Harrison's* (Powers 2012), "DM is classified on the basis of the pathogenic process that leads to hyperglycemia". The first mechanistic model - in which the

disposition towards hyperglycemia is due to autoimmune destruction of beta cells - defines type I diabetes, while the second model - in which hyperglycemia is due to insulin resistance and insulin insufficiency - defines type II diabetes. Doctors cannot prevent type I diabetes at present, but defining a pathogenic subtype allows researchers to study the etiologic mechanism in greater detail, and may yet reveal a new target on which to intervene.

Stroke, another NCD, is similarly subdivided into ischemic stroke and hemorrhagic stroke based on the etiologic mechanism (reduced blood flow in the local vasculature vs. bleeding from the local vasculature). Diseases like type I diabetes and ischemic stroke are defined using the constitutive model (as a disposition towards hyperglycemia and as an abrupt onset neurological deficit, respectively). However, for these two diseases the constitutive requirement is not sufficient, it is supplemented with the one mechanism requirement; together, the two principles define the disease.[37] Other diseases are classified solely using the constitutive model and may - for all we know - fail the one mechanism requirement. For example, over 80% of cases of hypertension can be classified as 'essential hypertension', in which the pathogenesis is unknown (Kotchen 2012).

We always have a choice of how much mechanism detail we want to represent in our model, and different models of varying detail will suit different purposes. One of these purposes is defining a disease category etiologically. By adding more detail into a mechanistic model, we can give the model a lesser scope and individuate an etiologically defined disease that is thereby rarer. Or by abstracting away detail from a model,[38] we can give it a greater scope and individuate a more common disease. Jeremy Simon (2008) illustrates this phenomenon with the example of cystic fibrosis. This chronic and noncommunicable disease involves dysfunction in a chloride ion transport system and is typically caused by a mutation in a particular gene. Simon argues that were we to encounter patients with the same dysfunction in the transport system but due to a different cause (say, an exogenous toxin), we would have a choice to make. We could represent two groups of patients separately using two different models, one including the gene and the other

---

[37] Brendan Clarke (2011) argues for a system of melanoma sub-classification according to mutated genes that play a role in a proposed etiologic mechanism for the tumour. He argues that "thinking about causes in this way suggests the means of intervening on the tumour – perhaps by modifying or blocking part of the pathological mechanism" (27).

[38] This kind of abstraction is sometimes called 'Aristotelian idealization' (Cartwright 1989).

including the toxin; or we could represent them with the same model and omit detail of the upstream cause. We could then decide whether we have one constitutively defined disease (cystic fibrosis), two etiologically defined diseases ('genetic chloride transport disease' and 'toxic chloride transport disease'), or – as with DM and stroke – two etiologically defined diseases nested within a third constitutively defined disease (cystic fibrosis, 'type I cystic fibrosis' and 'type II cystic fibrosis').

Another reason to abstract away or specify (de-abstract) detail from our model is that it can modify the scope of a hypothetical intervention. When we remove pathogen-related details but leave details of the pathogen's mode of transmission, mechanisms suddenly congregate into groups that transcend disease boundaries. The pathogenesis of waterborne diseases now falls under the scope of one model, and the pathogenesis of sexually transmitted diseases falls under the scope of another model.[39] We have increased not only the scope of the mechanistic model but also the scope of a hypothetical intervention on the model because preventing the waterborne transmission of pathogens (for example) will hypothetically interrupt the transmission of many types of disease. Thus, abstracting detail can allow us to visualize the potential impact of a wide-reaching intervention. Germ theory, as a highly abstract model of the pathogenesis of all infectious diseases, suggests that public health interventions that prevent the transmission of germs through many routes will have an enormous collective impact on population health. The hope of atherosclerosis research is to develop a mechanistic model representing the shared pathogenesis of many cardiovascular diseases, the equivalent of germ theory for the world of cardiovascular disease.

Broadbent (2009) criticizes multifactorial epidemiological thinking because it fails to provide general explanations for particular types of disease. The monomechanism ideal could fill this void. But Broadbent argues, "mechanistic explanation is neither necessary nor sufficient for the sort of general understanding I am advocating" (2009, 307); not necessary because general explanations are not always mechanistic, and not sufficient because mechanistic explanations are not always general. I agree with Broadbent that mechanistic

---

[39] When we abstract away enough detail from our model to capture the pathogenesis of multiple diseases, we do not violate the one mechanism requirement. The requirement asks that we represent the pathogenesis of a disease with one model; it does not require that we represent the pathogenesis of only one disease with each model.

explanations are not always necessary when explaining disease. But concerning sufficiency, the monomechanism requirement demands that the mechanism we have identified produces or explains every case of the disease in question; thus, the monomechanism ideal always achieves generality.

In answer to monocausal and multifactorial models of disease, Broadbent (2009; 2013; 2014) proposes a Contrastive Model. In the Contrastive Model, a disease D is defined through a contrast between *cases* and *controls*. The cases have certain symptoms (2009; 2013) or certain deviations from normal functional ability (2014) that are caused by a certain set of causes, while the controls lack the symptoms/deviations as well as at least one cause from among the set. While the monomechanism ideal is primarily an ideal of rational disease prevention and is for that reason not in direct competition with the Contrastive Model (which is model of general disease explanation), we can wonder how both models would fare as explanatory and preventive aids. The Contrastive Model and the monomechanism ideal are both capable of delivering general explanations for the onset of particular diseases; the former provides contrastive causal explanations (Broadbent, 2009), while the latter provides mechanistic explanations. However, the Contrastive Model is generally ill suited for enabling rational prevention because like the monocausal model it merely lists responsible causes without telling us how those causes produce the disease.

Smart proposes a 'causal classification of diseases (CCD)' in which "a disease is identified by its full-cause, that is, the disjunction of conjunctions of events jointly sufficient for contracting the disease" (2014, 257). Smart's model might fare better than the Contrastive Model as an aid to rational prevention because it contains details about how causes interact, at least as concerns their logical interactions. But a list of sufficient conditions is less helpful than a mechanistic model for two reasons. First, sufficient conditions usually contain a great many factors, including standing conditions and negating factors, not all of which will be relevant to prevention. Even if we left out some of these factors from our description of the full cause, there is a remaining problem: different sufficient causes might represent different pathogenic processes requiring different intervention strategies; if we lump these sufficient conditions together in the definition of a disease we might then run into the problem of multicausality that seems to complicate the prevention of our current chronic and noncommunicable diseases. To avoid this problem, we

would have to include only those sufficient conditions that represent a single mechanistic process. But then we might as well represent this process using a mechanistic model, which embodies greater detail, including the organization of the mechanism as well as *physical* interactions among causal components.

The one mechanism requirement is not the only requirement of the monomechanism ideal. Other desiderata for the ideal pathogenic model include a correspondence requirement, a completeness requirement, and a right levels requirement. While I cannot discuss these requirements in exhaustive detail here, I will mention a few important points.

The correspondence requirement asks that the model is accurate enough to suggest a successful intervention. A mechanistic model can plausibly be good model for a phenomenon in certain respects without being accurate; for instance, it might render the phenomenon intelligible and/or might be empirically adequate. But in order to be a good guide to intervention, there must *at least* be a correspondence between an element in our model on which we wish to hypothetically intervene and an element of the worldly mechanism on which we can actually intervene. Our theoretical intervention must correspond to some real intervention if we are to have any hope of preventing a real disease.

The completeness requirement asks that the model contain enough detail to recommend a successful intervention.[40] If our model is a mechanism sketch (Machamer et al. 2000), containing gaps between mechanism stages, those gaps should not fall in places where we wish to intervene. Nor should these gaps cover components of the mechanism that must be present or absent for our intervention to be successful. Even a mechanistic model that does not contain gaps between stages can fail to contain enough detail to facilitate rational intervention. To use Craver and Darden's terminology (2013), the mechanism schema might be 'superficial', representing the mechanism too abstractly. If the mechanism is represented too abstractly, we might not know how to intervene in the concrete world or what other concrete conditions must be satisfied for our intervention to work.

Failures to meet one or more of the above requirements – one mechanism, correspondence, completeness – might explain why therapeutics have been largely unable to stop the rising tide of multifactorial chronic and noncommunicable diseases. If our

---

[40] Jeremy Howick (2011b) imposes a completeness requirement on high-quality mechanistic reasoning.

pathogenic models represent only some of the mechanisms producing a given type of disease, or if they correspond poorly to the worldly mechanisms, or if they contain relevant gaps, then they might serve as a poor guide to therapeutic intervention.

Lastly, the right levels requirement asks that our mechanistic model contain the relevant levels of organization. Craver and Darden understand the concept of mechanism levels in terms of a "decomposition hierarchy" in which the 'lower levels' describe the entities and activities that compose larger entities and their activities at the 'higher levels' (2013, 21). A mechanism at the 'organ level' is composed of cellular mechanisms (the 'cellular level'), which are themselves composed of molecular mechanisms (the 'molecular level'). When mechanisms at various levels are known, it is sometimes more feasible or effective to intervene at one level compared to another.

When we say that chronic and noncommunicable diseases are 'multifactorial', we often add that they are caused by a myriad of genetic and environmental factors. Genes are involved in molecular mechanisms, and try as we might to interfere in processes like atherosclerosis with therapeutic molecules – drugs – we are often unsuccessful or only marginally successful. On the other hand, the kinds of environmental factors we often have in mind are causal risk factors, including modifiable behaviours like smoking and exercise, as well as social circumstances such as access to education and poverty. In recognizing that the etiology of chronic conditions is complex, the WHO points out that the determinants of chronic disease are "multifactorial *and multisectoral*" (2005, 15; my emphasis), and include behaviours and policies. The WHO also recognizes that 'social determinants of health' like literacy and employment contribute to health and disease as much as biological determinants do.

Medical science is at the stage of recognizing the existence of important social and behavioural etiologic factors, but as I have argued (several times now), etiologic factors are not enough. In order to prevent diseases by intervening at the personal and interpersonal levels, we must understand how social and behavioural factors participate in etiologic mechanisms. Given the important role that these variables are known to play in the pathogenesis of chronic and noncommunicable diseases, and considering that many of our historically most successful medical interventions have been public health initiatives, intervening at the behavioural and social level(s) may be the most feasible and effective

strategy for preventing modern multifactorial disease. A failure to satisfy the right levels requirement might explain why public health has struggled to prevent chronic and noncommunicable diseases. Successful public health interventions like health promotion activities and healthy policies benefit from a rich understanding of social-behavioural mechanisms; such an understanding is largely absent from medicine.

Intervening on social and behavioural variables might sometimes necessitate representing the etiology of disease at the social or behavioural level, but we need not limit ourselves to models that represent mechanisms *only* at the social and behavioural levels. The best guide to rational prevention might sometimes be a multilevel model that includes several levels of organization ranging from the social to the molecular. For instance, a multilevel model of the pathogenesis of many instances of CVD might include a lack of health promotion (social level), sedentary lifestyle (behavioural level), hypertension (organ system level), inflammatory cells (cellular level) and LDL-cholesterol (molecular level), all at once.

My suggestion is roughly in agreement with what Leen De Vreese and colleagues call "*methodological explanatory pluralism*—the view that *the best form and level of explanation depends on the kind of question one seeks to answer by the explanation* and that one needs more than one form and level of explanation to answer all questions in the best way possible" (2010, 372). In contrast, medicine's current models for the pathogenesis of diseases typically include only entities and activities below the level of the whole person, consistent with what De Vreeze and colleagues refer to as "methodological reductionism" (2010, 372). As I have mentioned, *reductionism* is a frequent criticism of the medical model. Reductionism impacts medicine's prevention strategies by directing the medical gaze to lower levels in the decomposition hierarchy.

Because social and behavioural risk factors are not universal causes of our current disease types, a mechanistic model including these risk factors might have insufficient scope. The model might not represent the pathogenesis of all cases of a particular disease; it would flunk the one mechanism requirement. As we have seen, we can define new diseases based on etiological mechanisms, and have sometimes done so to good effect. Defining a disease based on a specific etiologic mechanism that contains a particular causal risk factor is not the same as defining a disease based on that risk factor. The two definitions could very well sort patients differently because a causal risk factor can participate in multiple distinct

mechanisms for a disease. Granted, we do not yet fully understand the social-behavioural mechanisms leading to chronic and noncommunicable diseases. Our models remain sketchy. But defining a disease based on a mechanistic model – even a mechanism sketch – can initiate the kind of 'etiological research program' with genuine promise.

Preventing disease is not the only priority of the disease taxonomist. The monomechanism ideal must be balanced with ideals of disease treatment, among other objectives. We labour under the ultimate dream of the old medical model: to cure disease. However, I noted in the previous chapter that chronic diseases are mostly incurable. The only way to curtail their rising prevalence – *the growing burden of chronic diseases* - is to curb their incidence. Achieving this goal may require a new etiological standpoint. Its battle cry? *Mechanisms, not monocauses!*

3.5 Conclusion

Koch's etiological standpoint, the monocausal disease model, is often thought to have initiated a revolution in our conception of diseases and in the control of infectious diseases at the turn of the Twentieth Century. Chronic and noncommunicable diseases, our modern multifactorial scourges, are thought to defy the monocausal ideal. But because causal necessity in the monocausal model is a matter of stipulation rather than discovery, chronic and noncommunicable diseases are non-monocausal by definition rather than by their nature. On closer inspection, even infectious diseases elude the monocausal model. Instead, modern diseases adhere to a constitutive model of classification.

While eliminating a necessary cause entails prevention, a necessary cause is not necessarily a good guide to intervention. Thus, the monocausal model is also a poor ideal for rational disease prevention. Instead, we need a monomechanism ideal in which each disease has a singular pathogenesis represented by a single mechanistic model. While etiologic mechanisms often line up neatly behind constitutively defined infectious diseases, achieving the monomechanism ideal for chronic and noncommunicable diseases will often require redefining them according to etiologic mechanisms. Preventing these diseases may necessitate intervening on social-behavioural mechanisms in which our most important causal risk factors figure. As medicine is still at the embryonic stage of focusing on social-

behavioural etiologic factors ('determinants of health') rather than social-behavioural mechanisms or multilevel mechanisms containing social-behavioural variables, a new etiological standpoint is needed with a focus on mechanisms instead of monocauses. Rather than multifactorialism, it is the failure to achieve the monomechanism ideal - particularly due to an over-reliance on reductionistic models of pathogenesis - that better explains modern medicine's struggle with chronic disease prevention, and thus the *growing burden of chronic diseases*.

Chapter 4 - Predicting the Results of Medical Interventions: When Mechanistic Models Misfire

4.1 Mechanisms and Medical Interventions

Modern medical interventions often exhibit *treatment effect heterogeneity*. As I discussed in the Introduction, effect sizes for newer interventions are typically far less than 50% in absolute terms, and the effect of treatment is often found to vary across patient subgroups. For instance, antidepressants are shown to work in the severely depressed, but not in the mild to moderately depressed (Fournier et al. 2010); and a reanalysis of two clinical trials found that carotid endarterectomy and aspirin both have qualitatively different treatment effects in different prognostic subgroups (Rothwell 1995).

Other times, a promising new intervention is ultimately found to be ineffective (*treatment futility*) or unsafe (*treatment harm*). Interventions frequently fail to clear pre-marketing trials for these reasons; and sometimes interventions are cleared for market but after months or years of widespread use are tragically found to have caused great harm. The diabetes drug rosiglitazone came to market in 1999 and was widely used despite concerns about potential adverse ischemic effects on the heart. By the time a 2007 meta-analysis of trials suggested that the drug substantially increases the risk of heart attack and death from cardiovascular causes (Nissen & Wolsky 2007), the Food and Drug Administration (FDA) estimated that it had already caused 83,000 heart attacks (USSCF 2007). Before interventions are tested in clinical trials, their effectiveness and safety are often predicted on the basis of our understanding of how the intervention works. In this chapter, I will explore why our mechanistic models too often fail to accurately and precisely predict treatment effect heterogeneity, treatment futility and treatment harm.

Contemporary drug discovery is usually 'rational' (Mavromoustakos et al. 2011), based on an understanding of underlying mechanisms of disease or treatment, including etiologic mechanisms like the ones discussed in the previous chapter. Typically, our models of mechanisms do not anticipate intervention failures or dissect the non-universal effectiveness of our interventions. If they did, we might not make the same mistakes or might

use interventions in more targeted patient populations. When interventions do not produce the outcomes we predicted, scientists often try to understand why, and may propose new mechanisms or revise their old models in order to reconcile what their models say with the effects that the intervention actually harbors.

Failures to predict effectiveness correctly and precisely are not always the direct result of a mechanism glitch. As we will see in the next chapter, the standard model of prediction in medicine extrapolates from population studies; it does not reason from knowledge of mechanisms. However, as many of these studies were run to explore the effect of an intervention that was initially proposed on the basis of a theoretical mechanism, our incorrect or imprecise forecasts are an *indirect* result of a prediction based on a mechanism. Further, our models are often unable to explain treatment futility or treatment effect heterogeneity, so even when we did not actually use them to make a prediction they *would* predict poorly, they are incapable of delivering a reliable forecast.

Sometimes mechanistic predictions *directly* lead to medical error. Jeremy Howick (2011a; 2011b) uses the tragic example of antiarrhythmic drugs in patients who have had a heart attack. The story goes that in the 1970s and 1980s antiarrhythmics were widely prescribed to prevent death in these patients based on mechanistic reasoning, but ultimately *caused* death in tens of thousands of people, apparently because the mechanisms linking the drugs to survival are not what doctors thought they were.

A concern that there are plentiful past failures to predict from mechanistic knowledge is probably the main reason why the evidence-based medicine (EBM) movement considers this knowledge to be a poor base for evaluating the effectiveness and safety of our medical interventions. Mechanistic rationale is either displayed near the bottom of the evidence hierarchy for treatment (Guyatt et al. 2008) or not included in the hierarchy at all (Balshem et al. 2011), representing the belief that knowledge of mechanisms is less reliable than knowledge of the results of a comparative group study such as an observational study or an RCT. Howick argues that complications with mechanistic reasoning "are, and are likely to remain, common" (2011a, 136), and that "instances of high-quality mechanistic reasoning are rare" (2011a, 144).

Orthogonal to the evidence-based turn in medicine, a mechanist turn has taken hold in the philosophy of science (Rob Skipper and Roberta Millstein (2005) call it the "new

mechanistic philosophy"). The mechanisms literature has exploded over the past couple of decades. Yet most authors exploring the functions of mechanistic thinking in science have focused on mechanistic explanation (e.g. Glennan 2002; Bechtel & Abrahamsen 2005; Craver 2007; Bechtel 2008; Illari 2013). Far fewer have probed the roles of mechanisms in intervention and prediction (exceptions include: Steel 2008; Moghaddam-Taaheri 2011; Howick 2011a; 2011b; 2013; Broadbent 2013; Cartwright 2012; Clarke et al. 2014).

As a result, several concepts and distinctions are in need of clarification, including the concept of mechanistic prediction, as well as the distinction between an 'intervention on a mechanism' and an 'intervention mechanism'. Moreover, while philosophers of science have sought to understand various unsuccessful interventions from a mechanistic perspective, there is no satisfactory unifying account of when and why mechanistic predictions go wrong. Howick (2011a; 2011b) and Brendan Clarke et al. (2014) describe various general problems that beset mechanistic prediction in medicine, including: the issue of psychologically compelling but evidentially unsupported mechanism stories; incomplete knowledge of mechanisms; and mechanisms that are stochastic, complex or are counteracted by other mechanisms. But notice that – at least at a glance – these explanations focus on different components of the mechanistic prediction triad. The first problem is framed as an issue with our cognitive activities or reasoning, the second as an issue with our knowledge or mechanistic model, and the third as an issue with the worldly mechanisms. We can certainly understand the problems of mechanistic prediction – and profitably – as originating in different corners of the triad. But it is also useful and illuminating to visualize all of these problems from the same vantage point, an objective I will pursue in this chapter.

I will start by drawing some important distinctions among concepts related to mechanistic prediction (section 4.2). I will then survey various problems with mechanistic prediction identified by philosophers of science (section 4.3). I will concentrate on mechanisms involving health interventions because interventions pose unique challenges in addition to problems that can infect any mechanistic prediction. I will show that the various problems are all instances of three more general shortcomings of mechanistic models (rather than difficulties with our reasoning or with the mechanisms themselves). Namely, problems in mechanistic prediction are due to either model incorrectness (4.3.1), model incompleteness (4.3.2), or the amount of abstraction in our model (4.3.3). Each of these shortcomings results

in failure to properly ascertain the scope of the model (4.3.4). My account explains when and why mechanistic models misfire, generating vague or incorrect predictions about the result of an intervention, or failing to predict anything at all. A general understanding allows us to see why our medical mechanistic models are so often unable to predict treatment effect heterogeneity, treatment futility or treatment harm.

4.2 Mechanistic Prediction: Distinctions Old and New

Mechanisms are defined differently by different authors, in part because authors tend to focus on mechanisms in particular scientific disciplines such as psychology or the life sciences. Howick summarizes important elements of definitions provided by Stuart Glennan (1996), Peter Machamer et al. (2000), William Bechtel and Adele Abrahamsen (2005), and Nancy Cartwright (2009)[41]: "Mechanisms are arrangements of parts/features that (allegedly) ensure a regular relationship between "inputs" and "outputs"" (2011a, 126). Likewise, Phyllis Illari and Jon Williamson provide a consensus definition based on definitions by Machamer et al. (2000), Glennan (2002) and Bechtel and Abrahamsen (2005): "A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon" (2012, 123). Their 'phenomenon' is Howick's 'output' of the mechanism, while their "entities and activities organized in such a way" are Howick's "arrangements of parts/features".

As in the previous chapter, I will consider a particular mechanism to be 'functionally individuated' (Glennan 1996). In other words, we identify a mechanism (token) and distinguish it from its surroundings according to the phenomenon that it produces. For instance, 'Professor Plum's coronary atherosclerosis' might refer to the entities, activities and organization *that produced the atherosclerotic plaque in Professor Plum's coronary artery*. The entities (objects or properties) include molecules such as LDL-cholesterol and immune cells such as macrophages; the activities (events or processes) include macrophages phagocytosing cholesterol molecules to form foam cells; and the organization includes the structure of the arterial wall, the spatial location of the entities relative to one another, and

---

[41] Cartwright (2009) refers to "nomological machines", but they nonetheless have many of the same elements as 'mechanisms' (as defined by leading mechanism accounts).

the temporal sequence of the activities. Where do we draw the line? Are causally relevant genes part of the mechanism? Is the transmission of these genes to Plum from his parents? Is the evolution of the genes in Plum's distant ancestors? As Bechtel (2015) argues, mechanism boundaries are artificial, a feature of the way we represent mechanisms rather than a feature of the mechanisms. Fortunately, we will see in section 3 that a mechanism representation need not be infinitely complete in order to suffice for prediction.

A mechanism must be distinguished from its representation; I will refer to the latter as a *mechanistic model* or *model of a mechanism* (Glennan 2002; Bechtel and Abrahamson 2005; Darden 2008). As Bechtel and Abrahamson explain (2005), we can represent mechanisms in various ways, including linguistically, mathematically, and diagrammatically. Diagrammatic and narrative/linguistic representations of mechanisms are especially important in the health sciences, and will be my modality of choice. Even within a representational modality, there is more than one way to represent a particular mechanism. Crucially, we can represent the mechanism with lesser or greater detail (when greater detail is known).

Figure 4-1 depicts a mechanistic model with very little detail.

$$X \longrightarrow Y \longrightarrow O$$

Figure 4-1. A model of an intervention mechanism.

X is the input and O is the output or phenomenon. Letters and arrows represent entities or activities. The only organization included in this model is causal priority, represented by the direction of the arrows; arrows point from causally prior entities/activities to causally posterior entities/activities.

A second distinction should be made between mechanistic explanation and mechanistic prediction. In *mechanistic explanation*, we take for granted the output (the explanandum) and explain how and why it is produced by reasoning through the model (the explanans); in Figure 4-1, we explain O by reasoning from X to Y to O.[42] Mechanistic

---

[42] This characterization accords with Bechtel's (2008) concept of "epistemic explanation" in which the model of the mechanism explains the phenomenon, rather than Craver's (2007) concept of "ontic explanation" in which the mechanism itself explains the phenomenon.

explanation allows us to account for the known occurrence of a real-world phenomenon corresponding to O by reasoning through a mechanistic model that has O as an output.

Mechanistic prediction is inferentially identical to mechanistic explanation; what differs are the components that are taken for granted. In *mechanistic prediction*, we take for granted the model and reason through the model that O will be produced. Mechanistic prediction allows us to forecast the occurrence of a worldly phenomenon, again by reasoning through a mechanistic model for the phenomenon (an additional assumption is that the mechanism's inputs either obtain or will obtain). Mechanistic predictions are made in preclinical research when scientists predict - however tentatively – from theory that an experimental intervention will produce a desirable outcome in clinical trials. Mechanistic predictions are also occasionally made in clinical medicine; for instance, in the off-label use of an intervention towards an outcome for which it has not been studied. While mathematical models are sometimes used to generate quantitative predictions in medicine (Thompson 2011a), because most medical models are diagrammatic, mechanistic predictions are typically qualitative.

I described mechanistic explanation and prediction as a matter of 'reasoning through the model'. *Reasoning through a mechanistic model* is a cognitive or inferential activity in which we simulate or animate the operation of a mechanism mentally (Bechtel & Abrahamson 2005). In comparison, Howick defines "mechanistic reasoning" as "an inferential chain (or web) linking the *intervention* (such as HRT [hormone replacement therapy]) with a *patient-relevant outcome*, via relevant mechanisms" (2011b, 929; my emphasis). Mechanistic reasoning, as Howick understands it, is reasoning through a mechanistic model *of a medical intervention*.

It is not entirely clear what kinds of inferential chains or webs Howick has in mind. One type of model is represented in Figure 4-1, where X is an intervention and O is an outcome. We can call this representation a model of an *intervention mechanism* because it depicts how an intervention produces an outcome. In contrast, Figure 4-2 below represents a model of an *intervention on a mechanism* because it shows how an intervention disrupts a mechanism by severing a causal connection within the mechanism.[43] Figure 4-2 might

---

[43] Peter Spirtes et al. (2000) use the formal causal Bayes nets apparatus to model and predict the results of ideal interventions that similarly sever causal dependencies among variables.

represent an etiologic mechanism producing a disease or a pathophysiological mechanism manifesting a symptom of disease.

X

E → W → Y → O
E → Z → Y
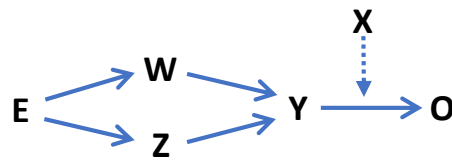
Figure 4-2. A model of an intervention on a mechanism.

In Figure 4-2, solid arrows represent productive causal activities, while the dashed arrow pointing away from intervention X represents a disruption of the causal activity producing O. This model contain two pathways. I will understand a *pathway* as a causal chain linking an input (farthest upstream element) with an output (farthest downstream element) in which exactly one arrow points away from each node (letter).[44] The first pathway in Figure 4-2 is from E-to-W-to-Y-to-O, while the second pathway is from E-to-Z-to-Y-to-O. This particular intervention has a greater chance of preventing the outcome because it disrupts the mechanism at a bottleneck. A *bottleneck* is a node through which all pathways pass. In this mechanism, the bottleneck occurs at Y.

Predicting an outcome through a model of an intervention mechanism (e.g. Figure 4-1) is a valid form of reasoning. If an intervention produces a phenomenon *and* that intervention obtains, then it follows that the phenomenon will be produced. Whether or not this reasoning is *sound* depends on whether the model is correct that the intervention will produce the phenomenon in the present circumstances.

Some authors have argued that knowledge of a mechanism does not entail knowledge of how that mechanism will behave under intervention (Howick 2011b; Broadbent 2013). This worry suggests that predicting (through a model) that an intervention on a mechanism will prevent an outcome might *not* be valid reasoning. However, when an intervention (successfully) disrupts the mechanism at a bottleneck (as in Figure 4-2), the mechanism *cannot* produce the outcome. Because the mechanism is responsible for bringing about the

---

Daniel Steel (2008) argues that the causal structure in causal Bayes nets approaches can be understood in terms of mechanisms.

[44] If the model is cyclic, elements occurring outside of the cycle will be inputs. If no elements occur outside of the cycle, we can arbitrarily consider any node to be an input.

outcome (full-stop), or maintaining the outcome, or contributing quantitatively to the outcome, preventing the operation of the mechanism is guaranteed to prevent the outcome, or terminate the outcome, or diminish its value. Therefore, predicting through a model of an intervention on a mechanism is also valid reasoning.

One might quickly object that there could be a second mechanism producing the phenomenon, and if the intervention only disrupts the first mechanism it will not prevent the outcome full-stop. But recall that a mechanism is functionally individuated; a mechanism is whatever entities, activities and organization produce that phenomenon on a given occasion. On this understanding of a mechanism, there can be only one mechanism for a given instance of a phenomenon, capturing all of the pathways leading to the outcome. Thus, if an intervention destroys the functioning of a mechanism for the phenomenon, it destroys the functioning of *the only* mechanism for the phenomenon. As we will see in section 3, the worry that the intervention might not have 'got all of the mechanisms' can be understood as a worry that our mechanistic model falls short, rather than a worry that our reasoning falls flat.

Before moving on, I will introduce a few final concepts that are important for understanding when and why mechanistic predictions miss. In the next section, I will characterize various prediction pitfalls in terms of the following four representational properties of a mechanistic model: correctness, completeness, abstractness and scope. Carl Craver and Lindley Darden (2013) call correctness a 'virtue' of a mechanistic model; they call the final three properties 'dimensions' of a mechanistic model.

Mechanistic models vary in their *completeness*. A mechanistic model is either a 'mechanism schema' or a 'mechanism sketch'. According to Machamer et al. (2000), in a mechanism schema the entities, activities and organization are all filled in so that there is continuity throughout the model, no gaps in any of its stages. In comparison, a mechanism sketch does have gaps in its stages. We can represent these gaps by placing 'black boxes' over missing entities, activities or organizational features in our model. So mechanism schemas are complete, while mechanism sketches are incomplete to varying degrees. Figure 4-3 represents the same mechanism as Figure 4-2, but one of the elements is black boxed in Figure 4-3.

X

E → ■ → Y → O

Z
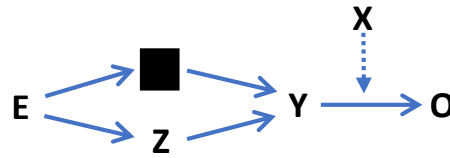
Figure 4-3. An incomplete mechanistic model.

Completeness is different from *correctness*. Craver and Darden (2013) understand correctness as a mapping or correspondence between elements of the model and elements of reality. A diagram of coronary atherosclerosis that depicts LDL-cholesterol is correct with respect to this element if real world coronary atherosclerosis involves low density lipoprotein-cholesterol molecules. It is also important that the causal relations represented in the model obtain in the mechanism. From a practical standpoint, the most important causal relation for an intervention mechanism is that between the intervention and the outcome. If the model represents a productive causal relationship between the intervention and the outcome when one does not exist, we might predict that the intervention is effective when in fact it is not. Note that an incomplete model can be correct (with respect to the elements represented), and a complete model can be incorrect (with respect to one or more elements). Like completeness, model correctness comes in degrees.

Abstractness for us will refer to the amount of detail in the mechanistic model. We can abstract away details from the model, leaving a model that is more abstract (this process is sometimes called 'Aristotelian idealization' (Cartwright 1989)). We might wish to have a model that represents how all cholesterol-lowering statin drugs work. To produce this model we must abstract away details specific to particular statin agents, including peculiarities of their structure or pharmacological properties. Alternatively, we might wish to show a model for a new brand of statin, in which case we can add in or specify details that are unique to that agent. Figure 4-4 represents the same mechanism as Figures 4-2 and 4-3, but with a greater amount of abstraction.

X

E → A → Y → O

Figure 4-4. A more abstract model of the mechanism modeled in Figure 4-2.

The elements W and Z in Figure 4-2 are replaced in Figure 4-4 with A, which nonetheless represents the same entities/activities as W and Z. Although it is more abstract than Figure 4-2, Figure 4-4 is not incomplete because there are no gaps in the model. To presage what will come, if we abstract away an important detail from a mechanistic model, we might be unable to discern when the mechanism will operate successfully.

Abstractness is related to a final representational property of a model: *scope*. The scope of a mechanistic model refers to all of the mechanisms in the world that the model faithfully represents.[45] Consider Figure 4-5.
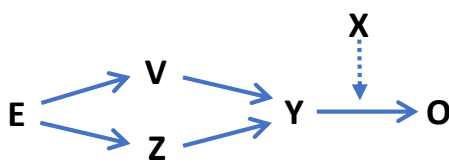
Figure 4-5. A mechanistic model with different scope compared to the model in Figure 4-2.

Figure 4-5 has a different scope compared to Figure 4-2. Figure 4-2 represents some E-to-O mechanisms that include W, while Figure 4-5 represents some E-to-O mechanisms that include V. Meanwhile, in Figure 4-4 the element A might represent 'W+Z *or* V+Z', in which case Figure 4-4 would have the combined scope of Figures 4-2 and 4-5. As Craver and Darden note (2013), more abstract mechanistic models often have a greater scope compared to less abstract models.

In the next section, I will apply these concepts and distinctions to the general problem of mechanistic prediction failure.

4.3 The Ways that Mechanistic Predictions Miss

In the previous section, I argued that there are valid ways of reasoning through mechanistic models, and thus of making mechanistic predictions. I also explained that there are two types

---

[45] The scope of a model might instead refer to the model's target. The target of a mechanistic model is the mechanism(s) that we *intend* for the model to represent, rather than the mechanism(s) that the model *correctly* represents. We sometimes construct a model in order to represent one mechanism, the model's target, but later learn that the model does not accurately represent its target. According to the sense of scope that I will use throughout this chapter, the target does not fall within this model's scope.

of mechanisms involving an intervention: an 'intervention mechanism' and an 'intervention on a mechanism'. Using a model of an intervention mechanism to predict that the intervention will produce the outcome is valid. So is using a model of an intervention on a mechanism bottleneck to predict that the intervention will prevent the outcome. As long as the prediction is valid, when a mechanistic prediction misses we cannot blame our reasoning, the first member of the mechanistic prediction triad.

Instead, we can attribute failure to the mechanistic model. In this section, I will argue that problems with predicting the results of an intervention using a mechanistic model can all be understood as a malfunction in either: model correctness, model completeness, or model abstractness. A problem with any one of these three properties makes it difficult to ascertain the scope of model, resulting in a false prediction, a vague prediction, or no prediction.

Philosophers of science explore a number of problems with mechanistic reasoning in healthcare, along with a number of notable examples of health interventions that flopped due to a mechanism glitch. I will now discuss several of these problems and examples to show how they can be absorbed into the general theory of mechanistic prediction failure presented above.

4.3.1 Correctness

Howick (2011a; 2011b) raises several historical examples of apparently incorrect mechanistic reasoning in order to build his case that mechanistic reasoning in medicine is usually poor quality. One of his recurring examples is the model of action for antiarrhythmic drugs that led to their widespread use in patients with a history of heart attack. The model predicted that antiarrhythmics would stabilize the heart's rhythm, which would in turn promote survival, especially through reducing the frequency of ventricular extra beats, which can lead to sudden cardiac death. However, in 1987 the Cardiac Arrhythmia Suppression Trial (CAST) enrolling patients with previous heart attack found *increased* mortality within the groups taking antiarrhythmic drugs (specifically, encainide, flecainide or moricizine) compared to the groups taking placebo (CAST Investigators 1989). Howick argues that our knowledge of the mechanism must have been wrong. While the stabilizing effect of antiarrhythmic drugs on heart rhythm was well established, the relationship between

stabilized heart rhythm and survival in patients with a history of heart attack was not. Because the model was wrong that a stabilized heart rhythm produces survival, it was also wrong that antiarrhythmic drugs produce survival (Howick 2011a; 2011b). Doctors now use an implantable cardioverter-defibrillator (ICD) to prevent arrhythmic sudden cardiac death in certain patients after a heart attack.

Howick construes the mechanism linking antiarrhythmics to survival as an instance of what I have been calling an intervention mechanism. We can model the mechanism fairly abstractly using Figure 4-1, where X represents an antiarrhythmic drug, Y represents stabilized heart rhythm, and O represents survival. Howick is essentially arguing that the arrow leading from Y to O is mistaken, and thus there is no pathway from X to O. The model misfires; reasoning through the model incorrectly predicts that X will cause O. If instead the model was mistaken about the identity of the intermediate element in the pathway (represented by the letter Y), the model would still be incorrect. But this mistake would not thwart mechanistic prediction so long as there is *some* real element that is caused by X and that causes O.

Incorrectness is a barrier to prediction when the model is incorrect about there being a pathway from X to O, when it shows an additional pathway from X to O that does not exist, when the overall effect of a pathway from X to O is represented falsely (e.g. as a productive pathway when the pathway is actually preventive), or when it misrepresents an element *off* a pathway from X to O (I will discuss *off*-pathway elements in the next section).

Returning to CAST, strictly speaking the trial revealed that antiarrhythmic drugs increased mortality. It may be true that the proposed positive effect of a stabilized heart on survival was mistaken, but this cannot be the whole story. X must inhibit O (prevent survival) in order to cause increased mortality in the trial. It could be that Y (stabilized heart rhythm) prevents rather than produces survival, in which case it is the arrow from Y to O that is incorrect (there should be an inhibitory arc there instead). Because we know that stable heart rhythms do not generally inhibit survival, it is more likely that X inhibits O through a separate pathway. This inhibitory pathway is not represented in the model. Thus, while Figure 4-1 might be incorrect, it is certainly *incomplete*.

In fact, the results of CAST do not rule out a link between a stabilized heart and survival; it could be that *in certain circumstances* stabilizing the heart's rhythm promotes

survival. These circumstances might even have been realized for some participants in the trial if the number of deaths caused by the drug swamped the number of deaths prevented by the drug. If this were the case, we could represent the mechanism as follows.
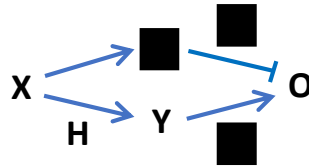
Figure 4-6. A model of an intervention mechanism with several black boxes.

H represents 'helping factors' that X needs to produce Y. The helping factors that Y needs to produce O are unknown. The second pathway according to which X prevents O is also unknown, including the helping factors needed in that pathway. If the helping factors for each pathway are not distributed uniformly in the population, X could produce O in some individuals and prevent O in others.

In short, while model incorrectness can account for many false therapeutic predictions, model incompleteness is at least as significant of a problem in medicine.

4.3.2 Completeness

Howick recognizes that our knowledge of mechanisms is often incomplete; there are "gaps in our knowledge of the inferential chain linking the intervention and the patient-relevant outcome" (2011a, 144). Not all gaps impede mechanistic prediction. Have a look at Figure 4-7 below.

Figure 4-7. An incomplete model of an intervention mechanism that permits prediction.

If Figure 4-7 was invoked to explain how X produced O, there would certainly be a problem: because the pathway through which X causes O is shrouded in a black box, the mechanistic model is a poor response to a '*how* question'. But black boxing letters on the pathway between X and O does not stand in the way of mechanistic prediction. The model has an

answer to the question, 'will X cause O?' (*yes*). Similarly, although Figure 4-8 below contains gaps, these gaps do not obstruct prediction. If both X and E occur, X will prevent O.
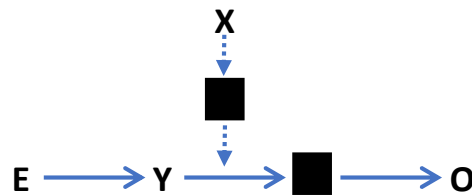
Figure 4-8. An incomplete model of an intervention on a mechanism that permits prediction.

When there is a black box instead of a letter on a pathway from the intervention to the outcome, it does not interfere with mechanistic prediction (though if there are many black boxes relative to known elements, a mechanistic *explanation* may be unsatisfying). On the other hand, black boxes off a pathway from intervention to outcome *can* pose difficulty for mechanistic prediction. In Figure 4-6, the helping factors above the arrows linking the intervention to the outcome are unknown. As a result, it is difficult to predict when X will produce O, and when it will prevent O.

In addition to gaps above the pathway from intervention to outcome, gaps in places where we wish to intervene can also foil prediction. In Figure 4-8, if Y were black boxed then we would not know which entity or activity to disrupt; the model's guidance would amount to: 'intervene somewhere along the etiologic pathway'. Alternatively, if there was a black box over the intervention then the model would suggest that we 'intervene somehow' on the mechanism, which is equally unhelpful advice.

Another form of sketchiness that can lead mechanistic predictions to miss is incompleteness with respect to the mechanism's organization. Consider Cartwright's (2012) much discussed example of the Tamil Nadu and Bangladesh integrated nutrition programs. In low income countries, malnutrition is a major source of childhood death, as well as chronic health conditions later in life. In the 1980s, the World Bank funded an integrated nutrition program in the Tamil Nadu region of India that focused centrally on educating mothers about childhood nutrition. The common sense theory was that educating mothers about nutrition would lead them to make healthier food choices for their household, and would result in better nutrition for the children. The program was a success, improving age-indexed weight and reducing malnutrition in the region's children (World Bank 1995).

Encouraged by the success of the project in Tamil Nadu, World Bank funded a similar program in Bangladesh. However, the Bangladesh program did not have the same positive impact on nutrition. The mechanistic model (which predicted that educating mothers about nutrition would produce better nourished children) predicted correctly in Tamil Nadu but poorly in Bangladesh. Several philosophers have discussed why the Tamil Nadu program's success did not carry over to Bangladesh (Cartwright 2012; Broadbent 2013; Howick 2013). One way to understand the problem is as a failure of the model to represent the required social structure, or social *organization*. The 'maternal education-to-childhood nutrition' mechanism requires a society organized such that the mothers shop for food and are responsible for the children's diet. This organization obtained in Tamil Nadu, but not in Bangladesh, where the husband does the shopping and in many cases the mother-in-law is responsible for childcare (Cartwright 2012). Had the mechanistic model been more complete with respect to the required organization, it would not have predicted success in Bangladesh.[46]

While it is undeniable that incompleteness can explain many mechanistic prediction difficulties, some philosophers attribute certain problems to the mechanisms themselves. I will now argue that we can assimilate many of these classic problems under the heading of incompleteness. One potential worry concerns 'fail-safe mechanisms'. One kind of fail-safe mechanism involves negative feedback. Consider the intervention mechanism represented in Figure 4-9.
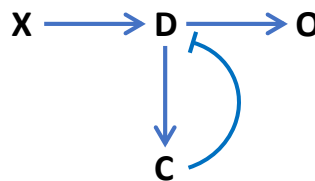
Figure 4-9. A model of an intervention mechanism including a negative feedback pathway.

The fail-safe in this case, the negative feedback loop, is not its own mechanism; rather, it is part of one pathway within the overall mechanism for O. This model could represent – albeit

---

[46] The World Bank may have predicted the success of the nutrition program in Bangladesh by extrapolating from success in Tamil Nadu. Nonetheless, the general principle that inspired the trial in Tamil Nadu predicts poorly in Bangladesh, due in part to incompleteness.

simplistically - why dieting sometimes works poorly. A resolution to cut calories (X) results in dieting (D), which produces weight loss (O). But dieting also causes a craving for high calorie foods (C), which inhibits dieting. When the craving is strong enough, it overcomes the resolve to diet, and the person cheats on their diet. Intense craving can cause over-eating, which is especially counterproductive to weight loss.

One might think that the negative feedback pathway is problematic for prediction because it impedes weight loss, and thus would make our weight loss prediction false or exaggerated. However, a careful predictor will only make a false or exaggerated weight loss prediction if they are unaware of the negative feedback loop; in other words, if the model is sketchy with respect to the negative feedback pathway. Knowing the complete model should lead them to temper their expectations of their New Year's resolution.

A second kind of 'fail safe mechanism' is a case of causal overdetermination like the one represented in Figure 4-10. The *causal overdetermination problem* is a special problem for an intervention on a mechanism.
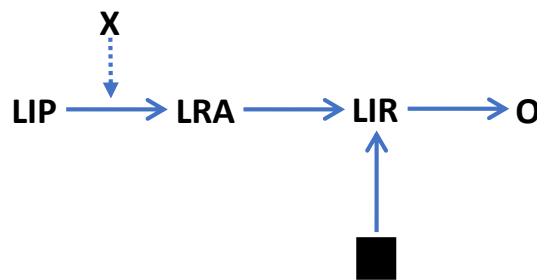
Figure 4-10. A model of an intervention on a mechanism with a causally overdetermining pathway.

Again, causal overdetermination is not due to a separate mechanism but is due to a separate pathway within the same mechanism. Let's say that the mechanism on which X is intervening is the mechanism producing hyperglycemia in an individual with insulin-dependent type II diabetes. Low insulin production (LIP) leads to low insulin receptor activation (LRA), which leads to a low intracellular response to insulin (LIR), which – through several pathways - causes hyperglycemia (O). Providing exogenous insulin (X) severs the connection between low insulin production and low insulin receptor activation; LIP no longer produces LRA. However, this type II diabetic might still have hyperglycemia because of a second, poorly understood 'insulin resistance pathway' that promotes low intracellular response to insulin. The insulin resistance pathway (black boxed in Figure 4-10)

overdetermines O, and thus the intervention on the 'low insulin production pathway' will not fully prevent O.

If I predicted from Figure 4-10 that exogenous insulin injections would fully prevent hyperglycemia I would not be a valid mechanistic reasoner because the intervention on the mechanism does not occur at a bottleneck. The only way a valid reasoning agent would predict that exogenous insulin completely prevents hyperglycemia is if the entire black boxed insulin resistance pathway is left out of the model. Only then does the intervention occur at a bottleneck. Once more, it is model incompleteness that explains why mechanistic predictions can fail in the face of a fail-safe mechanism.

The causal overdetermination problem arises when the intervention interrupts *only one* of several pathways. In comparison, the next problem occurs when the intervention activates *more than one* pathway. Perhaps the most widely discussed mechanistic prediction pitfall in the philosophy of science, I will call it the *paradoxical mechanism problem*.

Brendan Clarke et al. describe "the Masking Problem": "finding one mechanism linking *A* and *B* does not prove that there are *no other mechanisms operating*… If there are multiple mechanisms operating, they may impact on each other, and one or more may mask the effects of the mechanism you have discovered." (2014, 350). In the Masking Problem, the effect of an unknown 'mechanism' overpowers the effect of a known mechanism (presumably in the same individual). Howick describes a related challenge: "consider one potential source of unpredictability known as a paradoxical response, whereby the same drug has one effect in some cases and the opposite effect in others, presumably by activating different mechanisms or by causing a different response in the same mechanism" (2011b, 936). We encountered this difficulty already in Figure 4-6. A paradoxical response occurs when an intervention has different effects in different individuals. As a consequence, it may be difficult predict the direction – positive or negative - of the net effect in a population.

The Masking Problem and paradoxical responses are described above as difficulties with counteracting *mechanisms*. However, both issues can be understood as distinct manifestations of the paradoxical mechanism problem, a problem with our mechanistic *model*. Consider the paradoxical mechanism represented in Figure 4-11.
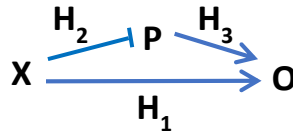
Figure 4-11. A paradoxical intervention mechanism.

Clarke et al. and Howick discuss the problems that can occur when unique 'mechanisms' have opposite effects. Because we are functionally individuating mechanisms, or carving out an individual mechanism as all of the pathways relevant to the outcome, the paradoxical mechanism problem occurs when there are counteracting pathways within *the same* mechanism. In other words, X both produces O through one pathway, and inhibits O through another pathway. In Figure 4-11, X inhibits O by inhibiting P.

One classic example of a paradoxical mechanism discussed in the probabilistic causality literature is the link between birth control pills and blood clots (Hesslow 1976; Cartwright 1989). Oral contraceptive pills containing estrogen increase the risk or probability of blood clots among women (by a small amount) due to deep vein thrombosis (van Hylckama Vlieg et al. 2009; Stegeman et al. 2013). Oral contraceptives also prevent pregnancy, which itself increases the probability of blood clots (again, by a small amount). So in certain subpopulations, estrogen-containing oral contraceptives may actually decrease the probability of blood clots by decreasing the probability of pregnancy. Figure 4-11 might explain what is going on here. Oral contraceptives (X) occasionally produce blood clots (O) with the help of factors H_1. They also inhibit pregnancy (P) with the help of factors H_2. And pregnancy occasionally produces blood clots with the help of factors H_3. Because the various helping factors are distributed non-uniformly in the population, X may produce O in some women and inhibit O in others.

Perhaps all three sets of helping factors will be present in the same individual, in which case the productive influence of oral contraceptives on blood clotting might mask their preventive influence. Or perhaps H_1 will be present in some women, while H_2 and H_3 will be present in other women; consequently, the pill will cause blood clots in the first group and prevent clots in the second group. Of course, if we knew all of the helping factors and their distributions in the population, we could predict the outcome in particular individuals. The paradoxical mechanism problem occurs when parts of the paradoxical mechanism are unknown. Perhaps one or more of H_1, H_2 and H_3 are black boxed (which was the issue in

Figure 4-6). As a result, we cannot predict from the model when X will cause O and when X will prevent O. More often, philosophers worry that an entire pathway has been left out of the diagram, and thus we are not even alerted to the serious risk of a counteracting effect. Consequently, we might predict that X will exclusively have one effect, when in certain cases it will have the opposite effect.

Examples of paradoxical mechanisms are abundant. Daniel Steel (2008) mentions the case of exercise and weight loss: exercise can produce weight loss by mobilizing fat stores, but it can also prevent weight loss by increasing appetite and food consumption. Alex Broadbent (2013) describes the example of low tar cigarettes: reducing the tar content of cigarettes hypothetically reduces exposure to tar in the lungs, but this effect is counteracted by another pathway in which the taste of low tar causes smokers to inhale more deeply. The blood thinner warfarin is another good example: warfarin prevents deaths by preventing fatal blood clots; it also (far less commonly) causes deaths by producing major gastrointestinal bleeding (Carrier et al. 2010). Howick (2011a) cites Manfred Hauben and Jeffrey Aronson (2006), who list 67 drugs that can have paradoxical adverse effects.

Many paradoxical drug effects are produced through drug-disease and drug-drug interactions, which are especially prevalent in patients with *multimorbidity*. In a drug-disease interaction, a drug might have paradoxical effects on the same outcome by worsening a pre-existing disease. For example, aspirin might prevent death by preventing a fatal cardiovascular disease event; it might also cause death by causing bleeding from a pre-existing gastric or duodenal ulcer. In a drug-drug interaction, a drug can have paradoxical effects on one outcome by altering the pharmacokinetics or pharmacodynamics of a second drug. An ACE inhibitor might lower blood pressure and prevent death by cardiovascular disease; but it might also cause death through interacting with non-steroidal anti-inflammatory drugs (NSAIDs), leading to fatal kidney failure.

Clarke et al. discuss an independent difficulty that is seemingly an issue with a worldly mechanism, the "Complexity Problem: Even where a mechanism linking *A* to *B* is well established and known in some detail, it can be hard to infer whether *A* has a positive effect on *B*, or *A* prevents *B*, or indeed whether *A* has any net effect on *B* at all. This is particularly true in cases where the mechanism is complicated: where there are several links on a pathway from *A* to *B* or where there are several pathways from *A* to *B*" (2014, 350).

116

Clarke et al. wisely point out that even if all of the components of the mechanism are known, it might be hard to infer the net effect. Returning to Figure 4-11, if X, H_1, H_2 and H_3 all occur, will O follow? The reason that the answer is difficult to predict is that the model does not represent the *strength* of the causal relationships. We could represent the functional dependence among the variables using equations if the functional dependence were known. As it stands, the diagram in Figure 4-11 does not include enough information to allow us to predict the net outcome; despite containing no obvious gaps, it is still incomplete.

Howick (2011a; 2011b) also argues that mechanism complexity – again, referring to the mechanism itself - is a major concern for mechanistic reasoning. One type of complexity he mentions is paradoxical responses, but as I just argued we can understand the problem of the paradoxical mechanism as a problem of model incompleteness. Complexity explains why we need to represent more information in our model: there are multiple pathways with qualitatively opposite effects, and thus we need to know which pathway is active on a given occasion or which pathway is 'stronger' when both pathways are active. However, it is the lack of this information in our model that directly explains why our prediction misses.

Another challenge with mechanisms according to Howick (2011a; 2011b) is that they are often – almost always, he argues - stochastic, or chancy. If a mechanism is stochastic, it can be difficult to predict whether the outcome will occur on a given occasion, even with the knowledge that all of the helping factors are present. Furthermore, it can be difficult to predict the size of the effect in a population. However, if the change in probability of the outcome due to the intervention were incorporated into the mechanistic model, stochasticity would no longer plague prediction. As Howick suggests, our reasoning can be reliable ('high quality') if we take stochasticity into account. [47]

---

[47] Howick (2011a) acknowledges that there are epistemological problems with complexity and stochasticity, problems with our knowledge of the mechanisms. However, he also argues that mechanist philosophers assume that a mechanism produces a *stable* input-output relationship. It is not entirely clear what an 'unstable' input-output relationship would amount to. If the *chance* of the output given the input is 'unstable', then the mechanism would behave neither deterministically nor stochastically but *erratically*. Howick argues that stability must be established rather than assumed. For the time being, since the existence of mechanism instability is not established either it is unclear whether this further *metaphysical* problem for mechanistic prediction is a practical problem or whether it is merely hypothetical.

Not all forms of model incompleteness foil mechanistic prediction. But when certain entities or activities, the organization of the mechanism, overdetermining or paradoxical pathways, or the strength or stochasticity of various activities are left out of the model, it can imperil our prediction.

4.3.3 Abstractness

Too much abstraction in our model can also explain why some mechanistic predictions miss. In fact, we can often translate the problems of model incompleteness into the language of over-abstraction. Recall that when we remove detail from a mechanistic model the model becomes more abstract. If the lack of this detail explains why our mechanistic prediction missed, then we could blame model incompleteness *or* model abstractness.

Sometimes incompleteness itself is represented in a model using a convention such as drawing black boxes in place of unknown elements (e.g. Figures 4-3, 4-6, 4-7, 4-8, 4-10). When a model contains black boxes, we might correctly predict that the intervention will not always work, but we might be unable to predict *when* it will not work. Other times, incompleteness is not acknowledged in the model. In Figure 4-6, instead of a black box along the top pathway from intervention to outcome, we could draw a single inhibitory arc from X to O; we could also omit the black box above each pathway. When black boxes are omitted, we might predict that the intervention will work in general but may not know whether this generalization holds universally; or we might predict that the intervention will *always* work and might predict wrongly in certain circumstances. Other times, we omit the mechanism's organization, entire pathways, or the strength/stochasticity of various activities within the mechanism. In any of these cases we may have abstracted away too much detail from the model.[48]

---

[48] Machamer et al. (2000) suggest that omitted factors can be part of the ceteris paribus clauses implicit in the model. The problem is not that a particular mechanism will operate successfully only when certain circumstances are satisfied ('ceteris paribus') – this is an inescapable characteristic of all mechanisms. The problem is that these circumstances are not represented in the model. Put another way, the 'ceteris paribus' condition is too abstract to enable precise prediction.

Certain mechanistic prediction malfunctions are best viewed as problems with the amount of abstraction in the model. As Craver and Darden (2013) explain, a mechanistic model can be 'superficial'. Descriptions that are superficial are so abstract that they are unhinged from the concrete world. If helping factors are represented superficially (e.g. as 'H$_1$', "H$_2$', and so on), we can infer that there *are* helping factors, but not the identity of those helping factors. Thus, we cannot predict when the intervention will succeed.

A superficial rendering of an intervention is a representation that is so abstract that we cannot translate it into a real world intervention – we don't know what to predict. Take the following model of the mechanism producing hyperglycemia in diabetes mellitus.

**interrupt
realization**

**disposition towards
hyperglycemia** → **hyperglycemia**
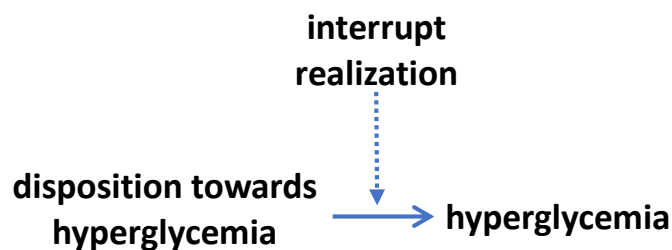
Figure 4-12. A superficial model of an intervention on a mechanism.

The mechanism is represented correctly; by definition, a disposition towards hyperglycemia manifests in hyperglycemia, and according to current theory the disposition and the hyperglycemia both exist (in patients with diabetes). The model is incomplete because the circumstances in which the disposition is manifest are omitted. But adding in these details would not tell us what constitutes a successful intervention because 'interrupt realization' is a superficial description. The intervention is described too abstractly to be of any use to a predictor.

If instead the intervention was represented with the words 'insulin injection', a problem of abstractness would remain. A disposition towards hyperglycemia is multiply realized by type 1 diabetes and by type 2 diabetes, among other subtypes, yet the model does not tell us which type to disrupt. Unfortunately, insulin injections might be less successful for type 2 diabetes than for type 1 diabetes because - as we have seen - in type 2 diabetes hyperglycemia is causally overdetermined by an insulin resistance pathway.

Cartwright (2012) notes that because abstract causal principles are multiply realized, a more abstract version might be true in a certain setting while a more concrete version might be false. This phenomenon explains why a more abstract model ('educating food buyers and

providers produces childhood nutrition') would have been predictive in Bangladesh, while a more concrete model ('educating mothers produces childhood nutrition') was not predictive. The description 'food buyers and providers' is multiply realized by mothers (in Tamil Nadu) and by husbands and mothers-in-law (in Bangladesh).

A model of the etiologic mechanism for a disease can also be overly abstract. Kenneth Rothman and Sander Greenland characterize 'multicausality' as follows: "A given disease can be caused by more than one causal mechanism" (2005, S145). Figure 4-13 contains two mechanistic models, each representing a different etiologic mechanism for the same type of disease. The disease D – like many modern chronic diseases - exhibits multicausality.

$$E_1 \longrightarrow Y_1 \longrightarrow D \qquad E_2 \longrightarrow Y_2 \longrightarrow D$$

Figure 4-13. Two mechanistic models representing two etiologic mechanisms for the same disease.

One challenge is that we cannot predict whether a successful intervention on one mechanism will prevent all cases of disease D because 'D' is so abstract that it represents instances of disease produced by different mechanisms. A successful intervention on the first mechanism might not be successful if applied to the second mechanism. Thus, we might predict that the intervention will always prevent D when in fact it will not, or we might not be able to predict how often the intervention will work. Both models might be correct and complete. Yet predicting that the intervention will prevent D in an individual case might turn out to be wrong. As an example, peptic ulcer of the stomach or the duodenum is often caused by the bacterium *H. pylori*, and in these cases antibiotics are usually an effective cure. Prescribing antibiotics to all patients with peptic ulcer though will not cure all cases of the disease because peptic ulcer sometimes results from a distinct pathogenic mechanism (e.g. drug induced injury). Similarly, statins and antihypertensive drugs prevent cardiovascular disease in some but not all cases, perhaps because cardiovascular diseases exhibit multicausality.

We can also understand the problem of multicausality in the language of the monomechanistic ideal of chapter 3. Both models in Figure 4-13 flunk the one mechanism requirement; neither model represents the pathogenesis of all instances of disease type D. Thus, an intervention on one of the models might prevent some but not all cases of D. If

instead we defined two subtypes of disease – D_1 and D_2 – based on the etiologic mechanisms of disease (the E_1 mechanism and the E_2 mechanism, respectively), then each model *would* pass the one mechanism requirement. An intervention on the first mechanism would prevent all cases of D_1, and an intervention on the second mechanism would prevent all cases of D_2.

In summary, we can understand many problems of model incompleteness equally well as problems of model over-abstractness. Other pitfalls of over-abstraction occur when a mechanistic model is superficial, when the model is multiply realized by distinct mechanisms, or when the phenomenon represented exhibits multicausality.

4.3.4 Scope

When a mechanistic model is incorrect, incomplete or overly abstract, a predictor may struggle to determine the scope of the model, the real-world mechanisms that the model represents. Consequently, a prediction – even a valid one - may go wrong in one of several ways.[49]

Model incorrectness can be a serious issue when the model is incorrect with respect to causal relationships within the mechanism because then the model might be mistaken with respect to the intervention's overall effect on the outcome. When a mechanistic model represents an overall positive effect of the intervention and is wrong, the scope of the model is empty; there will be no instances of successful intervention.[50] The model predicts that the intervention – perhaps a promising new drug - is effective, when in fact the intervention is futile or even harmful with respect to the outcome. Model incorrectness can account for many of the difficulties that researchers have in bringing a new drug to market; their model predicts that the drug will be effective, but human trials suggest otherwise. It might also

---

[49] Not every failure to determine a model's scope is fatal to prediction. For instance, if a model is totally wrong about an element on a pathway from intervention to outcome, strictly speaking the model may have no scope because it does not correctly represent any real mechanisms. But while the mechanistic model is incorrect, its error may not lead us to predict falsely.

[50] In the health sciences, researchers often construct mechanistic models to represent phenomena in the basic science lab. When an intervention works in cultured cells or animal models but not in humans, the problem is that the model's scope includes mechanisms in cultured cells or animals but not in humans.

account for certain examples of drugs that did make it to market but were later found to be ineffective or harmful, either towards the outcome for which they were originally indicated, or in their off-label use.

Model incompleteness is often problematic when the model neglects to represent an element off a pathway from intervention to outcome or when it neglects to represent an entire pathway. As a result, we are unable to specify the scope of the model.[51] If the model represents these unknown elements with black boxes, we might predict vaguely or imprecisely ('the intervention sometimes works'), or we might honour our ignorance by refraining from making a prediction. By acknowledging the existence of unknown helping factors the model alerts us to the possibility of treatment effect heterogeneity, but is unable to predict when the treatment will work and when it will not.

Finally, when a model is too abstract it omits details relevant to prediction. If the model is superficial, we may not know what to predict. If the model drops an element that sits above a pathway, or if the model or phenomenon is multiply realized, we are once again unable to correctly predict the model's scope. In this case, the model disguises treatment effect heterogeneity. Reasoning through the model and unaware of gaps in our knowledge, we might predict wrongly in a situation that falls outside the scope of the model.

Recall that many of our modern medical interventions are treatments of limited effectiveness, with effect sizes well below 50% in absolute terms. A mechanistic model representing one of these interventions is typically unable to specify its scope (which exact individuals will experience benefit) - perhaps because the model is incomplete, perhaps because the model is overly abstract. As a result, researchers test their interventions in standardly defined study populations (patients with a relevant diagnosis who meet certain eligibility criteria). Unable to discriminate which patients will experience benefit and which patients will experience no benefit or harm, doctors then use these interventions in wide target populations (patients with the relevant diagnosis and without contraindications), relying directly on the results in the study to predict the *overall* treatment effect in the target population, including the effect size. I will analyze this inference in the next chapter.

---

[51] When a model is incomplete only with respect to the mechanism's stochasticity, the trouble is not with determining the model's scope. The trouble is with determining the change in probability of the outcome due to the intervention.

Table 4-1. A General Account of Mechanistic Prediction Failure

General Problem	Instantiations of the Problem	Influence on Determination of Scope and on Prediction
Model is incorrect	• Model shows pathway that does not exist. • Model represents overall effect of a pathway falsely. • Model misrepresents an element off a pathway.	Model has no scope (e.g. model predicts that intervention is effective when it is ineffective or harmful). Believing that model has scope, reasoner may predict wrongly.
Model is incomplete	• Helping factors off a pathway are black boxed or omitted. • Organizational features are black boxed or omitted. • A pathway is black boxed or omitted (e.g. negative feedback pathway, causal overdetermination problem, paradoxical mechanism problem). • Strength or stochasticity of a pathway is omitted.	Model has scope, but reasoner is unable to specify model's scope. They may determine that intervention has heterogeneous effects, but because they are unable to discern when mechanism will produce outcome they may predict vaguely, imprecisely, or not at all.
Model is too abstract	• Model is superficial. • Model is multiply realized by distinct mechanisms. • Phenomenon represented is caused by multiple distinct mechanisms.	Model has scope, but reasoner is unable to specify model's scope. They may not know what to predict, or may predict wrongly in a situation that falls outside model's scope.

If a mechanistic model is incorrect, incomplete or too abstract, an agent reasoning through the model may be unable to correctly assess the model's scope. As a result, they may predict *wrongly* (when the model has no scope or when its limited scope is disguised by abstraction), *vaguely* (when the model's limited scope is acknowledged by black boxes), or *not at all* (when black boxes dissuade the agent from predicting or when the model is superficial). Table 4-1 summarizes my account of mechanistic prediction failure.

4.4 Conclusion

Mechanistic models of medical interventions often misfire, generating poor predictions, especially in the context of chronic health conditions and multimorbidity. When a prediction goes wrong, we might blame: the predictor reasoning through the model, the mechanistic model, or the mechanism. Philosophers of science discuss many classic problems for mechanistic prediction (including the problem of the paradoxical mechanism) and many noteworthy examples of faulty predictions in healthcare, often attributing failure to a different member of the mechanistic prediction triad. In this paper, I presented a general theory of mechanistic prediction failure. So long as the predictor is reasoning validly, we can understand prediction problems as originating in the mechanistic model, due to either incorrectness, incompleteness or abstractness. Unable to ascertain the scope of the model, the predictor may predict wrongly, vaguely, or not at all. A mechanistic model may fail to predict treatment futility or harm, resulting in the trial or off-label use of an ineffective or harmful intervention. Or the model may fail to predict treatment effect heterogeneity, meaning that the best we can do is predict that the treatment will work for some but not for all.

PART TWO: EVIDENCE-BASED MEDICINE

Chapter 5 - The Risk GP Model: The Standard Model of Prediction in Medicine

5.1 Introduction

Predictions are central to medical practice. Doctors want to know what will happen to the patient in the future given their present condition (prognosis), and how treatment or prevention might alter the natural course of events (intervention). We saw in the previous chapter that forecasts in medical research and practice sometimes come in the form of mechanistic predictions, but that evidence-based medicine (EBM) considers mechanistic reasoning to be of poor quality. Is there another model of prediction in medicine, a *standard model* in which trainees are schooled and according to which doctors practice? What we are after is a prediction scheme similar to other models of prediction in the philosophy of science, the most classic and well-known of which is the Deductive-Nomological Model (DN Model) of Carl Hempel and Paul Oppenheim (Hempel & Oppenheim 1948).[52]

Such an idealization is not to be found in medical textbooks. In fact, textbooks tend not to use the term 'prediction' to label a major category of clinical inference, but instead divide inferential activities into the traditional medical categories of diagnosis, prognosis, therapy and harm (Guyatt et al. 2008). Yet prognostic, therapeutic and harm-related inferences typically involve predictions, hypotheses about future outcomes. Even diagnosis can be conceptualized as a predictive activity; clinical textbooks speak of "clinical prediction rules" for diagnosis and the "positive predictive value" of a diagnostic test (Guyatt et al. 2008, pp. 491-505; Fuller et al. 2013, 580). Diagnosis is predictive in the wider sense of inferring an outcome that is not definitively known (i.e. the presence of a particular disease). It will be profitable to examine the shared structure of these distinct types of clinical inference.

An important clue to the existence of a standard model is that there seems to be a common target of several critiques of medical prediction, some of which will be explored in

---

[52] Another notable model of prediction is found more recently in the work of Peter Spirtes et al. (2000). Their approach uses directed graphical modelling to predict the probability distribution resulting from a targeted intervention on one or more variables.

sections 5.4 through 5.6. However, the received model lacks an explicit philosophical reconstruction – or a reconstruction of any sort, for that matter. Without a clear representation, it remains a nebulous target.

Here, I reconstruct and examine the Risk Generalization-Particularization (Risk GP) Model, the standard model of prediction in medicine. Risk GP is standard in that it represents the dominant prescriptive model in contemporary practice (the gold standard), as well as the model that many practitioners implicitly rely upon when making evidence-based decisions. Risk GP is an epidemiological model, relying centrally on aggregate outcomes in populations. Like the science of epidemiology, the model is relatively new when framed against the long history of medicine, although rational approaches to prediction have been around since at least the time of Hippocrates (460-370 BCE). The Risk GP Model actually consists of two inferences in series: a *generalization* of a risk measure from a study population to a target patient population of interest; and a *particularization*, a transformation of this measure to yield probabilistic information about a patient within the target population.

There are well-known problems at both stages. Most worryingly, the necessary assumptions for generalization and particularization may not hold widely, and even when they do hold, we might not have evidence to warrant them. These problems are not an inevitable challenge for clinical practice, or even for epidemiological predictions, but are peculiar to the Risk GP Model. Of course, most models are imperfect, and their ideal assumptions will sometimes fail to represent reality. Those circumstances demand flexibility; we should not commit ourselves to a one-model-fits-all approach, but should be model pluralists instead.

5.2 Models of Prediction in Historical Perspective

A few distinctions will be useful upfront. Alex Broadbent identifies a "process/product" ambiguity in the concepts of prediction. He distinguishes two senses of the term: prediction as a *claim*, and prediction as an *activity* (2013 89-93). The first sense of 'prediction' is a claim or hypothesis, such as: 'it will rain tomorrow'. The second sense of 'prediction' is an activity or argument, such as: 'following many previous weeks like this one it rained the next

day; therefore, it will rain tomorrow'. Prediction activities are inferences involving prediction claims.

In cases like the meteorological prediction just mentioned, prediction activities are inferences with a prediction claim, a definite forecast, as their conclusion. We can call these prediction activities *predictive inferences* to distinguish them from prediction activities that do *not* have a definite prediction claim as their conclusion. For instance, take the inference: 'on 60% of previous weeks like this one it rained the next day; therefore, the probability that it will rain tomorrow is 60%'. The conclusion is not a prediction claim; in asserting it, we are not placing a bet or committing ourselves to the occurrence of some future event. If instead it was clear skies without any approaching storm fronts, the meteorologist would conclude that the probability of rain is low, which is obviously not a prediction that it will rain tomorrow. Yet we might still want to call this statistical inference a 'prediction activity' because the conclusion tells us the probability *of* a prediction claim.[53]

As previously alluded, there are at least two, non-exclusive types of prediction claims in natural language and medical discourse. The more inclusive type encompasses all hypotheses about unknown (unobserved) events or outcomes. It includes diagnostic hypotheses like: 'the patient has heart disease'. Meanwhile, the less inclusive type of prediction claim is a subtype of the former, and includes only hypotheses about the future (e.g. 'the patient will experience a cardiovascular disease event over the next ten years'). In these cases, the outcomes are unknown specifically because they have not yet occurred (Broadbent calls this less inclusive type "narrow prediction" (2013, 93)). As we will see, the standard model can account for predictions in the broader sense. But since prognostic and therapeutic predictions are usually predictions in the narrow sense (they are hypotheses about what will happen in the future to the patient), narrow predictions will be our main focus.[54]

---

[53] Predictive inferences in medicine are typically also 'probabilistic' in that they warrant the definite prediction claim inductively. The essential difference is that a predictive inference concludes that the outcome or event will occur, while this statistical inference merely derives the probability of its occurrence.

[54] Narrow prediction claims include subjunctive conditionals, or 'counterfactuals' ('if T, then O'); specifically, counterfactuals in which the consequent refers to some future outcome or event. In order to decide on a course of action, especially when multiple alternative courses are open, physicians must often predict what will happen before the antecedents of the

An informative prediction scheme would model both the prediction claims and the associated prediction activities in a given field. The DN Model (Hempel & Oppenheim 1948) provides a good illustration. Given a physical phenomenon to be explained (the explanandum), we supply the laws of nature and particular facts that jointly entail it (the explanans). To explain why an object is accelerating at a particular rate of $1/2$ m/s^2, we can deduce the rate from Newton's Second Law and some initial conditions:

$$\text{Acceleration} = \text{Force/Mass}$$
$$\underline{\text{Force} = 1 \text{ N, Mass} = 2 \text{ kg}}$$
$$\text{Acceleration} = 1/2 \text{ m/s}^2$$

In the DN scheme, explanation and prediction are symmetrical activities; a prediction is an explanation in which the explanans (above the line) is known but the explanandum (below the line) is not. So the DN Model is also a model of prediction in the wide sense. The entire model represents a prediction activity, while the conclusion represents a prediction claim about an unknown variable.

Unfortunately, the DN Model is of limited use in characterizing modern medical prediction.[55] Few universal laws are used in clinical practice, and there is no unifying theory akin to Newton's Laws. Yet the absence of any grand theory in contemporary medicine is peculiar from a historical perspective. The miasma and contagion theories of disease persisted well into the Nineteenth Century (Gillies 2005), and from Ancient Greek medicine until the Renaissance, the Hippocratic Theory of the Four Humours, a paradigmatic example of a unifying medical theory, provided a theoretical basis for medicine (Duffin 2010, 42-45).

In the canonical interpretation of humoural theory, the balance of four bodily fluids – blood, phlegm, black bile and yellow bile – determines a person's state of health or disease. When each of the four humours is in equilibrium, the person is healthy; when any are in disequilibrium, the person is diseased. It follows that reversing disequilibrium in disease restores health. Thus, for bilious patients (with excess bile) and phlegmatic patients (with excess phlegm) the Hippocratic *Affectations* makes the following prescription: "In cleaning,

---

outcome are established. For instance, what will happen in the future to the patient *if* they are treated?

[55] Hempel (1962) also proposed a statistical model analogous to the DN Model that at first glance might seem more relevant, but because the model is intertwined with Hempel's interpretation of probability I will not discuss the details of the statistical model here.

employ medications according to the following principle: when patients are bilious, give medications that clean out bile; when they are phlegmatic, give medications that clean out phlegm" (Potter 1988, 43). Bloodletting, a therapy commonly used for thousands of years and for a wide range of diseases, was initially founded on a similar principle (Duffin 2010, 72).

We can model this kind of theoretical or mechanistic prediction as follows.

$$\frac{\text{The quantity/quality of a humour is imbalanced to degree } x \text{ (disease)}}{\text{Intervention: } -x}$$
$$\text{The quantity/quality of the humour is balanced (health)}$$

The humoural scheme is a model of prediction for treatment that follows the DN pattern. The prediction activity represented by the model is a predictive inference because we directly infer the prediction claim.

Aside from humoural theory, Hippocratic medicine was also notable for its emphasis on clinical observation. Hippocrates and his followers realized that diseases have a predictable natural course that can be abstracted from observation of similar cases. The Hippocratic *Prognostic* dictates that: "He who would make accurate forecasts as to those who will recover, and those who will die, and whether the disease will last a greater or less number of days, must understand all the symptoms thoroughly...For it is by the same symptoms in all cases that you will know the diseases that come to a crisis at the times I have stated" (Potter 1988, 48).

We can represent the reasoning as follows.

$$\frac{\text{Patients } a, b, c \text{ had fever lasting more than } x \text{ days and then died}}{\text{Patient } d \text{ has fever lasting more than } x \text{ days}}$$
$$\text{Patient } d \text{ will die}$$

In contrast to the humoural scheme for treatment, the clinical observation scheme is a model of prediction for prognosis. Greek for 'the process of knowing before', prognosis is the act of predicting a future clinical outcome in a treated or untreated case. Once again, the prediction

activity is a predictive inference. However, rather than deduction from theory, this time the inference is an enumerative induction from past experience.[56]

Throughout subsequent history, the two distinct kinds of prediction that coexisted in Hippocratic medicine – inference from theory and induction from experience – continued to coexist, though sometimes as rival rather than complementary approaches. Robyn Bluhm and Kirstin Borgerson describe "two traditions in medicine": the "rationalist" and "empiricist" traditions (2011, 204). Inference from theory is most aligned with the rationalist or mechanist tradition in the history of medicine. The mechanist tradition championed the establishment of medical theory or mechanisms of clinical causation as the route to medical knowledge; we can identify monumental figures like Paracelsus (1493-1541) and Claude Bernard (1813-1878) with this strain. In comparison, induction from experience is better linked to the empiricist tradition, which emphasized observation, classification, counting, and comparison, and within which Thomas Sydenham (1624-1689) and P.-C.-A. Louis (1787 –1872) can be placed.

Several antecedents were important in the rise of the standard model of contemporary medical prediction, which grew out of the empiricist tradition. With the development of epidemiology and statistics in the early Twentieth Century came new methods for classifying, counting and comparing, as well as for inductive inference. Then in the 1960s, the science of *clinical* epidemiology was born, which sought to apply these new tools to clinical research (Bluhm & Borgerson 2011). Through carefully quantifying outcomes in a population of patients, one can predict outcomes or determine their probability in future patients in a more precise way than is offered by simple enumerative induction.

While early Twentieth Century remedies typically aimed to cure disease or relieve symptoms, the second half of the century saw a shift towards disease prevention, especially the prevention of cancer and heart disease, the leading killers in developed societies (WHO 2011). Making use of their new tools, epidemiologists and clinical researchers began to classify risk factors, quantify the risk of disease associated with these factors, and measure the effectiveness and safety of new interventions to prevent disease.

---

[56] The models I have presented are necessarily idealizations of the Hippocratic approach; 'necessarily' because all models in the present sense are - by definition - ideal representations.

Meanwhile, a new era of consumer protectionism was beginning, including in healthcare. In the USA, the federal government granted increasing powers to the FDA to regulate the pharmaceutical marketplace, eventually requiring that all new drugs demonstrate safety and efficacy in clinical trials prior to market approval (Peltzman 1973). This legislation guarantees the existence of population studies for all new drugs marketed in the USA. However, in the early 1990s a group of clinical epidemiologists primarily located at McMaster University in Canada began to express concern that despite the abundance of tools from clinical epidemiology as well as epidemiological and clinical research evidence, medical practitioners were not using the tools or keeping up with the evidence. From their efforts, the EBM movement sprung into existence.

EBM represents the maturation of the medical empiricist tradition; it advocates applying the results of population studies over mechanistic reasoning or induction from personal clinical experience in diagnosis, prognosis and therapy (Howick 2011a). It was defended as a new "paradigm" for medical practice (EBM Working Group 1992; Djulbegovic et al. 2009), and has become dominant in medical research and education, accepted by leading medical schools and all of the major medical journals. With the increase in evidence-based clinical practice guidelines late in the Twentieth Century in terms of number and influence (Upshur 2014), EBM set the new standard for clinical reasoning.

These developments led to the emergence and ascendance of a new model of medical prediction, which I will introduce in the next section.

5.3 Introducing the Risk GP Model

Two approaches – inference from theory and induction from experience –characterized medical prediction from Hippocratic medicine down to the present. In the Twentieth Century, owing to the development of clinical epidemiology, a shift towards disease prevention, tighter drug regulation, the emergence of the EBM movement, and an increase in practice guidelines, a new epidemiological model – the Risk GP Model - emerged. A refinement on the 'induction from experience' approach, it has become the standard model of prediction in contemporary medicine.

Part of the advantage of an epidemiological or statistical model of prediction over simple enumerative induction is that patients with the same observed clinical features develop different outcomes. One patient with high blood pressure will have a heart attack, while another patient will not. This difference might arise because the causal factors responsible for heart attack in a patient with high blood pressure are indeterministic. The difference might also occur because patients with high blood pressure vary in factors (including unknown factors) that contribute to heart attack.

As a result of differing outcomes among patients, the frequency of an outcome in a population will rarely be zero or one. Nonetheless, public health practitioners may wish to predict the frequency in order to plan health services or enact health policy. To these ends, we can measure the frequency of the outcome in a study population and use the result to predict the future frequency in a particular target population.

Population-level predictions are only indirectly relevant for bedside medicine because physicians treat individual patients, not populations. The physician and their patient want to know the particular patient's prognosis and whether the patient will respond to the proposed treatment. Reasoning about the mechanism that produces the outcome in order to answer these questions undoubtedly faces challenges, which we explored in broad strokes in the previous chapter. For instance, because of mechanistic model incompleteness we usually do not know in exactly which patients the mechanism will operate successfully to produce the outcome, or else we would define our target populations more narrowly than at present to include just those individuals. Of course, neither does the frequency of the outcome in the target population tell us which patients will develop the outcome. But this frequency *can* help us to quantify our uncertainty; we can determine the probability that a patient in the target population will develop the outcome.

Frequencies are not the only aggregates we can measure in a population. We could also calculate the average for a clinical variable with a range of values greater than two, and often do. But frequencies are sometimes easier to measure and often easier to interpret. Furthermore, frequencies – but not averages – can tell us the probability of the outcome for a patient. Perhaps for these reasons, frequencies are so often used in medical prediction, and will serve as our paradigm example of an aggregate outcome. The *absolute risk* (AR) is the epidemiological term for the frequency of outcome O in a population:

$$AR = (\text{number of } O \text{ in population})/(n \text{ of population})$$

The absolute risk figures in prognostic prediction claims. For instance, according to the widely used Framingham Risk Scale, the ten-year untreated risk ($AR_{\neg X}$) of cardiovascular disease (CVD), including a heart attack or stroke, in a woman with 19 'CVD points' (assigned based on a number of risk factors) is about 25% (D'Agostino et al. 2008). In other words, we *predict* that in a population of untreated women with 19 CVD points, 25% will develop a CVD outcome over the next ten years.

In patients that are part of a "high-risk" population, defined as a population in which the untreated risk of CVD is 20% or greater (Genest et al. 2009, 570), a cholesterol-lowering drug called a statin is often prescribed. From clinical trial evidence (Baigent et al. 2005) we might predict that the *treated* risk (AR_X) in the high-risk group described above - the frequency of CVD among patients treated with a statin - is approximately 20%.[57] This figure alone tells us nothing about the effectiveness of the drug; the untreated risk might be the 25% cited above, but it might also be 20% (the treatment prevents no CVD on net) or even <20% (the treatment *causes* CVD on net). To discover a treatment's overall effectiveness, we must compare the AR_X to the $AR_{\neg X}$ using some measure of effect size (ES) such as the *relative risk* (RR):[58]

$$RR = AR_X/AR_{\neg X}$$

Measures of effect size are used in therapeutic (effectiveness or harm) prediction claims. For example, the relative risk of a CVD event due to statin therapy is 0.8 (RR = 20%/25%); as an effect of the statin, we predict that four fifths as many treated patients will have a heart attack or stroke over the next ten years compared to untreated patients.

We could also quantify the effect size using the *absolute risk reduction* (ARR), also known as the *risk difference* (RD):

$$ARR = AR_{\neg X} - AR_X$$

---

[57] The figure cited is for a 1mmol/L reduction in LDL cholesterol, which will be assumed from this point on. Therapeutic effectiveness is expected to vary by the amount of reduction in LDL cholesterol achieved.

[58] 'Effect size' sometimes refers exclusively to measures in which the difference in outcome between groups is divided by the variability (Guyatt et al. 2008, 782). Here, I use the term to denote any measure of causal association between exposure and outcome.

The absolute risk reduction of CVD due to statins in women with 19 CVD points is 5% (ARR = 25% - 20%); in other words, we predict that the difference in frequency of CVD events between treated and untreated patients attributed to the statin will be 5%.

In short, predictions in the standard model involve *risk measures*: either the absolute risk or a measure of effect size derived from the absolute risk such as the relative risk or absolute risk reduction. For prognostic predictions (involving the absolute risk), the standard model can be represented as follows.

<div style="text-align:center">

Prognosis:
In the study population, AR = r for O
—— ... ——
In target population *F*, AR = r for O
—— ... ——
For patient *a* in target population *F*, p(O|F) = r

</div>

From the absolute risk of outcome *O* measured in a study population, we predict the absolute risk in a target population defined by clinical features 'F', which can be filled in with whatever prognostic factors we choose. From this prediction claim, we then infer the probability of *O* for a patient from the target population. The prediction activity represented by the model is thus a serial inference consisting of two sub-inferences. As in all serial inferences, the conclusion of the first sub-inference (the first three lines) is a premise in the second sub-inference (the final three lines). Furthermore, both sub-inferences are enthymematic because they rely upon other premises or assumptions (represented by the ellipses) that are suppressed. As we will see, different assumptions might validate the model equally well; but its core structure – the expressed premises and the ultimate conclusion - is unchanging.

As an illustration, in determining that the patient I described earlier has an untreated CVD risk of 25%, we implicitly relied on this model. Our judgment required that we place the patient in a well-defined target or reference population (women with 19 CVD points) that has an associated absolute risk of 25%, as reported in the clinical literature (Genest et al. 2009). 'Women with 19 CVD points' describes a very large (perhaps temporally unbounded) population. The researchers who developed the Framingham prognostic model certainly did not predict the absolute risk in this population by measuring properties of each of its members. Rather, our 25% absolute risk prediction is an extrapolation from a study that

followed a sample of women (and men) over time and measured the frequency of cardiovascular disease events that accrued (D'Agostino et al. 2008).

Within the Risk GP Model, O can refer to a future event or outcome (narrow prediction), or it can refer to any unobserved event or outcome (wide prediction). The prediction activity that leads me to call 'heads' when the coin is in the air is just as sound if the coin has already landed but is covered by your hand. Thus, the version of Risk GP presented above can also be used for diagnostic predictions. In diagnosis, the outcome has already developed (several patients in the target population already have the disease O) but the frequency is not definitively known. Rather than prognostic factors, 'F' now refers to signs, symptoms and diagnostic test results.

For therapeutic predictions, the standard model is essentially the same, but involves the effect size rather than the absolute risk.

Therapy (Effectiveness or Harm):
In the study population, ES = r for O,X
$$\overline{\overline{...}}$$
In target population F, ES = r for O,X
$$\overline{\overline{...}}$$
For patient a in target population F, $\Delta p(O|F) = r$

We start with the effect size for an outcome O due to exposure X (a treatment), measured in a study population. From the effect size in the study population, we predict the effect size in target population F. Then from the effect size in the target population, we derive the change in probability of O for a patient from the target population upon exposure to X.

For example, we might want to treat our patient who has a 25% untreated risk of CVD with a statin drug. The medical literature (Baigent et al. 2005) reports that we can reduce the patient's risk of CVD by one fifth in relative terms (from 25% to 20%). Recall that the untreated risk of CVD varies by reference population, and is not 25% in all patients. Thus, in order to derive the 25%-to-20% risk change we again had to place our patient in the target population consisting of women with 19 CVD points. The treatment effect size we predict for this target population is extrapolated from clinical trials enrolling a much smaller study population.

In summary, standard prediction in medicine is a two-stage process. The first stage or sub-inference is a *generalization* or extrapolation of a risk measure from a study to a target

136

population of our choice. The second stage or sub-inference is a *particularization* or transformation of the value of the risk measure to yield probabilistic information about a patient from the target population. Hence, we call the standard model the *Risk Generalization-Particularization (Risk GP) Model* (the abbreviation is fitting considering that most medical risk management takes place in general practice (GP)). Generalization can be seen as the epidemiology, public health or clinical practice guideline stage because the conclusion is a population-level prediction. On the other hand, particularization is best seen as the clinical medicine stage because the conclusion concerns particular patients.

To be sure, the Risk GP Model is not the only prediction model used in medicine. Enumerative induction is still important in diagnosis and prognosis as physicians develop clinical experience and an acute sensitivity to patterns of disease progression. It may also underlie the prediction that a patient will benefit from a treatment or experience certain side effects based on their previous response to the treatment. On the other hand, mechanistic reasoning resembling the humoural scheme may be used to justify treatment for diseases of excess or deficiency.

Yet the Risk GP Model is aptly considered the standard model because it is held up as a normative ideal by medical authorities, if sometimes only implicitly. In other words, it is the gold standard. The *Users' Guides to the Medical Literature*, an authoritative evidence-based medicine textbook, claims that applying study results to patient care requires asking whether you can "*generalize*" or "particularize" the results to your patient (Guyatt et al. 2008, 6). Paul Glasziou and David Mant explain what this assessment involves in the context of treatment decisions: "The first stage involves an assessment of the transferability of the trial evidence [to your care setting]; the second deals with the application by the clinician to the individual" (2007, 88-89).

Other textbooks, surveys and articles also identify generalization or particularization as the gold standard approach (Szklo and Nieto 2007, 376; Post et al. 2013; Fuller 2013a; Glasziou & Irwig 1995; Goodman 1999; Djulbegovic et al. 2011; Straus et al. 2011). For instance, Piet Post et al. (2013) did a systematic review of the medical literature to identify approaches to generalizing efficacy results. They recommended that decision-makers generalize the relative effect size (e.g. relative risk) found in randomized clinical trials to the target patient population, a proposal that was reflected in most of the sources they identified,

including several EBM guides. In order to determine how a treatment will lower a particular patient's risk, we then need to consider their particular untreated risk, which we infer from a prognostic study (Glasziou & Irwig 1995; Glasziou & Mant 2007).

Moreover, Risk GP is the inferential model implicit in evidence-based practice guidelines, which set the *standard* of medical care. I (2013a) surveyed clinical guidelines recommending the use of common prescription medications and found that generalization from clinical trial results was used every time to support treatment recommendations for a guideline's target patient population. Clinicians who follow the advice of evidence-based guidelines thus rely – if unknowingly – on the soundness of an extrapolation inference. They may not always interpret the effect sizes reported in guidelines probabilistically, but in the case of preventive treatment, the benefits of the recommended therapy are often described as a lowering or reduction of a patient's risk of the undesired outcome (Genest et al. 2009; Papaioannou et al. 2010; Rabi et al. 2011); the use of this language promotes a probabilistic interpretation. By establishing the standard of care, clinical guidelines play a substantial role in directing clinical practice; this is especially the case when the standard is reinforced through schemes that reward providers for following guideline recommendations or through computerized prompts that encourage adherence to recommendations at the point of care. Through authoritative guides and practice guidelines, Risk GP is a powerful standardizing influence on modern clinical practice. The Risk GP Model thus potentially promotes the kind of *cookbook medicine* that physicians often decry.

The Risk Generalization-Particularization Model is the new standard. It tries to overcome the challenges with other models of prediction by using outcomes in study populations as a basis for prediction activities involving target populations and individual patients. In the next two sections, I will explore each constitutive inference – generalization and particularization – in turn. I will also rehearse some old problems and raise some new concerns with the standard model.

5.4 Generalization

5.4.1 The Inference Scheme

In the study population, AR = *r* or ES = *r*

$$\overline{\overline{...}}$$

In target population *F*, AR = *r* or ES = *r*

Medicine (broadly construed to include public health) makes population-level predictions. For instance, the World Health Organization predicts that the number of annual worldwide cardiovascular deaths will be 23.3 million in 2030 (WHO 2011; Mathers & Loncar 2006). Prognostic predictions are typically based on the frequency of the outcome measured in study populations. Thus, the Framingham Heart Study measured cardiovascular disease outcomes over twelve years in a cohort of 8,491 adults with no history of major CVD (D'Agostino et al. 2008). It was from these data that the Framingham Risk Scale, an algorithm to predict the absolute risk in targeted risk groups, was developed.[59]

Similarly, in treatment we use the Risk GP Model to predict the effect size in the target from the effect size in a study. Rather than the frequency or average value of the outcome, the effect size quantifies the *change* in the frequency or average. Our earlier effect size prediction for statin therapy was based on the relative risk reported in the Cholesterol Treatment Trialist meta-analysis of randomized controlled trials, which analyzed data from 90,056 trial participants (Baigent et al. 2005). The inference involved in predicting the effect size is variously called a 'generalization', 'extrapolation' or 'transportation' of the treatment effect (Horton 2000). We can represent the extrapolation of either the absolute risk or the effect size using the inference scheme above. We can also identify the scheme as a predictive inference because we directly infer a prediction claim about a risk measure in a target population. Of course, we should not demand that this prediction is perfectly accurate because we never expect that the risk measure in the target will *exactly* equal the risk measure in the study (perhaps the true relative risk is 0.8 in the study and 0.85 in the target). The prediction succeeds so long as the absolute risk or effect size we predict by extrapolating

---

[59] The development of the Framingham scale involved statistical modelling, but the algorithm still extrapolates from outcomes measured in a (cohort) study.

from the study is *close enough* to the true value in the target (though we usually do not specify how close is 'close enough').

Treatment effect predictions depend upon the prior judgment that the association we are extrapolating is indeed an *effect* of the treatment rather than a correlation with a different explanation. This prior 'causal inference' is distinct from the 'causal prediction' that we have so far been considering. The causal inference begins with a correlation between exposure and outcome of magnitude r in the study population; it concludes that the correlation is the effect of the exposure in the study. In contrast, the causal prediction starts from this effect size of magnitude r in the study, and concludes that the effect size in a target population will also be r. We will explore causal inference in epidemiological studies in the next chapter.

Interpreting the effect size, telling a causal story based on the numbers, is more difficult than it first appears. Recall that we predict a CVD absolute risk reduction of 5% due to a statin medication in a population with an untreated risk of 25% ($ARR = AR_{\neg X} - AR_X = 25\% - 20\% = 5\%$). It is tempting to conclude that the statin will prevent CVD in exactly 5% of treated patients, but this conclusion does not follow from the numbers alone. As Broadbent argues (2013, 117-121), an exposure may cause or prevent the effect with a *greater* frequency than is quantified by the measure.

This scenario can occur if the exposure replaces other causes that would have otherwise produced the outcome. In a study examining the effects of one hour/day exercise on quality of life, busier participants might substitute the exercise for another activity they were previously enjoying. The exercise regime might cause wellbeing for these participants but in lieu of their previous activities, so there would be no difference in overall effect to quantify. Furthermore, as is well recognized, a 5% net difference in outcome is perfectly consistent with the exposure preventing the outcome in more than 5% of exposed participants if it also causes the outcome in some other exposed participants (Cartwright 2010). In neither of these two scenarios does the effect size reveal the frequency with which the exposure caused or prevented the outcome. The effect size simply reveals the net difference in outcome causally attributable to the exposure (Broadbent 2013, 53-54).

Returning to the generalization inference, a crucial assumption is represented by the ellipsis in the scheme above: a representativeness assumption, stating that the study population is *sufficiently similar* to the target population. With this assumption we are

justified in expecting a similar effect, but evidence is required to support it. The evidence might consist in certain methodological features of the study, or in empirical evidence for causal comparability. As we will now see, we should be concerned that we often do not have the needed evidence, that the representativeness assumption fails in typical extrapolations, and that the assumption is poorly articulated in medicine.

5.4.2 The Trouble with the Generalization

Having described its structure, I will now discuss the main problem with the Risk GP generalization scheme: trouble with its representativeness assumption. As Broadbent argues (2013, 107-108), the entire work of extrapolation is done by this assumption. Yet several philosophers and medical commentators are worried that in extrapolating – in particular, in extrapolating treatment effects – we often fail to take account of the assumption, even when the representativeness of the study should be in doubt.

Evidence for sufficient similarity is typically one of two kinds: methodological or causal-empirical. Methodological evidence for the representativeness of the study considers how the study population was sampled; in particular, if the study population was a large random sample from the target population, we might assume (perhaps after checking a few variables known to be important) that the effect size statistic in the study represents the effect size statistic in the target. This judgment also demands that the study protocol creates study conditions - including lifestyle circumstances and environmental exposures - that are representative of the conditions the target population will encounter.

Unfortunately, study populations are not always samples from the target population. In extrapolating the effect size measured in a clinical trial, our target might be a population that was ineligible for the trial, such as older patients or patients with other concurrent diseases. If the study population is not a sample from the target population, then it cannot be a *random* sample from the target.

Even when the extrapolation is from a sample to the population sampled – what Rudolf Carnap (1945) called the "inverse inference" – clinical trials, the studies from which we most commonly extrapolate effect size estimates, seldom sample randomly (Bluhm & Borgerson 2011). Instead, strict inclusion and exclusion criteria are used to determine the wider

population eligible for enrollment, and factors such as proximity to the trial site further decide which patients are actually enrolled. Thus, in extrapolating the effect size, methodological reasons for assuming representativeness are rarely satisfied.[60]

In the absence of a methodological warrant for extrapolation, we need some form of causal-empirical evidence. In general, this evidence consists of two parts: knowledge of the variables appearing in the causal mechanisms from exposure to outcome, and empirical data on the distribution of these variables in the study compared to the target. For instance, assuming that the probability of an effect is fixed by its causes, Nancy Cartwright articulates sufficient conditions for extrapolating the effect size from study population X to target population θ: (a) X and θ are the same with respect to "[t]he causal laws affecting O", and (b) "[e]ach 'causally homogeneous' subclass has the same probability in θ as in X" (2011a, 754). The latter condition (b) specifies that the causal variables are distributed identically in the study and the target, while the former condition (a) ensures that if the causal variables are distributed identically, their contribution to the outcome will be the same. Cartwright emphasizes how epistemically demanding these conditions are. She argues that this depth of knowledge is never obtained in reality, but that moreover, we should not expect any set of sufficient conditions for extrapolation to hold widely, so the effect size will seldom be transportable.

Similarly, Daniel Steel argues that "similarity in all relevant respects may be required for extrapolating an exact, quantitative causal effect" from a study or "base population" to a target population (2008, 80).[61] Like Cartwright, Steel concludes that because background causes always differ between two populations, quantitative causal effects will rarely be

---

[60] There have been recent calls for conducting 'pragmatic trials' (vs. 'explanatory' or 'efficacy trials') in medicine, embodied in a series of articles published in the *Journal of Clinical Epidemiology* (Zwarenstein & Treweek 2009). Unlike less representative efficacy trials, pragmatic trials are designed to inform treatment decisions by deliberately enrolling typical patients in typical care settings. They are thus ideal candidates for studies that provide methodological evidence for representativeness.

[61] Steel also defines a less restrictive sense of representativeness in which the base and target are "cell representative" (2008, 205-208), however the conditions under which we can extrapolate according to the cell representativeness criterion are only slightly less demanding.

replicated in a target population, even if the target is closely ancestrally related to the study population.

I present the proposals of Cartwright and Steel only to illustrate the kinds of causal-empirical knowledge that could plausibly allow us to extrapolate the effect size. It is possible to conceive of other conditions for generalizing. However, if these other conditions are anywhere near as demanding, we must concede that the kinds of knowledge and data needed will be difficult to come by.

Several authors in the medical literature have also presented considerations for extrapolation that attempt to locate causally relevant similarities and differences between two populations (Rubins, 1994; Cowan and Wittes, 1994; Dans et al., 1998; Dekkers et al., 2010; Post et al., 2013). Unfortunately, these suggestions typically lack rigour, and fall far short of spelling out the conditions needed to validate the generalization inference. For instance, under the heading of "Process of evaluation of compelling reasons that might limit generalizability", Post and colleagues present a list of six questions adapted from the "User's Guides" to assessing the applicability of clinical trial results (Dans et al., 1998). The questions include: "Are there [biological] patient differences that may diminish the treatment response?", and "Are there important differences in patient compliance that may diminish the treatment response?" (Post et al. 2013, 642). Their approach to assessing group comparability is literally question begging; it provides us with no answers as to when patient differences will modify the effect size. Thus, the standard model of generalization is *too* elliptical, incomplete as a model for medical prediction.[62]

The claim made by Cartwright and Steel that effect sizes are seldom transportable to different populations is an empirical one; its truth depends on how often in reality our study and target populations are causally comparable in all of the right ways. As we saw in the Introduction, many medical commentators seem to share their concern; they worry that our *unrepresentative trials* fail to predict the results of intervening for our target populations. Thus, the "external validity" or "generalizability" of these trials – the extent to which the trial results predict the results of intervening elsewhere – might be poor (Feinstein & Horwitz 1997; Black 1998; Rothwell 2005; Upshur 2005; Rawlins 2008; Campbell-Scherer 2010). By

---

[62] Cartwright (2007) forcefully criticizes our effectiveness predictions for their lack of rigour.

formulating the conditions under which study results travel to a target population, Cartwright and Steel can be seen as elucidating the metaphysical basis for these worries. In Chapter 7, I question whether the relative risks measured in trials generally generalize to target populations.

Given these concerns, we should in the very least be cautious about extrapolating effect sizes from typical trials to typical target populations. Since the conditions for sound extrapolation are poorly articulated in medicine, we might worry that practitioners reasoning within the Risk GP generalization scheme would default to the conclusion without careful attention to the causal context of study and target.[63] In these cases, the assumption of representativeness – however we choose to formulate it – may fail, and we may be led into error.

5.5 Particularization

5.5.1 The Inference Scheme

$$\text{In target population } F, \text{AR} = r \text{ or ES} = r$$
$$\overline{ \dots }$$
$$\text{For patient } a \text{ in target population } F, \text{p}(O|F) = r \text{ or } \Delta\text{p}(O|F) = r$$

Generalization is a predictive inference in which we project either the frequency of O or the effect size from a study population onto a target population. The term 'generalization' might be preferred by some authors to emphasize that it is an inference from the less general (a relatively small population, often atypical with respect to its clinical features) to the more general (usually a much larger, more typical patient population). For the sake of contrast, I call the subsequent inference a 'particularization' because it is an inference from the less particular (a population) to the more particular (a patient). Philip Dawid (2015) calls the inference from group frequencies to individual probabilities 'individualization' or the 'group to individual (G2i) inference'.

---

[63] As we will see in Chapter 7, EBM's approach of 'simple extrapolation' indeed treats generalizability as a default.

The same clinical sources that discuss generalizing or extrapolating often do not describe the subsequent inference from group aggregate outcomes to individual patient outcomes. The inference is suppressed – but some prediction activity must be at work if medicine is to have anything to say about a patient's prognosis or about the potential benefits and harms of treatment for that patient. If we do indeed make inferences about patients in clinical medicine, then what might those inferences look like?

I will first consider whether the standard approach in prognosis is to infer the full-stop claim that 'patient a will O'. Carnap considered this type of induction a special case of the "direct inference" – the inference from a population to a sample drawn from the population – in which the sample has an n of 1 (1945, pp. 84-85). If the absolute risk of O is less than 100% but still high in target population F (most Fs will O), we might predict that *this F* will O. We can construct the n-of-1 direct inference as follows.

$$\text{Most } F\text{s will } O$$
$$\underline{a \text{ is an } F}$$
$$a \text{ will } O$$

When the outcome is very common, physicians may indeed reason according to this scheme. One might worry about the correctness of the inference. For example, it does not preclude the possibility that a is also a G, and that most Gs will $\neg O$ (an instance of the reference class problem, discussed further in section 5.5.2). Yet even if the inference were unproblematic, the descriptive accuracy of the model for medical prediction is certainly *not* unproblematic. Medicine often deals with outcomes that are not very common but are still important – either they are highly valued or highly disvalued. The Framingham Risk Scale considers as a high-risk group (Fs) a population in which the frequency of a CVD event (O) is 20%. Despite the fact that most Fs in this group will not develop CVD, a physician will often act to prevent CVD by prescribing a statin medication. If we assume that physicians have *some* justification (however sound or unsound) for treating high-risk patients with statins, we must search for the justification beyond the n-of-1 direct inference scheme.

Instead of a full-stop, definite prediction claim, physicians are instructed to communicate the patient's risk of the outcome (Goodman 1999). The slip from talking about risk in a population to talking about a patient's risk is subtle but not trivial; it signals a shift in meaning from 'risk' as frequency of the outcome to 'risk' as probability of the outcome.

The Framingham Risk Scale does allow us to predict the future prevalence of CVD in an untreated high-risk population (risk as frequency). But the Risk Scale is primarily intended as a tool for bedside medicine, so it seems unlikely that the ultimate aim of the scale is to predict the frequency of CVD in a population. Rather, the investigators of the Framingham Heart Study note that the purpose of their study was "to formulate a single multivariable risk assessment tool that would enable physicians to identify *high-risk candidates* for any and all initial atherosclerotic CVD events using measurements readily available at the clinic or office" (D'Agostino et al. 2008, 744; emphasis added). The purpose of the Framingham Risk Scale is to estimate a patient's risk of CVD (risk as probability) so that therapy can be rationally considered.[64]

The transition from frequencies to probabilities often goes unnoticed, and thus the two-stage inference involved in standard medical prediction is sometimes presented as a single step (Guyatt et al. 2008, 6). However, particularization is an inference distinct from the generalization from study population to target population. In prognosis under the Risk GP Model, particularization is an inference from the premise that the frequency or absolute risk of the outcome in a target population *F* is *r* to the conclusion that the probability of the outcome for a patient - given that the patient is an *F* - is *r*. Communicating a patient's risk in standard practice means expressing a probability equal to the absolute risk we predict in a reference population within which the patient can be placed (Goodman 1999; Djulbegovic et al. 2011).[65]

---

[64] Many authoritative sources use the concept of risk as probability. Paul Glasziou and Les Irwig write, "To identify patients who should expect benefit to be greater than harm, we need to predict each patient's risk" (1995, 1358). The *Users' Guides* claim, "Clinicians require studies of prognosis— those examining the possible outcomes of a disease and the probability with which they can be expected to occur" (Guyatt et al. 2008, 511). In the Preface to a *Lancet* volume called *Treating Individuals*, Peter Rothwell raises a key question that he claims is frequently asked by clinicians applying study results: "How can I judge whether the probability of benefit from treatment in my current patient is likely to differ substantially from the average probability of benefit reported in the relevant trial or systematic review?" (2007, ix). Dawid (2015) investigates how standard interpretations of probability fair in describing individual risks.

[65] I should emphasize that the probability generated by Risk GP is a conditional probability: p(*O/F*). The p(*O/F*) is a probability 'for a patient' insofar as that patient is an *F*. It should not be confused with some single case probability or chance that is fixed by all of the physical facts relevant for that patient. If such a non-trivial single case probability exists, its value will

The ellipsis in the particularization scheme above represents further needed assumptions. One indispensable assumption is that the patient under consideration is indeed a member of the reference population: *patient a is an F*. If the patient is not a member of population *F*, then there are no grounds for measuring the probability of the outcome for the patient from the frequency of the outcome in *F*.

A second important assumption is invoked to support the transition from frequencies to probabilities: the probability of each member of target population *F* is the same ($F = \{x_1, x_2,..., x_n\}$, $p(x_1) = p(x_2) =...p(x_n)$). No member of *F* is any more or less likely the patient whose individual risk we are presently estimating.[66] To illustrate this idea, imagine that all members of the target population were registered in a large central database. Imagine also that we had a randomizer - perhaps a computer program - that randomly selected a patient from the list, so that each patient had the same probability of being chosen, equal to $1/n$ (where *n* is the size of the population). If the absolute risk or frequency of *O* will be 2% and there are 100 patients on the list, we predict that 2 patients will have the outcome. The probability of randomly selecting one of these two patients that will develop the outcome is $p(O|F) = 1/100 + 1/100 = 2/100$, which is equal to the absolute risk. However, if the randomizer was broken and as a result one of these two patients was ten times as likely to be chosen compared to any other patient on the list, then $p(O|F) = 10/100 + 1/100 = 11/100$, which is *not* equal to the absolute risk. The assumption that the probability of each member of the target population is the same is crucial for the particularization inference. In the next section, we will consider the extent to which this assumption and the assumption that patient *a* is an *F* are reasonable in clinical medicine.

Just as physicians are often interested in uncommon but meaningful outcomes in prognosis, in treatment they are often interested in small but clinically important effect sizes. Recall that the CVD risk difference attributable to a statin is 5% in a target population with an untreated risk of 25% ($ARR = AR_{\neg X} - AR_X = 25\% - 20\% = 5\%$). If we had to predict anything definite, we would predict that the patient will not have the outcome either way

---

most likely diverge from the $p(O/F)$ because the target population *F* is typically heterogeneous.

[66] To use Ian Hacking's terminology, the chance setup - for us the target population along with the mechanism for selecting patients for risk assessment - must be *unbiased* (Hacking 2001, 24-25).

because the frequencies of *O* among treated and untreated patients - 25% and 20%, respectively - are well below 50%. Yet guidelines recommend treating this population with a statin (Genest et al. 2009) because CVD events are so undesirable and statins are thought to decrease the patient's risk.[67] Similar to prognostic particularization, treatment particularization under the Risk GP Model is not a predictive inference to outcomes that we predict will occur but an inference to probabilities of occurrence. The same two assumptions needed in prognosis are thus also needed in treatment: the patient is a member of the reference population, and the reference population is an unbiased probability-generating setup.

In treatment, particularization involves the effect size. We saw in the previous section that interpreting the effect size is tricky. It might be an error to believe that the absolute risk reduction reveals the frequency with which the exposure will cause the outcome in the target population. It might thus be equally erroneous to infer the probability that the treatment will cause the outcome from the absolute risk reduction.

Instead, we infer the *change* in probability of the outcome due to the exposure ($\Delta p(O|F) = r$) from the effect size attributable to the exposure (ES = r). If the effect size is measured as the relative risk, then the relevant change in probability is $p(O|X\&F)/p(O|\neg X\&F)$ because RR = $AR_X/AR_{\neg X} = p(O|X\&F)/p(O|\neg X\&F)$. However, if the effect size is measured as the absolute risk reduction, then the relevant change in probability is $p(O|\neg X\&F) - p(O|X\&F)$ because ARR = $AR_{\neg X} - AR_X = p(O|\neg X\&F) - p(O|X\&F)$. In deciding whether or not to treat, the ideal comparison modeled by Risk GP is between the probability of *O* given treatment (for *O*) and the probability of *O* given no treatment (for *O*).[68]

In deciding whether or not to treat, it is important that the change in probability of the outcome reflects the effect size of the treatment rather than a mere correlation between

---

[67] Prudential judgments – those involving a joint consideration of the probability and desirability of the outcome – are essentially judgments of the expected utility of a course of action, even when the reasoning is qualitative or not made fully explicit. Peter Schwartz (2009) argues that it may be rational to consider information beyond the desirability of the outcomes and the patient's personal risks; namely, it may be rational to consider information about how the patient's risk differs from the 'average risk' in a wider reference population.
[68] In deciding between two alternate treatments (*X* and *Z*), the comparison is instead between the probability of *O* given *X* (and not *Z*) and the probability of *O* given *Z* (and not *X*).

treatment and outcome. Switching your mailing address on a heart health survey from your home to your workplace address might switch your reference population for the study. If your new estimated probability of CVD qua member of the new reference population was different, you would not conclude that changing your address on the form had any effect on your cardiovascular health. Analogously, a physician should not recommend a potentially harmful treatment if all it will do with respect to the desired outcome is shuffle the patient around into a different reference class. They should recommend treatment to a patient only if it is reasonable to believe that the treatment might make a positive causal difference in the outcome for that patient.

An error in our upstream causal inference (e.g. concluding that there is an effect size due to mailing address from a confounded correlation between CVD and mailing address) can lead to an error in our downstream decision-making. More generally, an error in *any* of our upstream inferences, including the generalization or the particularization, might infect our ultimate conclusions. An inferential chain is broken if there is a chink in any one of its constitutive links. The particularization scheme is certainly not without its weaknesses. As with generalization, there are problems with its core assumptions, which I will now describe. I will then argue that standard particularization fails to address a further issue that probabilistic inferences involving single cases (e.g. patient *a*) must confront: the reference class problem.

5.5.2 The Trouble with the Particularization

In contrast with generalization, particularization is less often discussed in the medical literature; it is largely an implicit step in our medical prediction activities. Yet probabilistic reasoning in general *has* received serious attention in the philosophy of science, and as compared to generalization, the two assumptions needed for valid particularization are easier to formulate.

First, we must assume that patient *a* is an *F*, or else the probability of the outcome in population *F* is not directly relevant for *a*. Though the assumption might seem trivial from the perspective of the inference scheme (it is presupposed by the conclusion), it is non-trivial from the perspective of medical practice. Careful work must be done to establish that the

assumption is true. If 'F' reflects some parameter that we can ascertain with relative ease such as a high-risk score on the Framingham Risk Scale, we might be fairly certain that the patient is an F. However, if 'F' reflects a diagnosis for which we do not have decisive evidence, then our certainty in the truth of the assumption will be significantly less. Our uncertainty as to whether the patient is an F should then influence our uncertainty as to whether the patient will have the outcome.

For the sake of illustration, say that you have an urn containing spotted or freckled balls (F), as well as balls that are not freckled ($\neg F$). Additionally, say that half of the freckled balls are orange (O). Meanwhile, none of the not-freckled balls are orange. So $p(O|F) = 0.5$, while $p(O|\neg F) = 0$. Finally, precisely half of all of the balls in the urn are freckled, $p(F) = 0.5$. What is the probability that you will randomly draw an orange ball from the urn, $p(O)$? It would be arbitrary (and mistaken) to set your probability of O to $p(O|F)$ because we cannot assume that the ball you draw will be freckled; it is just as likely that the ball will be not-freckled. The correct probability is the probability of O given a ball of any colour. Since a quarter of all balls in the urn are orange, $p(O) = 0.25$ ($p(O) = p(O|F)p(F) + p(O|\neg F)p(\neg F) = (0.5)(0.5) + 0(0.5) = 0.25$). From the equation for $p(O)$ just used, we see that it will only be the case that $p(O) = p(O|F)$ when: (i) $p(F) = 1.0$[69] (we are certain that the ball will be F), or (ii) $p(O|F) = p(O|\neg F)$[70] (the probability of O is the same regardless of whether or not the ball is F). Otherwise, how closely $p(O|F)$ approximates $p(O)$ depends on how close is our certainty to 1.0 that the ball will be F, and how close the $p(O|\neg F)$ is to the $p(O|F)$.

In short, if we lack certainty about an individual's membership in F, we might not want to assume it for the sake of inference. Since we are often relatively uncertain about a patient's diagnosis or about their membership in a risk group given fallible evidence, it would often be imprudent to set our probability of O to $p(O|F)$. The Risk GP

---

[69]
$$\text{Proof: If } p(F) = 1, \ p(O) = p(O|F)p(F) + p(O|\neg F)p(\neg F)$$
$$p(O) = p(O|F)(1) + p(O|\neg F)0$$
$$p(O) = p(O|F)$$

[70]
$$\text{Proof: If } p(O|F) = p(O|\neg F), \ p(O) = p(O|F)p(F) + p(O|\neg F)p(\neg F)$$
$$p(O) = p(O|F)p(F) + p(O|F)p(\neg F)$$
$$p(O) = (p(F) + p(\neg F))(p(O|F))$$
$$p(O) = (1)(p(O|F))$$
$$p(O) = p(O|F).$$

particularization does not accommodate diagnostic or classificatory uncertainty, and provides no direction when confronted with this frequent phenomenon. When both p($O|F$) and p($O|\neg F$) can be estimated, it would be better to reason according to the equation for p(O) (total probability) above.

Particularization also relies upon a second assumption that is sufficient for deriving p($O|F$) from the frequency of O in target population F: the probability of each member of target population F is the same. Each member of the target population must have an equal probability of having their risk assessed. We modelled this situation with a database that included all patients in F and a computer program that randomly chose a patient from the list.

We might wonder how well this probability model represents our target populations in medicine. We can start by considering the target populations defined by clinical practice guidelines. Guidelines are often produced by national professional groups and are usually disease-specific, so their target populations will often consist of all patients with a certain diagnosis in a particular country. We might consider as a target population all patients on the UK's National Health Service (NHS) with a certain diagnosis. If guidelines are supposed to guide individual clinicians, then the relevant unknown is the probability of the outcome for a patient on this list who is assessed by a given clinician. However, clearly not all patients in the NHS have an equal chance of being seen by the same doctor; those patients who live in the north of Scotland have a very low chance of being seen by a physician in the south of Wales. Thus, the probability model assumed by Risk GP is a poor fit for an obvious example of a target population setup: a clinician applying a practice guideline.

Perhaps instead we should define target population F to include only those patients that are likely to be assessed by a given clinician; for instance, everyone listed in a community physician's regular patient roster. Even then we should doubt that each patient in the target population has an equal probability of being assessed; some patients are rarely seen by their physician at all. So our probability model might also inadequately represent a physician assessing their registered patients. If assessment is correlated with the outcome, then the probability of the outcome for patients who are assessed will differ from the frequency of the outcome in the overall target population. For example, it is reasonable to suspect that patients with vague, undiagnosed cardiovascular symptoms like occasional shortness of breath or minor chest pain are more likely to have their risk of CVD assessed by

their physician using the Framingham scale. Since these symptoms sometimes indicate underlying CVD, assessment might be correlated with CVD, even among Framingham high risk patients (the target population). The assumption that the probability of each patient is the same might not hold in the context of our most natural examples of target population setups, which may lead us to greatly misestimate the probability of the outcome.

Aside from the trouble with its two assumptions, there is one final weakness of the particularization scheme worth visiting, a problem with its conclusion. Most often target population *F* will be a broadly defined group such as patients with a particular disease. It is doubtful that the risk measure calculated in a study would be generalizable to many more narrowly defined target populations as the relevant features according to which we might want to narrow our target population are the same features that would ostensibly modify the value of the risk measure. But when we can confidently predict how a risk measure will differ according to other features particular to the patient, we should. Many authors in the medical literature have decried that our approach based on 'study averages' often loses sight of the relevant particularities of individuals (Feinstein & Horwitz 1997; Tonelli 1998; Kravitz et al. 2004; Greenhalgh et al. 2014; Tanenbaum 2014). This criticism is an aspect another problem we encountered earlier, which I called *the tyranny of aggregate outcomes*. Because standard medical prediction is hostage to the aggregate outcomes we have measured in a population, we might miss crucial information relevant to predicting *individual* outcomes in particular patients.

The perennial reference class problem looms in the background. As the problem goes, a member of a population is also a member of many subgroups of that population. A patient who is a member of a high-risk group for CVD might also be a member of the subgroup of women at high risk of CVD, as well as the subgroup of women at high risk of CVD who enjoy Eighties music, and so on. The problem is that the probabilities might differ in different reference groups, and since there is no coherent answer to the question of which group is *the* reference group for this individual, there seems to be no coherent answer to the question of which probability is *the* probability for this individual. However, for practical purposes one response makes intuitive sense: we should make decisions based on the probability in the narrowest informative reference class for which we can form a reasonable

probability judgment (Flores 2014).[71] If your high CVD risk patient is a member of the subgroup of high-risk patients that currently report crushing chest pain, then you are well advised to act based on the probability of a heart attack in this more narrowly defined group.

Another way to frame the response is as a matter of more evidence. We might have good evidence for the prediction claim beyond the study from which we extrapolated. This evidence might consist of the results of a different epidemiological study, or non-epidemiological evidence such as the clinician's experience suggesting that patients similar to the one under consideration fair differently on average. Sensibility (in addition to a body of philosophical literature) dictates that we should condition our belief in the hypothesis on all of the available evidence, and not only on the results of a single study, however rigorous that single study may be.[72]

The final problem with the particularization is thus that it is not particular enough; it fails to consider other features beyond 'F' particular to the patient, even when we have good evidence that the probability given 'F' plus these other features diverges from the probability given 'F' alone. Within the Risk GP prediction model, we are - at least sometimes - ignoring information relevant to prediction.

5.6 Trouble with the Standard Model and the Case for Model Pluralism

Confirmation that there is indeed a standard model of medical prediction is provided by the fact that there seems to be a common target of several compelling critiques in medicine and in the philosophy of science. There are serious problems with the Risk GP Model, especially with its assumptions, which are often difficult to warrant with evidence and will often fail in practice. These problems are not challenges to the model's validity, but rather to its

---

[71] I cannot defend the intuition here, but see discussion in Gillies (2000, 119-123). Early notable 'narrowest reference class' solutions to the reference class problem were given in Keynes (1921) and Ayer (1963). As Brendan Clarke et al. argue (2013), features suggested by the relevant causal mechanisms may help us to locate the relevant narrower reference class.

[72] For instance, see Good (1967) and Ramsey (1990). Jacob Stegenga (2011) criticizes the lack of evidential inclusiveness in medical treatment meta-analyses, studies from which we often extrapolate an effect size estimate.

soundness in many instances of its use, and thus to its privileged status as the standard model.

Since the assumptions are peculiar to the Risk GP Model, so too are the problems with the assumptions. The model is not a constraint on medical prediction because - as we have seen - there are other ways of predicting. Other models of prediction rely on different assumptions, so that another model's assumptions might hold when Risk GP's assumptions fail. Yet the relative lack of attention that other models receive in medicine suggests a certain inflexibility when it comes to prediction.

Medical prediction-makers are best served by adopting a pluralist approach to prediction: exploring all of the options to decide which model best fits the situation at hand. Model pluralism is one antidote to cookbook medicine. The first step towards pluralism is to recognize that there are other ways of predicting. To that end, I introduce just a few alternate models of prediction suggested by the preceding discussion. No doubt they each have their own weaknesses. I present them simply to make the case that model pluralism is a practicable alternative to a one-model-fits-all approach.

We saw that two distinct approaches to prediction have coexisted throughout history, one relying on theory or mechanisms, and the other relying directly on experience. In the previous chapter, I argued that mechanistic predictions can miss in several ways. Mechanistic prediction thus relies on assumptions of its own: that the mechanistic model is correct enough, that it is complete enough, and that it is not too abstract. Because it does not depend on aggregate outcomes in populations, mechanistic reasoning may be useful when *any* of Risk GP's assumptions fail or when we lack human studies altogether. It is most dependable when the relevant mechanisms are simple and well-established. As an example, osteoporosis medications are commonly prescribed to prevent bone fractures (especially hip fractures) in older women. A physician may worry that a particular patient is less likely to benefit from osteoporosis medication than the average target patient, or may have good reason to believe that the patient will benefit but may worry about harmful side effects. In hopes of preventing fractures, they may put aside the empirical evidence on osteoporosis medications and instead reason that osteoporotic bone fractures are almost always caused by falls. They might then suggest sensible measures to prevent falls in the home, and thus intervene on a common mechanism producing fractures in the elderly.

154

As another example, in medical conditions caused by nutrient or hormone deficiency (e.g. iron deficiency, hypothyroidism), we may reverse the condition very simply by supplying the nutrient or hormone in which the patient is deficient. We rarely have clinical research evidence demonstrating the effectiveness of various doses of replacement for various magnitudes of deficiency. In lieu of extrapolating from research studies, the physician might tailor the replacement dose to restore a particular patient's levels to the statistically normal range, especially if the deficiency is asymptomatic (and thus the dose cannot be adjusted based on patient feedback). This mechanistic reasoning is based on the simple principle that if a deficiency causes a condition, curing the deficiency cures the condition.

In contrast to mechanistic prediction, predicting from personal clinical observations is a model that often relies on an induction from previous experience. There are two contexts in which it may be useful: predicting from experience with previous patients, and predicting from experience with the same patient. Predicting from experience with previous patients may be useful when the research literature lacks data on a relevant patient subgroup into which the present patient falls; for instance, when the relevant subgroup is hard to operationalize. Based on behavioural cues from their patient (the way the patient describes their symptoms, their affect, their non-verbal communication), a physician may experience the intuition that the patient is seriously unwell and at risk for further deterioration in their health (a poor prognosis). It is difficult to systematically study the subgroup 'patients whose behaviour triggers the clinical intuition that they are at risk of deterioration in their health', and an expert physician may have to rely on their own experience with such patients in order to make a prognostic inference.

On the other hand, predicting from experience with the same patient is a useful approach in treatment when the patient's response is observable; for example, in treating symptoms. From previous response to treatment, we predict future response to treatment; or from previous superior response to one treatment compared with another, we predict future superior response. We must assume that the patient's current physiology is sufficiently similar to their previous physiology, but this judgment is often easier to make and more reliable than the assumption that a study population is sufficiently comparable to a target population. The former assumption might be warranted when the latter assumption is not.

155

Even if the former assumption is satisfied, the inference is not infallible; for instance, perhaps the patient's previous symptoms spontaneously remitted soon after starting treatment, and the physician is wrong to infer that the treatment will relieve the patient's symptoms the next time they recur. Thus, the inference is strongest when the patient's untreated symptoms or health status are stable rather than fluctuating. The *n*-of-1 trial can be seen as a refinement of this approach to address its weaknesses (Hankey, 2007). In an *n*-of-1 trial, a patient alternates between treatment and no treatment (or between different treatments), outcomes are systematically recorded, and effectiveness and safety are inferred, sometimes with the aid of statistics.

Other models that maintain a role for controlled studies in predicting the effect of an exposure or intervention but eschew extrapolating the effect size include the proposals of Steel and Cartwright. Although both of their approaches require detailed background knowledge, in contrast to the generalization inference scheme, the conditions for sound prediction are explicitly formulated and thus may be easier to assess. According to Steel (2008), even if we are unable to extrapolate the effect size from a study, we might still be able to extrapolate the "positive causal relevance" of the exposure. After formulating the notion of positive causal relevance in terms of ideal causal Bayes nets interventions, he develops a "mechanisms approach" to extrapolating positive relevance. Oftentimes decision-makers wish to know an exposure's effect size to help decide among alternate courses of action. Other times, predicting that the exposure will have a positive (greater than zero) effect in the target provides enough information. To illustrate, Steel develops his theory mainly in the context of extrapolating from animal studies, especially extrapolating harmful effects like carcinogenicity. We may lack good epidemiologic data but possess data from controlled animal experiments showing that an exposure causes serious harm in animals. If we can extrapolate that the exposure will also cause serious harm in at least *some* humans, this conclusion might warrant public health action; we might disseminate a public advisory warning, regulate the use of the exposure, or ban its use altogether.

Meanwhile, when we have controlled human studies but they are not sufficiently representative of the target, they might serve as a proof of principle that the exposure *can* make a difference rather than a base from which to extrapolate the effect size. Cartwright (2012) advises asking whether the causal principle or mechanism that operated in the

controlled study will also operate in the target. Predicting that the intervention "will work for us" also requires knowing that certain support factors are present in the target context to enable the intervention's causal power. Her account sheds special light on predictions involving policy or social interventions. Extrapolating from clinical studies of behavioural or complex care interventions is often difficult because the healthcare context strongly determines effectiveness and will often vary across sites, even within the same health system. An educational intervention to promote diabetes self-management will only work if the patient has the local resources to make lifestyle changes. Similarly, the effectiveness of an interprofessional team intervention aimed at developing comprehensive care plans for patients with multiple chronic conditions might depend on the composition of the care team and the patient's particular constellation of diseases.

Probabilistic models to rival the Risk GP particularization scheme are also available. As argued in section 5.2, in prognosis when we know the $p(O|F)$ and the $p(O|\neg F)$ but are relatively uncertain as to which group (F vs. $\neg F$) the patient belongs, we can calculate the probability of the outcome according to the formula for total probability: $p(O) = p(O|F)p(F) + p(O|\neg F)p(\neg F)$. This model may be useful when we lack a definitive diagnostic inference that can decide whether the patient is F or $\neg F$ - a not uncommon scenario. Diagnostic tests often lack sensitivity (given a negative result we are not confident that the patient lacks the disease) or specificity (given a positive result we are not confident that patient has the disease). But from the negative or positive finding and the reported sensitivity or specificity we can calculate the probability that they have the disease, $p(F)$, and the probability that they do not have the disease, $p(\neg F)$. If we also know the outcome rates among individuals with and without the disease, we can determine the $p(O|F)$ and the $p(O|\neg F)$, respectively. Finally, we can calculate the patient's prognosis, the $p(O)$, from the total probability equation, based on our current diagnostic evidence. We can then determine whether they fall into a useful risk category for treatment purposes.

On the other hand, we saw that the standard model neglects evidence for the probability of the outcome beyond the study population from which it extrapolates. We might know that the patient is an F and the $p(O|F)$, but are well-advised to update our probabilities in the light of further evidence or further features particular to this patient (assuming that the evidence is credible). We could choose to update our probabilities

quantitatively and formally, using - for instance - the model of Bayesian conditionalization sometimes used in diagnosis. The Framingham Risk Scale tells us the probability of a CVD event given a set of risk factors 'F', the p($O|F$). We might additionally learn that the patient has a family history of CVD in a first-degree relative, 'H', and may wish to revise the probability of a CVD event accordingly, since patients with a family history of CVD are more likely to have CVD themselves. Bayes' Theorem tells us that p($O|H$) = [p($H|O$)p(O)] / p(H). We can treat the risk of CVD among patients with risk factors 'F' as the base rate or prior probability of CVD, p(O). We can then use other epidemiologic evidence to determine the probability of having a family history of CVD among patients with CVD, p($H|O$), as well as the probability of having family history of CVD in general, p(H). Finally, we can calculate the probability that the patient will have CVD given the new information that they have relevant family history, the p($O|H$), using Bayes' Theorem.[73]

Other times, probabilities may be difficult to estimate; and besides, physicians and patients might struggle to make meaning out of the precise numbers. Thus, we might instead choose to update our probabilities informally, or even qualitatively. We might expect a patient's magnitude of benefit from a treatment to be lower if they are poor adherer to the treatment. We may not know exactly how much lower it will be, but depending on our purposes, it might still be useful to expect a lower risk reduction than inferred by Risk GP. Among patients that achieve some reduction in LDL cholesterol using statin therapy but fall short of evidence-based targets, we might expect the relative risk to be higher than the 0.8 predicted from trial evidence (less of a risk reduction). Thus, the reduction in probability of CVD due to statin therapy will also be lower. Among patients that achieve *half* of the recommended reduction in LDL cholesterol, we might predict that the relative risk will be roughly 0.9 (half as many CVD events prevented). For the patient with an untreated risk of 25%, this prediction suggests a reduction in probability of CVD from 25% to 22.5%, compared to the reduction from 25% to 20% predicted by Risk GP. Modifying probabilities derived from aggregate study outcomes is a revolt against the tyranny of aggregates.

---

[73] Guidelines for CVD risk assessment instruct physicians to double the risk computed by the Framingham scale when CVD is present in a first-degree relative younger than age 60 (Genest et al., 2009), which is good advice if the p($H|O$) / p(H) is roughly equal to 2.

Considering that alternate models are possible, we should question whether we need a standard model of prediction in medicine, a strategy that seems to promote the inflexibility of reasoning (cookbook medicine) I caution against. Medicine would better accommodate the context-sensitivity of its prediction activities if it were instead standard to carry along a plurality of models and to match the model to the circumstances. Unfortunately, authoritative medical textbooks and guides that teach the gold standard of prediction fail to mention many of the alternatives. Other models, such as mechanistic prediction and predicting from personal experience, receive variable recognition and often scant exposition. Crucially, because the standard model is elliptical – its assumptions remain largely unarticulated in standard practice – we are often unable to recognize when the model fails, are often unable to appreciate the importance of alternate models, and often fail to take model pluralism seriously. Yet when it comes to medical prediction, many models are surely better than one.

5.7 Conclusion

'Prediction' refers to prediction activities as well as prediction claims, broadly construed to cover unknown events and outcomes, or narrowly construed to cover only future events and outcomes. In medicine, narrow predictions are made in prognosis and treatment, while broad predictions are made in diagnosis. Induction from experience is one rational approach to prediction used throughout medical history from at least the time of Hippocrates. It matured with the ascent of epidemiology and statistics, giving rise to the Risk Generalization-Particularization Model, the standard approach to prediction in contemporary medicine. The generalization involves extrapolating a risk measure (either the absolute risk or the effect size) from a study population to a target population, while the particularization involves probabilizing the risk measure for a patient from the target population.

The standard model is not without its troubles. Risk GP models medical prediction incompletely; further assumptions are needed in place of the ellipses. First, generalization requires a representativeness or sufficient similarity assumption. Unfortunately, in treatment the effect size may not generalize widely due to our unrepresentative trials. Moreover, the conditions for sound extrapolation are poorly articulated in standard practice, resulting in an overly elliptical generalization scheme. Meanwhile, particularization depends upon two

assumptions. The first assumption, that the patient under consideration is a member of the reference population, is questionable whenever there is significant diagnostic uncertainty. The second assumption, that the probability of being assessed is the same for each patient, is tenuous in the case of perhaps our most natural target populations. Finally, standard particularization is frequently not particular enough, failing to consider further features that locate the patient in a narrower informative reference class, and thus allowing aggregate outcomes to dominate the prediction of individual outcomes. There are of course other models of prediction we could use. Rather than inflexibly committing ourselves to a standard medical prediction cookbook, it should be standard to embrace model pluralism in medical prediction.

Chapter 6 - The Confounding Question of Confounding Causes in Randomized Trials

6.1 Confounders and Causes

The reasoning behind classical controlled experiments is simple. The experimenter sets up two conditions so that they are as alike as possible in every causally relevant way except for one experimental factor. If there is a difference in effect, then logic compels us to accept that the experimental factor is a cause of the effect. John Stuart Mill called this inference scheme the 'method of difference'.

Regrettably, human studies, including experiments involving human subjects, are not so simple. Comparative group studies, in which researchers compare human populations, are especially defiant of Mill's ideal as human populations are heterogeneous in terms of variables that might be relevant to the measured outcome but that the researchers cannot manipulate (for example, age). When these variables are imbalanced among the study groups, statisticians, epidemiologists and social scientists tend to call them *confounding factors* or *confounders.* One clever way of 'controlling for' confounders is by allocating participants to the study groups such that the confounders are balanced among the groups, similarly represented within each. The study groups are thereby (hopefully) rendered adequately comparable.

One a posteriori strategy for achieving balance in comparative group studies involves looking for factors we suspect to be relevant and ensuring that the groups are 'matched' for each of these factors. This strategy is harnessed in observational group studies, a mixed bag of study designs in which subjects are generally followed in the course of routine life.

The a priori strategy involves *randomly* allocating participants to the study groups. If the study sample is large enough (so the rationale goes), we can rest assured that confounders, including the ones we do not suspect, will be distributed similarly in each group. In other words, we run a *randomized* controlled trial. From the second half of the Twentieth Century onward, RCTs have become increasingly numerous in the human sciences. The evidence-based medicine and evidence-based healthcare movement began in

the early 1990s, and the evidence-based policy (EBP) movement in education, social planning and other areas followed on its heels. Both movements recommend turning to RCTs over observational studies in decision- and policy-making whenever possible. According to EBM, RCTs are "so much more likely to inform us and so much less likely to mislead us" in assessing the effectiveness of therapy (Sackett et al. 1996, 72).

A treatment or policy intervention is effective only if it causes some outcome. Researchers thus measure effectiveness through a causal inference. In the previous chapter, I remarked that causal inference is indispensable if we are to make causal predictions by generalizing from study results. In an RCT, randomization is thought crucially important for the causal inference; one influential claim sometimes made on its behalf is that randomization controls for *all* of the confounding variables, including those that are known (suspected) as well as those that are unknown (unsuspected).

In a paper that is already a classic in the philosophy of medicine, John Worrall (2002) considers various arguments for the superiority of the RCT and concludes that it is overrated within EBM. In the Introduction, we encountered objections similar to this one, which accuse modern medicine of undue *RCT worship*. Notably, Worrall rejects the claim that randomization controls for all confounding variables, even probabilistically speaking. Far from being probable, he argues that it might be highly improbable that all confounders are controlled in an RCT. Yet Worrall worries that epidemiologists (2002) and philosophical accounts of RCT causal inference (2007) commit to the assumption that - at least in the ideal - RCTs control for all confounding variables.

Philosophers often worry about confounders as other causes of the study outcome – what I will call *confounding causes* - that could explain a difference in outcome between study groups if they too are distributed differently between groups. Jeremy Howick (2011a) argues that a fundamental epistemic and scientific principle justifies this worry: plausible confounders provide alternate explanatory hypotheses for the study's result, and we must eliminate or rule out plausible rival hypotheses before we can accept the hypothesis that the study exposure explains the result. If RCT causal inference demands balance in confounding causes yet RCTs cannot supply it (with a reasonable probability), confounding causes seem to present a confounding conundrum at once philosophical and scientific.

In this article, my aim is to probe the importance of confounders in RCT causal inference; doing so will throw light on the role of randomization, which is often at issue in debates about whether randomized studies are epistemically superior to other kinds of group studies. I call the assumption that all confounding causes are balanced among the study groups *the balance assumption*. Some argue that: RCTs, our gold standard studies, are unable to balance all confounders; the balance assumption fails (*first commitment*). Yet, intuitively: the balance assumption is an important condition for comparative group study causal inference (*second commitment*). It thus seems that either our gold standard studies typically fail to warrant the conditions for a causal inference, or one of the previous two commitments is mistaken.

I will accept the first commitment in section 6.2, and deny the second commitment in section 6.3 - the balance assumption is the wrong logical ideal. In section 6.4, I will propose an alternate ideal, along with required conditions for causal inference. The ideal ought to guide the design of comparative group studies, while the required conditions ought to guide the interpretation of group study results. In section 6.5, I will distinguish two concepts of 'confounder'. I will argue that confounders are primarily important not as causes but as correlates of the ultimate 'other cause', which I call *C*. The role of randomization is not to balance confounding causes, but to prevent systematic imbalances in these confounding correlates at baseline. Finally, in section 6.6, I will apply lessons learned in this chapter to a problem of central concern: the exclusion of patients with comorbidities from the majority of randomized trials.

6.2 The Balance Assumption

There are three parts to the balance assumption in need of clarification: what a confounding cause *is*, what it means for a confounding cause to be *balanced* in a study, and what it means for *all* confounding causes to be balanced.

The concept of epidemiological confounding has undergone several historical revolutions (Morabia 2011), and is a highly confused concept even its modern form (Greenland & Robins 1986; Pearl 2009). The confusion may be partly due to a clustering of multiple related yet distinct concepts under the term 'confounder'. One sense of 'confounder'

or 'confounding factor' is an alternate cause of the study outcome (neither the exposure nor one of its mediating causes) that researchers must control for in order to avoid a faulty causal inference. This *direct causal* concept of confounder is widely used in the philosophical literature on RCT causal inference (Papineau 1994; Worrall 2002; Howson & Urbach 2006; Cartwright 2007a; Howick 2011a; La Caze 2013). As Worrall argues, "The effects of the factor whose effect is being investigated must be "shielded" from other possible confounding factors…There is, however, clearly an indefinite number of unknown factors that *might* play a causal role" (2002, S321-S322). Meanwhile, Howick describes conditions that a factor must satisfy in order to count as a confounding factor, including: "*The factor potentially affects the outcome*." As an illustration, Howick suggests that "[m]ore people in the experimental arm of the trial might have red hair, but since red hair is unlikely to affect most outcomes, on the basis of current knowledge it does not satisfy the first condition for being a confounder" (2011a, 35).[74] Adam La Caze suggests that "many possible causal factors differentially distributed in the study groups may influence the observed effects of treatment. The investigators of observational studies…attempt to minimize the effects of such possible confounding factors" (2013, 358). Finally, Colin Howson and Peter Urbach define 'prognostic factors' (a term often used interchangeably with 'confounding factors') as "those respects that are causally relevant to the progress of the medical condition under study" (2006, 184).

An inventory of common confounding factors makes for a heterogeneous list. For instance, Howick mentions exercise, age, health, social class, and placebo effects as potential confounders (2011a, 34-35). Exercise is believed to play a preventive causal role in mechanisms that produce cardiovascular outcomes like heart attack and stroke, and for this reason might be associated with these outcomes in a study. Yet it is not clear that the next two factors on the list are causes. Age and health status are certainly stock examples of confounding variables that randomization is supposed to disarm. Yet age, the time elapsed since a participant was born, seems no more causally efficacious than today's date on the

---

[74] Howick's second condition for a confounder is an orthodox one: "*The factor is unequally distributed between experimental and control groups*" (2011a, 35). If this condition is not established or if it fails, we can call the factor a *potential* confounder. The 'potential' qualifier should be taken as implicit if not explicitly stated in what follows whenever it is not known whether the factor is balanced or when the factor is known to be balanced.

Gregorian calendar. Similarly, 'health' seems to refer to some (unmeasured) composite of an individual's total fitness or functioning – a construct rather than a physical entity. Age and health status are often associated with outcomes like heart attack or stroke, but it is not clear that they causally affect those outcomes.[75] These factors seem to fail Howick's own direct causal criterion for confounders.

There thus appear to be variables that defy the direct causal concept of confounder yet are considered paradigmatic confounding factors nonetheless. I will distinguish confounders that are causes of the study outcome (*confounding causes*) from those that are not. The balance assumption I will examine here refers exclusively to confounding causes. For now, think of confounding causes as causes of the outcome (other than the exposure and any of its downstream mediating causes) that are imbalanced among the study groups. In section 4, I will fill in this sketch by exploring the logical relationship between confounding causes and study outcomes. Then in section 5, I will turn to a distinct concept of confounder and explain its relevance for group study causal inference.

Worrall (2002) presumes that the notion of balance used by the methodologists he cites is statistical: a factor is imbalanced when its distribution is highly skewed and balanced otherwise. In other words, a variable is balanced when its average value or relative frequency is not significantly different between groups. What does it mean for *all* confounding causes to be balanced? For Worrall's methodologists, it means that *each* confounder (suspected or unsuspected) is balanced. A balanced distribution of each confounder is appealing because at first glance it is the kind of ideal comparability that warrants a causal inference in a comparative group study. Therefore, we will understand *the balance assumption* as maintaining that *each potential confounding cause is distributed similarly among the study groups*. Although Worrall attributes to his sources the idea that randomization achieves

---

[75] How might age and health status come to be associated with health outcomes that they do not cause? Age increases linearly over time simply because is a temporal variable, and many health outcomes tend to worsen over time for other reasons; for instance, tissue damage accumulates over the lifetime (sometimes leading to disease or injury), and atherosclerosis accumulates in arteries over the lifetime (sometimes leading to heart attack or stroke). If health or health status is a composite measure of various health outcomes, it will similarly worsen over time, and will be correlated with any specific health outcomes that follow the same trajectory.

balance in *all* variables, some confounders (e.g. 'placebo effects') are principally controlled through other common RCT design features such as blinding.

Worrall (2002) reconstructs various arguments made by methodologists and philosophers as claiming that balance in all potential confounders probably (rather than certainly) obtains in an RCT.[76,77] After all, even if there is only one confounding cause and the process determining its distribution in the study is random, there is still a small chance that the distribution of that one cause is significantly skewed.

Just how probable is it that the balance assumption will obtain in any given RCT? Worrall argues that it is potentially unlikely: "given that there are indefinitely many possible confounding factors, then it would seem to follow that the probability that there is some factor on which the two groups are unbalanced (when remember randomly constructed) might for all anyone knows be high" (2002, S324). He argues that it is a "quantificational fallacy" to infer that the probability of balance in indefinitely many confounders is high from a high probability of balance in any one particular confounder. Similarly, Howson and Urbach emphatically argue that if the number of factors in an RCT is unknown, all one can say for sure about the probability of an imbalance is that "it lies somewhere between zero and one!" (2006, 196). I accept their point that if there are an indefinite or unknown number

[76] For instance, Worrall (2002) cites D.P. Byar et al., who claim that "randomisation tends to balance treatment groups in covariates (prognostic factors), whether or not these variables are known" (1976, 75). Howick quotes *Bradford Hill's Principles of Medical Statistics* (1991) as claiming that "We can equalise only for such features as we can measure or otherwise observe, but we also need unbiased allocation for *all* other features, some of which we may not even know exist. Only randomisation can give us that" (Howick 2011a, 50; emphasis added). More recently, Edward Cox and colleagues state that "Randomization ensures reasonable similarity of the test and control groups and protects against various imbalances and biases that could lead to erroneous conclusions" (2014, 2350).

[77] In response to Worrall (2002, 2007), statistician Stephen Senn (2013) objects that his fellow statisticians are well aware that many potential confounders will be imbalanced in a randomized trial, and that the conventional analysis of trials assumes as much (see also La Caze et al., 2012). He further argues that Worrall's concern about baseline imbalances is irrelevant because "[i]t is not necessary for the groups to be balanced" (1442); to believe otherwise is to subscribe to a myth that Senn assumes "no medical statisticians believe" (1439). Perhaps Senn is right; perhaps the clearer, more careful, or more considered views of the authors that Worrall (and others) quote would agree with Senn's assessment of the probability and necessity of balancing all confounders. But because the balance assumption seems to function as an important ideal, it is worthy of examination anyhow.

of potential confounders, then the probability that all potential confounders are balanced is indefinite or unknown – and not necessarily high.

But perhaps we can be a bit more definite. Again, since Worrall describes the balancing for which randomization in particular is responsible, let us concentrate on the distribution of confounding causes at baseline. We can quantify the probability that all causes are balanced at baseline, p('all'). Assuming that the relevant causes are statistically independent of one another, p('all') = (p('one'))^n, where p('one') is the probability that one cause is balanced, and n is the number of unique causes. Let us also permit a weak degree of balance; say, a range of similarity in the distribution of the cause between groups that we would expect 95 times out of 100 when we randomize 'in the long-run', so that p('one') = 0.95. Then p('all') = ('0.95')^n. If $n = 14$, then p('all') = 0.49. Thus, if there are 14 or more unique confounding causes in an RCT, one or more will probably be *im*balanced at baseline, even if the probability of balance is high for any one given cause.

Fourteen is probably an underestimation of the number of statistically independent causes in a randomized trial; dozens of genes contribute to the endpoints in which health researchers and social scientists are interested, and most genes are inherited independently of one another. The upshot of this demonstration is that there is good reason to doubt that the balance assumption is true for even one of our best RCTs. In a well-designed and well-executed randomized trial, we should not assume that all confounding causes are balanced among the groups.

In his (2007), Worrall cites philosophers, including Nancy Cartwright (1989), who endorse the importance of controlling for all confounding causes in an RCT, cashed out in terms of probabilistic independence between group allocation and each cause. According to Cartwright, in an "ideal RCT", by definition the "assignment of individuals to either the treatment or the control group should be statistically independent of all other causally relevant features that an individual has or will come to have" (1989, 64). Might there be a link between (1) probabilistic independence between group assignment and each cause and (2) a balanced frequency distribution for each cause? Worrall conjectures,

> [If] we were to take the study population and divide it again and again by some randomizing device into control and experimental groups and keep a cumulative total of the relative outcomes in the two groups, then we would expect that in the indefinite

long run, the innumerable other possible causal factors would balance out [among study groups] (2007, 472).

Worrall is pointing out that in a 'long run RCT', the balance assumption might be satisfied. Along these lines, David Papineau – whose account of RCT inference Worrall (2007) also discusses – claims that randomization ensures that all other causes of the outcome are probabilistically independent of the treatment, which will "show up, not just in this sample, but in the long-run frequencies as the randomized experiment is done time and again (1994, 447). But as Worrall argues, whatever may be true in the long run or in an ideal RCT is not necessarily true in a real RCT. Thus, the probabilistic accounts Worrall surveys provide us with no further reason to believe that the balance assumption will obtain in reality.[78]

We must however question to what extent ideal RCT causal inference relies on the balance assumption. This question will occupy us in the next section.

6.3 The MOD-G Study

So far we have seen that balancing all causes in an RCT is a lofty ideal. But why bother with these confounded confounding causes in the first place? Cartwright makes explicit one powerful intuition favouring this strategy: "The underlying supposition is that differences in probabilities require a causal explanation; if the distribution of causes in the two groups is the same but for T yet the probability of O differs between them, the only possible explanation is that T causes O" (2011a, 751). This supposition is based on the intuition that a difference in effect implies a difference in cause. To harness this inferential machinery, one seeks out or sets up a situation in which there is a difference in effect or probability of effect, along with only one difference among a set of plausibly relevant causes. In other words, one runs a Mill's method of difference study.[79] The method of difference is the usual paradigm

---

[78] Nor should the accounts of Cartwright and Papineau be understood as describing the distributions of confounding causes that typically obtain in finite real-world RCTs, but as describing ideal sufficient conditions for causal inference. Cartwright, citing Worrall (2002), concedes that "[i]t is of course not clear how closely any real RCT approximates the ideal" (2010, 64). In section 6.4, I will examine the sufficient conditions that Cartwright proposes.
[79] Mill's classic description of the supposition underlying the method of difference reads: 'If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring in the

for classical controlled experiments, but Cartwright suggests that observational studies and RCTs involve the same logic (751).[80]

For instance, in a case-control study the investigators compare a positive case of the outcome with a control case in which the outcome is absent. The choice of control is not made arbitrarily; the investigators select a control that is similar to the case in causally relevant background circumstances. The investigators can then look for a potential cause of the outcome that was present for the positive case but absent for the control case. In a comparative group study, a balanced distribution of other causes seems to do the work that similarity in background causal circumstances does in a case-control study.

Of course, *complete* identity in all background causes, which the method of difference demands, is unlikely. Nonetheless, as J.L. Mackie (1965) argues, the method of difference is a logical ideal towards which scientists strive in their controlled experiments. Analogously, balance in each and every confounding cause, however improbable, seems to function as an ideal for our group comparisons, our confidence in the soundness of our causal inference increasing as the comparability of our groups increases. Embracing this idea, Howick recognizes that clinical trials are typically not sufficiently large to rule out *all* baseline confounders, but in response to Worrall (2002, 2007) he argues that this fact does not undermine the advantages of RCTs over observational studies. The former allow us to rule out a greater number of confounders than the latter, and it is on this basis that RCTs should be judged superior (Howick 2011a). Even though our randomized studies do not attain the ideal, if they are closer to the ideal than our non-randomized studies, then perhaps they are better after all.

The question that will presently occupy us is whether we should in fact hold up a logical ideal of balance in all confounding causes. In particular, is the balance assumption ever enough for sound causal inference in comparative group studies, and is it ever needed for sound causal inference? The first part of the question asks: is balance in all confounding causes *sufficient* for the conclusion that the exposure caused or prevented the outcome in a

---

former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon" (as quoted in Mackie 1980, 298).

[80] J. S. Mill would not agree; he denied that the Methods can be applied to group comparisons (Mill 1882; Morabia 2013).

study showing a difference in outcome between groups? The second part asks: is balance in all confounding causes *necessary* for the causal conclusion in a study showing a difference in outcome?

The following thought experiment will suffice to answer both questions. In the tradition of referring to clinical studies using a handy acronym, I will call it the MOD-G Study (Method of Difference Group Study). The studies I am about to describe could be controlled trials of an intervention, but they could just as easily be observational group studies examining any kind of exposure, harmful or beneficial. The MOD-G Study actually includes two group comparisons: MOD-G 1 and MOD-G 2.

6.3.1 MOD-G 1

MOD-G 1 will examine a hypothesis that any real comparative group study is designed to test: the hypothesis that the exposure (X) caused the outcome (Y). There are of course other causes of Y (confounding causes) for which the investigators must control. For simplicity's sake, we will restrict the number of relevant confounding causes to two: C_1 and C_2. Also to make matters simple, X, Y, C_1 and C_2 are all dichotomous variables; that is, each variable is either present or absent in an individual participant. Each variable is therefore measured as a frequency in the overall study group (i.e. as the absolute risk).

Finally, we will make two deterministic assumptions. The first assumption is that causes act deterministically, that the set of causes present for an individual fully determines whether or not that individual gets the outcome. This simplification will allow us to compare the finite frequency of Y between groups rather than the probability of Y, and rule out the possibility that any difference in frequency of Y between groups is due solely to chancy causation. The assumption that causes determine their effects is traditionally called 'determinism', and can be summarized by the slogan: same (complete) cause, same effect. But we can more precisely call it 'forward determinism' to distinguish it from a distinct deterministic assumption that we will also assume. The second assumption – call it, 'reverse determinism' – is that whether or not an individual gets the outcome fully determines whether or not there was a complete cause of the outcome. Reverse determinism adheres to the slogan: some effect, some (complete) cause. It discounts the possibility that any

difference in frequency of *Y* between groups is due solely to *Y*'s spontaneously popping into existence uncaused. Together, forward determinism and reverse determinism imply that any difference in frequency of the outcome between groups is proof of some relevant causal difference.

Despite these two simplifying assumptions, the lessons learned in this section will apply just as well to situations in which we are comparing probabilities. In those cases, rather than deterministic assumptions we must make the analogous assumption that the probability of *Y* is completely determined by the set of causes present; thus, any difference in probability of *Y* must be due to some relevant causal difference.

The MOD-G 1 investigators measure the frequencies of *Y*, C_1 and C_2 in a group exposed to *X*, as well as in an unexposed group. Their results are presented in Table 6-1.

Table 6-1. Comparative Group Study in which all Confounding Causes are Balanced				
	Y	*X*	C_1	C_2
Exposed	0.5	1.0	0.5	0.5
Unexposed	0	0	0.5	0.5

*Numbers are frequencies, groups are equal in size.

As is typical, the investigators have incomplete background knowledge. In fact, all they know is that *Y* is the effect of one or more of *X*, C_1 and C_2, which exhaust the variables that are plausibly causally relevant. Table 6-1 is similar in many respects to a Mill's method of difference table, but while a Mill table typically has a '+' or '-' representing the presence or absence of a factor for an individual, Table 6-1 includes a number representing the frequency of a factor in a study group.

Despite their ignorance of the relevant causal mechanisms, the investigators see from Table 6-1 that: (i) X is positively correlated with Y: an increased frequency of *X* is accompanied by an increased frequency of *Y*. They also see that: (ii) all confounding causes of *Y* are balanced in the study; each confounding cause is (perfectly) uncorrelated with exposure *X*. They suppose that: (iii) if all causes of *Y* are balanced except *X and X* is positively correlated with *Y*, then *X* must have caused *Y*. The investigators also happen to be

171

disciples of Mill, so they reason using a kind of method of difference inference scheme: from i, ii and iii, they conclude that X caused Y in the study.

At this point in our thought experiment, we will allow ourselves to be omniscient and find out what really happened at the individual level. In each participant, Y represents the presence of a clinically important protein biomarker. The presence of Y is fully determined by the conjunction of C_1 and C_2. The confounders C_1 and C_2 represent two other proteins, each coded by different genes. C_1 is a precursor for Y, while C_2 is the enzyme that catalyzes the conversion of C_1 to Y. The pathway is represented below (Figure 6-1).
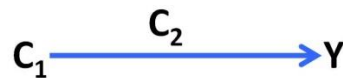
Figure 6-1. Mechanism producing Y in MOD-G 1.

X plays no part in this mechanism, which is the only mechanism that produces Y. Thus, X does *not* cause Y. How then can we explain the study results?

The key is that to say C_1 and C_2 are balanced is not to say all that much of use. Neither C_1 nor C_2 will cause Y without the other. Rather than the distribution of C_1 and the distribution of C_2 provided by Table 6-1, we need to know the distributions of $C_1 \& C_2$, $C_1 \& \neg C_2$, $\neg C_1 \& C_2$, and $\neg C_1 \& \neg C_2$. The frequencies for C_1 and for C_2 in Table 6-1 are consistent with a range of possible frequencies for these four conjunctions. It turns out that the actual frequencies are those reported in Table 6-2.

Table 6-2. Supplementary Data for Table 6-1						
	Y	X	$C_1 \& C_2$	$C_1 \& \neg C_2$	$\neg C_1 \& C_2$	$\neg C_1 \& \neg C_2$
Exposed	0.5	1.0	0.5	0	0	0.5
Unexposed	0	0	0	0.5	0.5	0

*Numbers are frequencies, groups are equal in size.

Table 6-2 reveals that in the exposed group, 50% of individuals were positive for both C_1 and C_2 ($C_1 \& C_2$), which explains the 50% frequency of Y in that group because C_1 and C_2 are jointly sufficient for Y. However, in the unexposed group, the 50% of individuals that were positive for C_1 (column '$C_1 \& \neg C_2$') were not the same 50% of individuals that were positive for C_2 (column '$\neg C_1 \& C_2$'). Because neither C_1 nor C_2 will cause Y without the other,

172

no one in the unexposed group was positive for *Y*. Exposure *X* plays no role in this causal story. Yet the researchers thought that it *must* because all confounding causes were balanced and there was a difference in outcome between the groups!

Even authors who doubt the likelihood of balancing all confounding causes in a randomized trial sometimes accept the sufficiency of this condition. For instance, Howson and Urbach suggest that a guarantee that the comparison groups are balanced for each prognostic factor "has at least the virtue that if it were true, then the conditions for an eliminative induction would be met, so that whatever differences arose between the groups in the clinical trial could be infallibly attributed to the trial treatment" (2006, 197). What the hypothetical MOD-G 1 study shows is that the balance assumption is not enough for sound causal inference. In a comparative group study with positive results, the finding that all confounding causes are balanced is *not* sufficient for inferring the causal conclusion. It may not be true that *X* caused *Y*, as demonstrated by the folly of our black box researchers.

6.3.2 MOD-G 2

Although the balance assumption is not sufficient, it might perhaps be necessary. It might be indispensable for sound causal inference, and thus a crucial consideration for triallists. To investigate this possibility, let us commission a second study: MOD-G 2. This time, let us hypothesize that another exposure (X^*) caused outcome *Y*. The researchers in this study are in much the same situation as before; they are told that C_1 and C_2 are the only plausible confounding causes and that they can make deterministic assumptions, but are given no other information. They measure the frequencies of X^*, C_1, C_2 and *Y* in a study population (Table 6-3).

Table 6-3. Comparative Group Study in which not all Confounding Causes are Balanced				
	Y	X^*	C_1	C_2
Exposed	0.5	1.0	0.75	0.5
Unexposed	0.25	0	0.25	0.5

*Numbers are frequencies, groups are equal in size.

Once more, the researchers observe that: (i) $X*$ is positively correlated with Y. They still maintain that: (iii) if all causes of Y are balanced except $X*$ *and $X*$ is positively correlated with Y, then $X*$ must have caused Y. However, it is now not that case that: (ii) all confounding causes of Y are balanced in the study; C_1 is extremely imbalanced. Thus, they do not conclude that $X*$ caused Y in the study. They were reasonable not to do so, given their epistemic disadvantage. We, on the other hand, are able to take a look inside the black box and find out what really happened.

The human pathway involving the four variables is represented below (Figure 6-2).
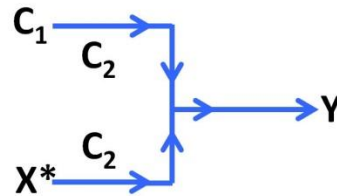
Figure 6-2. Mechanism producing Y in MOD-G 2.

As before, the C_2 enzyme always catalyzes the conversion of C_1 to Y. This time, $X*$ is very similar to C_1, so C_2 will also catalyze the conversion of $X*$ to Y. Table 6-4 explains the results of Table 6-3 in light of this mechanism.

Table 6-4. Supplementary Data for Table 6-3						
	Y	$X*$	$C_1\&C_2$	$C_1\&\neg C_2$	$\neg C_1\&C_2$	$\neg C_1\&\neg C_2$
Exposed	0.5	1.0	0.25	0.5	0.25	0
Unexposed	0.25	0	0.25	0	0.25	0.5

*Numbers are frequencies, groups are equal in size.

We know from the biological mechanism that what is most salient is the frequency of individuals who are positive for $C_1\&C_2$. This frequency is the same in both groups, $C_1\&C_2$ is perfectly balanced. In the unexposed group, the 25% of individuals who were positive for $C_1\&C_2$ fully account for the 25% frequency of Y. In the exposed group, the 25% of individuals who were positive for $C_1\&C_2$ accounts for half of all individuals who were positive for Y. The remaining 25% of Y-positive participants in the exposed group must have gotten the outcome via some other causal pathway. Since C_1 and C_2 on their own cannot

cause *Y*, the only other difference - *X** - must be involved. Indeed, *X** caused *Y* in the other 25% of participants with C_2 in the exposed group, who are found in the column '$\neg C_1 \& C_2$'.[81]

This reasoning reveals that the balance assumption is not needed, not necessary for sound causal inference. Even if the assumption fails, we might still be able to confidently infer that an exposure caused a difference in outcome between groups, if only we had the right information. In summary then, the balance assumption is not enough and not needed; in a study with positive results, the truth of the balance assumption is neither necessary nor sufficient for a sound causal inference.

Where did the black box researchers go wrong? The trouble was not with their first premise, the positive correlation between exposure and outcome; nor was it technically with their second premise, the balance assumption, which in the first study they correctly asserted and in the second study they correctly denied. The problem was with their third premise, a conditional that stated that if the first two premises are true, then the exposure caused the outcome. Notice that the overall argument is deductively valid (by modus ponens). Yet the argument is unsound because the third premise is not true. The trouble with their Mill's method of difference-type reasoning was their Mill's method of difference-type supposition, the conditional premise just described.

Salim Yusuf and colleagues note: "By using as an analogy experiments conducted in a test tube or in animals, it is often argued that all extraneous and confounding variables can and should be controlled" (1990, 77). Similarly, Annie Britton and colleagues suggest: "In the classical laboratory experiment, the effect of the variable of interest is isolated by controlling the values of other relevant variables…[T]here maybe a residual feeling that an RCT is a form of laboratory experiment" (1999, 117). Both articles argue that the quest to control all confounders is misguided, and perhaps arises through inappropriate analogy with classical Millian experiments.

We might wonder: what *can* we infer about a comparative group study with positive results in which we know that all confounding causes are balanced – but we know nothing else? I suggest that we can infer nothing of interest at all. The distribution of individual

---

[81] Depending on your intuitions about causal over-determination, you may want to say that *X** also caused *Y* in individuals in the exposed group with $C_1 \& C_2$ (both $C_1 \& C_2$ and $X^* \& C_2$ caused *Y*).

components of a causal pathway tells us nothing about the collective assemblages that regularly produce the effect, and confounding causes as we have illustrated them through examples - exercise, proteins - are no more than component parts. On its own, the balance assumption does not even lend inductive support to the causal conclusion. If I am handed a black box (this time, literally speaking) and told only that it contains a wheel and a rope, what justification do I have for believing it contains a pulley? You cannot build a pulley with only a wheel and a rope – the mechanism also requires an axle. If I lack any information about whether or not the black box contains this essential third component of the mechanism, it is arbitrary to assume that the mechanism is complete; I might just as well assume that it is incomplete. If you are inclined to believe that the balance assumption supports a causal conclusion given positive study results, perhaps you are making an important presupposition. For instance, perhaps you are presupposing that the complementary components of the causal pathway are evenly distributed in the population.

A balanced distribution of each confounding cause is the wrong logical ideal for comparative group study causal inference. Consequently, we should not appraise the epistemic worth of a group study – whether a randomized trial or non-randomized study - according to how closely it approaches that ideal. However, we are not done with confounders and causes just yet. As we will see in the next section, the idea of balancing 'other causes' is not too far off the mark.

6.4 Disjunction *C* and the Ideal Study

Returning to the confounding conundrum I posed at the outset, we accepted that in an RCT it is unlikely that all confounding causes are balanced among the study groups; the balance assumption probably fails. However, we need not worry yet because we have also seen that the balance assumption is not the ideal condition we might have thought it was for group study causal inference – it is neither necessary nor sufficient. Thus, RCTs may *not* fail to warrant the conditions for a causal inference.

Mill was familiar with the phenomenon of causal interaction – what he referred to as "intermixture" – in which causes modify the effects of one another (Morabia 2011, 298; Mackie 1980, 306). This idea describes what was happening earlier in the MOD-G Study.

The causes C_1 and C_2 interacted to produce Y. Rather than confounding causes, the thought experiment showed that comparing the frequencies of the $C_1\&C_2$ causal assemblage among the study groups allows us to make a sound causal inference. In this section, I will introduce two new concepts: 'complex causes' and 'C'. These concepts are useful for understanding comparative group study causal inference but also causation in epidemiology and the social sciences more generally.

6.4.1 The Ultimate 'Other Cause': C

It is well documented among philosophers as well as some epidemiologists (e.g. Rothman & Greenland 2005) that what we ordinarily call 'causes' are not complete causes of their effects, but are mere components of more complete causal conditions. This situation is absolutely characteristic of the causes that researchers might label as 'confounders' within a particular study. J.L. Mackie (1965, 1980) developed a nomenclature to describe the logical relations among causes and their effects. In Mackie's language, the conjunction A&B&C is minimally sufficient for D just when all components (A, B, C) are jointly sufficient for D, but no subset of the components (neither A, nor B, nor C, nor A&B, nor A&C, nor B&C) is sufficient for D (1980, 62). A conjunct in a minimally sufficient condition can also be a negated term (e.g. ¬A), representing an 'interfering factor' or 'counteracting cause'. What I will call a *complex cause* has similar properties; it can be described by a conjunction ($C_1\&C_2\&...$) of non-redundant single (negated or unnegated) terms, where $C_1\&C_2\&...$ is minimally sufficient for causing Y.

Although Mackie's minimally sufficient conditions fully determine the effect, our definition of a complex cause should not be understood as precluding the possibility of chancy causes, those that act indeterministically. A complex cause can be 'minimally sufficient for causing Y' even if the chance of Y occurring given the complex cause is less than 100% (think of a radioactive atom causing the stochastic emission of an alpha particle). As Papineau (1985) suggests, sufficiency can be expressed here in terms of the cause determining the *chance* of the effect rather than the effect itself.[82]

---

[82] I also do not intend to rule out the possibility that Y is a quantitative variable. Cartwright (2012) adapts Mackie's definition to account for quantitative or multivalued effect variables:

Confounding causes are generally components of complex causes. They are represented by the conjuncts. On their own, confounding causes are insufficient for causing the outcome in a study. Neither precursor proteins, nor enzymes, nor exercise, nor placebos are complete causal conditions. This fact explains why confounding causes are only derivatively important; they are important insofar as they constitute complex causes, but it is controlling for complex causes that is of ultimate importance in causal inference.[83]

Both Mill and Mackie were aware that there exists a "*plurality of causes*" for any phenomenon (Mackie 1980, 307). There are many complex causes - each uniquely constituted - of an effect Y. In the MOD-G Study, we considered only one complex cause that excluded the exposure (C_1&C_2). In any real study, there will be a great many more. At first glance, the plurality of complex causes might pose a problem for the comparability of our groups. If something like the balance assumption - but for complex causes rather than confounding causes - is desired, then we run into the same initial problem that plagued the balance assumption. Namely, given the number of unique complex causes in a group study, it might be improbable that each unique complex cause is balanced among the groups.[84]

Fortunately, what is needed in a sound group study is something much weaker than a balancing of each unique complex cause. To help understand why, let us define a disjunction C, where C includes all unique complex causes of Y *except* those that involve X as a conjunct: $C = (C_{1,1}\&C_{1,2}\&...) \lor ... \lor (C_{n,1}\&C_{n,2}\&...C_{n,m})$[85]. The conjuncts in this general

---

instead of being sufficient for the effect, a complex might be sufficient for producing a *contribution to the value* of the effect. If the black box researchers measured Y as the blood concentration of the protein biomarker, the C_1 precursor and C_2 enzyme might be jointly sufficient for raising the blood concentration of the biomarker by a certain amount.

[83] Judea Pearl (2009) examines traditional no-confounding criteria and similarly challenges the sufficiency and necessity of the assumption that all potentially relevant variables are unassociated with the exposure. Recognizing that confounding sometimes arises when variables work in groups rather than singly, Pearl proposes as a solution the notion of a "*non-trivial sufficient set*" (195) that bares similarities to our idea of a complex cause.

[84] Paul Thompson (2011b) argues a similar point. If there are 11 quantitative traits relevant to the outcome, each with 6 possible values, the number of unique possible combinations is 6^{11} = 362,797,056. He notes that this number is greater than the population of the United States in 2009. Thus, even in an RCT enrolling a population the size of the United States, the two groups would not be balanced for each unique combination (assuming that all or most possible combinations are present in the population).

[85] Or $C = \lor_i \land_j C_{ij}$. I am assuming that Y is a dichotomous variable. For quantitative effect variables there is an analogue of the dichotomous C term, which we can call 'quantitative C'.

formula (e.g. $C_{1,1}$) are potential confounding causes, while the disjuncts (e.g. $C_{1,1}\&C_{1,2}\&...$) are complex causes. $C_{1,1}$ is the 1st confounding cause of the 1st complex cause, while $C_{n,1}$ is the 1st confounding cause of the nth complex cause ($C_{n,1}$ may or may not be identical to $C_{1,1}$). Since C is a disjunction of complex causes, it is satisfied whenever an individual has any complex cause of Y (except those containing X, by stipulation).[86] C is the ultimate 'other cause'.

Assuming determinism, what matters for sound causal inference is not the distribution of each unique complex cause, because any complex will cause Y. What matters is the distribution of participants satisfying C, which abstracts away the particulars.

Think back to MOD-G 2 and Table 6-4. The complex cause $C_1\&C_2$ was perfectly balanced between the study groups, which allowed us to causally attribute the difference in outcome to the exposure (X^*). Since $C_1\&C_2$ was the only complex cause of the outcome that did not include the exposure, we can fill in the general formula for C using only one conjunction: $C = C_1\&C_2$ for this exposure and this outcome. If instead there were two complex causes, it would not matter if each complex was greatly imbalanced so long as the disjunction of both complex causes (C) was not.

6.4.2 The Ideal Comparative Group Study

In the MOD-G Study, causal conclusions follow deductively from: (i) the positive association between exposure and outcome, (ii) the premise that C is distributed equally between groups, and (iii) our two deterministic assumptions: forward determinism and

---

Quantitative C is a function of C_{ij}, while the quantitative Y variable is a function of quantitative C. We should expect that the form of the Y function will vary depending on the outcome that Y represents; for example, Cartwright (2012) models the relationship between Y and what I call quantitative C (she calls it 'W') with a linear equation. We should expect that the functional relationship between quantitative C and the C_i's will also depend on the outcome.

[86] The stipulation that C not include X is consistent with the observation that what normally counts as a confounding cause is relative to the study. A study will designate a certain factor of interest as the exposure (X), which will then determine what factors are to be considered potential confounding causes. One study's exposure is another study's confounding cause.

reverse determinism.[87] To see how, consider Table 6-5, representing ideal conditions for a comparative group study causal inference.

Table 6-5. Ideal Conditions for a Comparative Group Study Causal Inference			
	Y	X	C
Exposed	y_X	1.0	z
Unexposed	$y_{\neg X}$	0	z

*Numbers and variables are distributions, $y_X > y_{\neg X}$.

We need only consider three factors: Y, X and C. Forward determinism ('same (complex) cause, same effect') ensures that every instance of a complex cause produces the outcome, and thus the frequency of Y in the unexposed group can be no less than the frequency of C. Reverse determinism ('some effect, some (complex) cause') guarantees that every instance of the outcome is produced by a complex cause, so that the frequency of Y in the unexposed group can be no more than the frequency of C. Together, these two deterministic assumptions imply that the frequency of C in the unexposed group is equal to the frequency of Y: $z = y_{\neg X}$. Because C is perfectly balanced between the groups, the frequency of C in the exposed group is also equal to $y_{\neg X}$.

The final special feature of the ideal study represented by Table 6-5 is that the frequency of the outcome is greater in the exposed group compared to the unexposed group ($y_X > y_{\neg X}$). In other words, the study shows what we might call a 'positive result'.[88] Recall that we already worked out that the frequency of C in the exposed group is equal to $y_{\neg X}$ in our ideal study. Therefore, in the exposed group the frequency of Y ($= y_X$) is greater than the frequency of C ($= y_{\neg X}$). There must be some individuals who got the outcome Y but lacked a complex cause in C. According to the assumption of reverse determinism, these outcomes must have had *some* cause. If not C, each of these outcomes must have been caused by one of

---

[87] Comparative group study logic exemplifies eliminative inference because – as we are about to see – we conclude that the exposure caused the outcome by eliminating all other (complex) causal explanations for the difference in outcome. Mackie (1980) argued that all eliminative inferences are deductive.

[88] The finding that $y_X < y_{\neg X}$ could also be regarded as a 'positive result', suggesting that X *prevents* Y. I do not have space to discuss the logic of prevention here.

the complex causes excluded from C: complex causes involving X. Thus, exposure X is a cause of the outcome. Furthermore, we can causally attribute the difference in outcome to the exposure.

I have relied on a deterministic ideal only to make vivid the deductive validity of group study causal inference. Our reasoning is equally valid if we allow for indeterministic causation and measure the probabilities of Y and C. Instead of considering the distributions in Table 6-5 as frequency distributions, we can consider them as probability distributions. Instead of forward determinism, assume that the probability of Y is completely determined by the set of causes present for an individual (the probability of Y is zero in the absence of a complex cause and greater than zero in the presence of a complex cause). Then any difference in probability of Y between groups must be due to some relevant causal difference. If C is distributed equally (C is probabilistically independent of the exposure), then the relevant causal difference must involve the exposure, and we can causally attribute the difference in probability of the outcome to the exposure. In measuring the probability of C rather than the probability of each complex cause in C, we are assuming (for simplicity) that each complex cause determines the same probability of Y. I will have more to say about this assumption shortly.

Balance in all confounding causes is a confounded (confused) logical ideal. On the other hand, an equal distribution of C among study groups is an unconfounded ideal, free from confusion *and* free from epidemiological confounding. In a comparative group study showing an association between exposure and outcome, the premise that C is distributed equally between groups is sufficient for our causal conclusions (so long as we make the necessary metaphysical assumptions).[89]

Cartwright (2010) also articulates sufficient conditions for causal inference to undergird her 'ideal RCT'. She starts by specifying subpopulations defined according to their causally relevant factors (confounding causes). Each subpopulation is represented by a conjunction over all relevant confounding causes, with each cause appearing as a negated

---

[89] If C is distributed equally and there is *no* difference in outcome between groups, it is not necessarily true that the exposure did not cause the outcome. It could be the case - however unlikely - that the exposure caused the outcome in some participants and prevented the outcome in just as many participants.

term (if absent) or an unnegated term (if present). In our MOD-G thought experiment, in which there were two confounding causes, there were four unique subpopulations: $C_1\&C_2$, $C_1\&\neg C_2$, $\neg C_1\&C_2$, and $\neg C_1\&\neg C_2$. According to Cartwright, "In an ideal RCT each K_i [subpopulation] will appear in both [study] wings with the same probability" (2010, 64). Then if there is a higher probability of the outcome in the treatment group ($p(Y|X) > p(Y|\neg X)$), these conditions entail that the treatment causes the outcome in at least one of the subpopulations. In MOD-G 2, X* caused Y in the $\neg C_1\&C_2$ subpopulation (X*&C_2 was a complex cause of Y).

Cartwright's conditions for causal inference are sufficient by the lights of my account as some of the subpopulations will contain complex causes in C, and these subpopulations will each be equally distributed between groups. However, Cartwright's ideal RCT is more demanding than my ideal study (Table 6-5) for two reasons. First, it requires that unique subpopulations *not* containing complex causes in C are each distributed equally, which is not needed. In MOD-G 2, two of the subpopulations not containing complex causes were greatly imbalanced (Table 6-4), yet we could still conclude that the exposure caused the outcome. Similarly, in Cartwright's ideal RCT each unique subpopulation containing a complex cause in C is distributed equally, which is also stronger than needed. If we assume that each complex cause determines the same probability of Y, all that matters is that the probability of C is the same for both groups.[90]

If we allow that two unique complex causes might fix two unique probabilities of the outcome (a reasonable allowance), then we should instead demand that 'C's contribution to the probability of Y' is the same for both groups. C's contribution to the probability of Y is the average probability of Y among 'C subpopulations', multiplied by the total probability of C subpopulations. We can think of it as the force that C exerts on Y. If C's contribution to the probability of Y is the same for both groups yet the total probability of Y – the *net* force on Y - is greater in the exposed group ($p(Y|X) > p(Y|\neg X)$), then X is causally responsible for

---

[90] In her (2012), Cartwright describes the causal principle governing the RCT population using a linear equation in which *Y* is a function of *X* and *W*. *W* represents the net contribution of causes that are not interacting with *X*, and is thus analogous to my dichotomous C variable. The ideal conditions that Cartwright stipulates include the assumption that *X* is probabilistically independent of *W*, which is equivalent to my requirement that the probability of C is the same for both groups.

this difference in probability.[91] However, 'C's contribution to the probability of Y' is harder to visualize. So for the rest of the paper I will refer to the condition that 'C is distributed equally between study groups' as the key requirement of an ideal study. This condition is good enough whenever C acts deterministically or whenever each of C's complex causes determines the same probability of the outcome; but whenever two unique complex causes fix two unique probabilities of the outcome, we should instead demand that C's contribution to the value of Y – as previously defined – is the same for both groups.

6.4.3 Required Conditions for Causal Inference

So far I have argued that an equal distribution of C is sufficient for sound causal inference in a positive group study (with the above caveat). I have not shown that a perfectly balanced distribution of C is *required* for the causal inference, and in fact it is not. To conclude that X caused Y, all we require is that C's distribution is *less unequal* than Y's distribution. In the probabilistic case, the ratio of C's probability in the exposed group to its probability in the

---

[91] The proof of this principle requires some work. We must first define a 'C subpopulation' as a homogeneous subpopulation with at least one complex cause in C and without any complex causes that are not contained in C. We can then define 'C's contribution to the probability of Y ($C_{p(Y)}$)' as the (weighted) average probability of Y among C subpopulations, multiplied by the total probability of C subpopulations. If there are two unique complex causes ($complex_1$ and $complex_2$):
$C_{p(Y)} = p(Y|C) \times p(C)$
$C_{p(Y)} = [p(Y|complex_1)p(complex_1) + p(Y|complex_2)p(complex_2)] \times [p(complex_1) + p(complex_2)]$.
The total probability of Y in a study group is the sum of C's contribution to Y and ¬C's contribution to Y, or:
$p(Y) = p(Y|C)p(C) + p(Y|¬C)p(¬C)$
$p(Y) = C_{p(Y)} + p(Y|¬C)p(¬C)$.
The $p(Y|¬C) = 0$ in the *unexposed* group because the ¬C partition of the unexposed group contains no complex causes (none in C, none involving X), and - by the assumption of reverse determinism - Y can only ever occur when it is caused. If $C_{p(Y)}$ is the same in both groups yet the total probability of Y is greater in the *exposed* group, then $p(Y|¬C) > 0$ in the exposed group. This is only possible if the ¬C partition of the exposed group contains some complex causes involving X, in which case X causes Y in one or more subpopulations of the exposed group.

unexposed group must be less than the ratio of Y's probabilities ($p(C|X)/p(C|\neg X) <$ $p(Y|X)/p(Y|\neg X)$).[92]

To see this, we can insert any values we like for Y and C in Table 6-5 (easing the requirement that C is perfectly balanced), and suppose any values we want for the probability of Y given C ($p(Y|C)$) and the probability of Y given $\neg C$ ($p(Y|\neg C)$). Let's say that there is a twofold difference in the probability of Y between groups, and that the probabilities of Y are 0.2 in the exposed group and 0.1 in the unexposed group. Let's also say that the probability of Y given C is 0.5 and that the probability of Y given $\neg C$ is 0. The probability of C in the unexposed group must then be 0.2 to account for the 0.1 probability of Y.[93] If there is a twofold difference in the probability of C between groups, then the difference in C fully accounts for the twofold difference in Y[94] (the probability of C in the exposed group would be 0.4, which accounts for the 0.2 probability of Y). But if the imbalance in C is less than twofold, then the difference in C cannot fully account for the difference in Y, which leaves only the exposure to account for part of the difference.

Although it may be unrealistic and unnecessary for sound causal inference, an equal distribution of *C* serves an important function as a regulative ideal for the *design* of a comparative group study. The various techniques and tricks used in comparative studies – randomization, double-blinding, stratification, matching – are an attempt to bring the distribution of C closer to the ideal, so that we can feel more confident that any difference in outcome is due to the exposure. However, we should not despair that real studies typically fall short of the ideal; all that is required in the *interpretation* of the study's result is that C is less unequally distributed than the outcome.

---

[92] I am again assuming that each complex cause in C determines the same probability of the outcome, and that the probability of the outcome in the absence of a complex cause is zero (reverse determinism).

[93] $p(Y) = p(Y|C)p(C) + p(Y|\neg C)p(\neg C)$
$0.1 = (0.5)p(C) + (0)p(\neg C)$
$0.2 = p(C)$

[94] In a notable example of epidemiologists reasoning in this manner, Richard Doll and Richard Peto (1980) write that selection bias can plausibly account for a twofold difference in a study's outcome. In the next section, we will see how selection bias influences C's distribution.

In summary, rather than a balance in confounding causes, group study causal inference depends upon an equal distribution of *C*. We should have more confidence in our causal inference whenever we are more confident that C is less unequally distributed than the outcome. Any of the complex causes in *C* are sufficient for causing the study outcome, and – by stipulation - none of the complex causes include the exposure. Meanwhile, confounding causes are generally only conjuncts of complex causes in *C* – on their own they are *not* sufficient for causing the outcome. Are philosophers of science then wrong about the importance of confounders in group studies? We will now see that they are not wrong; they have simply exaggerated the importance of confounders *as causes*. Recognizing the importance of confounders as correlates of C helps clarify the role of randomization in group studies.

6.5 Confounders as Causes, Confounders as Correlates

Though we have relieved the tension between the implausibility of balancing each confounding cause and the apparent need to do so in a group study, there is something left unresolved: why statisticians are so worried about confounding variables. So far we have examined the role of confounders as causes of the study outcome, a role that philosophers often assume in their accounts of RCT causal inference. In this section, we will analyze a concept of confounder distinct from the 'direct causal concept': *confounders as correlates*. More precisely, confounders are important as correlates of *C*.

Earlier, I noted that philosophers typically cast confounders as causes, but that some of our stock examples of confounders do not seem to fit the bill. For example, age and health status are not straightforwardly causal variables. They are however associated with outcomes of interest; for instance, older patients are more likely to have atherosclerosis that accumulated in their arteries over time, and are thus more likely to have a heart attack or stroke.

In contrast to the philosophers, methodologists in epidemiology and the social sciences have historically worried about confounders not simply because they are (sometimes) causes but crucially because they are correlates of the outcome. Kenneth Rothman and Sander Greenland state: "In general, a confounder must be associated with both

the exposure under study and the disease under study to be confounding" (1998, 120). Meanwhile, the *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice* defines a confounder as: "A factor that is associated with the outcome of interest and is differentially distributed in patients exposed and unexposed to the [exposure] of interest" (Guyatt et al. 2008, 777). The textbook discusses the importance of balancing prognostic factors (potential confounders): "If prognostic factors—either those we know about or those we do not know about—prove unbalanced between a trial's treatment and *control groups*, the study's outcome will be biased" (Guyatt et al. 2008, 70). In other words, confounders can lead us to falsely conclude that the exposure caused or prevented the outcome when in fact it did not.

Where must a factor be 'associated with the outcome of interest' before it qualifies as a confounder? Presumably, in the comparative group study if its imbalance is to bias the study's results. However, an association between an imbalanced variable and the study outcome is not enough. If the treatment causes an excess of the outcome in the treatment group, any imbalanced variable will be automatically associated with the outcome; but the result is not thereby biased. What the *Users' Guides* are truly worried about are variables that are associated with instances of the outcome not caused by the treatment. But to say this is just to say that they are worried about variables associated with instances of the outcome caused by complex causes in C. In fact, we often suspect that a certain variable might be a 'confounder' in a comparative study like a trial after we have observed its association with the outcome in a prognostic study, a study searching for correlations between variables like age and outcomes like heart attack in a population that is *not exposed* to the trial treatment. If we assume that the heart attacks in this untreated population all had some cause, those causes must be found in C. Thus, if age is associated with the outcome in this untreated population, it is associated with C. If the association between age and C also holds in the trial and age is imbalanced between trial groups, we should worry that C is imbalanced. In short, confounders like age are important not as causes of the outcome but as correlates of C.

I have referred to the concept of a confounding cause as a 'direct causal concept' because on this interpretation a confounder is causally relevant to the outcome in a direct sense: it causes the outcome. While the distinct concept of 'confounder as a correlate of *C*' is clearly an associational concept, unlike some associational definitions of a confounder that

attempt to do away with causation (as much as possible), our associational concept should not be considered non-causal because *C* is a causal variable. However, it is neither a *direct causal* concept because a correlate of *C* (such as age) does not necessarily cause the outcome. Instead, we can consider it an *associational-causal* concept of confounder: confounding variables are factors *associated* with complex *causes* of the outcome in *C*. To avoid ambiguity, we might reserve the term 'confounding factor' for the direct causal concept, and label the associational-causal concept with the term 'covariate' or 'prognostic factor'. While the *Users' Guides* (2008) treats 'prognostic factor' and 'confounding factor' terms as synonyms, the textbook *Evidence-Based Medicine* (Straus et al. 2011) distinguishes confounders, as causes of the outcome, from prognostic factors, as correlates of the outcome that need not be causes of the outcome. From now on, I will exercise a similar distinction between confounding causes and prognostic factors. (Of course, many variables are both a prognostic factor/covariate and a confounding cause.)

Though I have argued that the importance of balancing confounding causes in a group study has been exaggerated, I do not wish to deny a role for the concept of confounding causes in our causal reasoning. It may be background knowledge that reveals a potential imbalance in *C* rather than epidemiologic data. The study investigators may believe based on biomedical theory that the variable is a potential confounding cause of the outcome, and moreover, that the other confounding causes needed to build complex causes are common. In a study of the side effects of a new drug in an older patient population, they may notice that a higher proportion of patients in the drug group report frequent alcohol consumption compared to the control group. They know that alcohol sometimes interacts with drugs to produce side effects/adverse events. Because older patients are often prescribed multiple medications, the investigators might further suspect that complex causes involving alcohol and prescription drugs but having nothing to do with the experimental drug might be overrepresented in the experimental group. Their background knowledge of the interactions involving alcohol as a confounding cause might thus alert them to a potential imbalance in *C*.

How might a prognostic factor, a correlate of C wind up imbalanced between the study groups? One mechanism involves chance: in a randomized study, randomization can throw up a highly unequal distribution. There are also non-random or systematic ways in which covariates come to be imbalanced. In a non-randomized trial or in an observational

group study, if the investigators or care providers can select which patients will receive the treatment of interest and which patients will not receive the treatment, for conscious or unconscious reasons they might treat patients with the new drug that are on average younger or healthier. Then age or health will be associated with the drug, and as a result the drug might be associated with C. Or, if patients can self-select for treatment, then patients with a sedentary lifestyle may be less likely to opt for a new exercise regime, and sedentary lifestyle will be (negatively) associated with exercise (which might consequently be associated with C). In either case, it is said that the study suffered from 'selection bias'.

Properly executed randomization prevents selection bias in a controlled trial by preventing selection.[95] Study investigators and participants are prevented from deciding group assignment; instead, group allocation is determined by a random process. Thus, randomization militates against systematic imbalances in C by barring systematic imbalances in prognostic factors at baseline. Of course, the possibility of a large baseline imbalance due to chance remains, and as we saw in section 2 a chance imbalance is probable if we think that the number of relevant variables is great enough. Fortunately, just as we should not worry about balancing all confounding causes, we should not be concerned about the dismal prospects of balancing all of C's correlates, of which there will be many. Like confounding causes, prognostic factors – as correlates of C – are only derivatively important. They serve an epistemic function, alerting us to imbalances in C. It is the distribution of C itself that is of ultimate concern in causal inference, and C is only one variable. Once we have controlled for all systematic sources of imbalance in a trial (through randomization and subsequent methods), the chance that C is distributed roughly evenly between groups is relatively high so long as the trial population is relatively large.

In an observational group study, the investigators typically do not dictate which patients receive the exposure because usually the data are collected in routine practice.

---

[95] After dismissing the argument that randomization controls for all confounding variables, Worrall accepts the claim that randomization controls for selection bias, calling it "a cast-iron argument for randomization" (2002, S325). Worrall then qualifies his agreement, claiming that randomization is merely one means to this end. La Caze (2009) agrees that randomization blocks certain sources of selection bias, but objects that alternative means of achieving this end through other unbiased allocation mechanisms are not available to observational studies.

Although they cannot prevent provider selection or patient self-selection, observational studies are not defenseless against selection bias. They can match the study groups for similar distributions of prognostic variables, stratify the study groups according to prognostic variables and analyze results within strata, or make statistical adjustments to the data based on observed imbalances in prognostic variables. However, as the common refrain goes: these methods can only control for variables that are observed and that we suspect are relevant. The methods leave open the possibility of a hidden association between exposure and C that is not predicted by the observed, suspected prognostic variables. If we get a positive study result we might therefore lack confidence that C is less unequally distributed than the outcome, or might have difficulty assessing how confident we can be. How worried we should be about unforeseen selection bias should depend on how complete our knowledge is about the relevant variables, which will vary by circumstance.

A virtue of my account of comparative group study causal inference is that it proposes a common logic for randomized and non-randomized studies. In comparison, a causal Bayes nets account of causal inference in clinical research that includes an 'ideal intervention' on the exposure variable (e.g. Steel, 2011) is more straightforwardly applicable to a randomized study compared with an observational group study because it requires that patient characteristics (endogenous variables) do not influence exposure. A unified account of comparative group study causal inference allows us to compare randomized studies with non-randomized studies directly (in general, and in particular cases). We can compare our confidence in studies that are quite distinct by holding them up to a singular ideal. In the design of any comparative group study, we strive towards the ideal of an equal distribution of C among study groups. In interpreting the data from any study, our causal inference is secure if the distribution of C is less unequal than the distribution of the outcome. The diverse methods of RCTs and observational studies – randomizing, matching - can be seen as promoting a balanced distribution of C or as correcting for imbalances in C (via prognostic factors).

When we examine the ideal, we see that the matter of whether or not a study was randomized is not a premise in our causal inference. Acknowledging that randomization is not necessary for sound causal inference – as well as recognizing the conditions that *are* necessary – might help to temper overzealous RCT worship. RCTs may in general bring us

189

closer to the comparative group study ideal, but even if this is so we should not let ideal be the enemy of sufficient, especially when other considerations (ethical, economic) are at stake. We can make a causal inference in any comparative study when the study groups are sufficiently comparable (regardless of how we go about generating and assessing comparability). In understanding just what kind of comparability is needed, confounding causes are merely pixels in a more complex causal picture.

6.6 Homogeneity and the Design of Randomized Trials

Having examined the requirements and methodology of sound causal inference in comparative group studies, we will now investigate a common practice in trial research to see whether it is justified. As is well documented, clinical trials – namely, efficacy trials - impose numerous exclusion criteria on prospective participants, rendering a large percentage of patients ineligible for the study. One common exclusion criterion that is highly significant for chronic disease care is the exclusion of patients with chronic comorbidities. A systematic review by Harriette Van Spall et al. (2007) discovered that 81% of RCTs exclude patients with common medical conditions. These exclusions result in the highly *unrepresentative trials* that many medical commentators worry about. Exclusions thus threaten the external validity of trials, especially for patients with *multimorbidity*. As I earlier noted, individuals with multimorbidity are paradoxically some of the most common yet least studied patients. By excluding vast subgroups of patients from randomized trials, we sacrifice information about the effectiveness and safety of our treatments for a large portion of the population. So why do we do it?

Britton and colleagues (1999) reviewed exclusions from RCTs and discussed several potential reasons, including: prior knowledge about the treatment effect in excluded subgroups (lack of equipoise), concern about risk of serious adverse events in excluded subgroups, increased statistical precision of the effect size estimate,[96] and reduced risk of bias. I cannot discuss all of these considerations here, but will focus on the final rationale.

---

[96] Britton and colleagues (1999) note an important inconsistency. In arguing that excluding certain subgroups improves statistical precision, authors assume heterogeneity of treatment effects (which influences variance). Yet those who extrapolate the effect size to the excluded

Researchers who argue that excluding certain patients reduces risk of bias often acknowledge that exclusions also reduce generalizability (or at least increase the risk of non-generalizability). But narrow eligibility criteria are defended nonetheless as priority is given to internal validity or unbiasedness. Charles Hennekens and Julie Buring (1987) argue that "[internal] validity should not be compromised in an effort to achieve generalisability, because generalisability can be inferred only from a valid result". Similarly, David Moher and colleagues argue: "Internal validity, the extent to which the design and conduct of the trial eliminate the possibility of bias, is a prerequisite for external validity: the results of a flawed trial are invalid and the question of its external validity becomes irrelevant" (2010, 20). This potentially appealing argument for excluding patients with comorbidities is that doing so is part of a necessary trade-off; we must sacrifice external validity for internal validity. To caricature the trade-off argument, in order to get a true conclusion about the target population out (external validity), we need to put in a true conclusion about the study population (internal validity). Thus, we must concern ourselves first and foremost with internal validity or else questioning whether we have external validity is 'irrelevant'; there is no sense in asking whether a wrong answer is generalizable.

Although this argument wisely warns of the danger of sacrificing internal validity, it unwisely deemphasizes the importance of external validity. While a substantially biased answer to the 'question of external validity' is unhelpful, an unbiased but non-generalizable answer to the question of internal validity is equally unhelpful because the question of internal validity is not the one to which physicians and patients outside of the trial ultimately want an answer. Hence, the results of an internally valid but externally invalid trial might also be irrelevant (and perhaps misleading).

Another problem with the argument is that it assumes that exclusion criteria do in fact improve internal validity. One thought is that certain patients are more likely to drop out from a study, and that drop-outs can bias the study results (Britton et al. 1999). Perhaps the patients who drop out do so disproportionately in the treatment group, and perhaps they are also correlated with C, so that their dropping out leads to a great imbalance in C. Intention-to-treat (ITT) analysis, in which a patient's outcomes are analyzed in the group to which they

---

patients assume homogeneity of treatment effects. In the next chapter, I will show that the assumption of homogeneity of treatment effects is prevalent and problematic.

were originally assigned, is supposed to mitigate this problem. Indeed, if patients are followed-up in their original group assignment then the distribution of *C* will not change. A detailed discussion of the problem of drop-outs and the solution of ITT analysis is beyond the scope of this chapter.

A distinct concern – one that I will investigate in more detail - is that including heterogeneous patients in the trial could compromise internal validity. According to La Caze, "Improvements in internal validity are achieved…by excluding participants who will complicate the analysis, that is, by ensuring the experiment is conducted on a relatively homogenous group of patients" (2009, 519). La Caze recognizes that ensuring external validity in these studies is an important issue for EBM. Similarly, Alvin Feinstein suggests, "By excluding patients with associated diseases from trials of therapy for a particular disease, statisticians can design the trial to contain a presumably 'homogeneous' population, but the results achieved in these 'purified' circumstances cannot be extrapolated to the heterogeneous world of clinical reality" (1970, 456). And Gudrun Bornhoft and colleagues write:

> To obtain a high IV [internal validity]…the study population should be as homogeneous as possible, while in evaluating EV [external validity] it is of great interest to what extent the intervention is also applicable among a heterogeneous population and under heterogeneous conditions, particularly with concomitant diseases and co-medications. Homogeneity within a group is usually attained by restrictive inclusion and exclusion criteria, homogeneity and comparability between groups by randomisation. With regard to IV it is the best method since randomisation is the only adequate means to reduce the risk of the unequal distribution of unknown confounding factors" (2006, pg. 4).

These authors believe that, while the practice might create problems for external validity, distilling a homogeneous trial population improves internal validity. Britton and colleagues suggest that this reasoning might be influenced by the tendency to treat RCTs as classical laboratory experiments that compare "homogeneous experimental units" (1999, 117). Yusuf and colleagues (1990) agree, and argue that homogeneous trial populations are both unnecessary and elusive. The demand for trial homogeneity may be driven by the desire to design a method of difference group study like the MOD-G study of section 6.3.

Let us apply my account of confounders and causal inference to the homogeneity rationale for excluding patients with comorbidities from trials. Homogeneous populations are those that have low variability in terms of potential confounding factors. In order to decrease

variability, we exclude patients with various comorbidities (who are often prescribed variable drug regiments) - we restrict the number of potential confounders. By doing so, we improve the chance that randomization will balance all potential confounders at baseline and generate homogeneous comparison groups. As we saw in section 6.2, under the assumption of statistical independence the probability that all potential confounders are balanced in an RCT at baseline (p('all')) is equal to the probability that one potential confounder is balanced (p('one')) raised to the number of potential confounders (n); or, $p('all') = (p('one'))^n$. Excluding unique potential confounders from the trial decreases n, which increases p('all'). Recall that for a p('one') of 0.95 and an n of 14, p('all') = 0.49. Excluding one, two, or three of these 14 potential confounders from the trial would result in a new p('all') of 0.51, 0.54 and 0.57, respectively.

This procedure only improves internal validity if balancing all potential confounders in a trial prevents us from misattributing a difference in outcome to the exposure. Yet as we have seen balance in all potential confounders – whether they are understood as confounding causes or as prognostic factors - is not what is ultimately needed for sound causal inference. Instead, we should strive towards a balanced distribution of C. If the study outcome is dichotomous, each multimorbid patient – despite their unique constellation of comorbidities and medications – will either have C or not. C is satisfied if and only if a patient has one or more complex causes. Decreasing the number of unique comorbidities in the study might decrease the total number of complex causes, but this will only serve – at best - to decrease the number of patients with C. Because C is just one variable, decreasing the number of patients with C will not increase the probability that C is balanced in the same way that decreasing the number of unique confounding causes (n) will increase the p('all') according to the above equation.

I do not mean to suggest that triallists should be unconcerned with imbalances in comorbidities. Certain comorbidities (or combinations of comorbidities) might be important prognostic factors, strong correlates of C. We can identify these factors through prognostic studies that measure their association with the outcome in untreated populations. Instead of excluding patients with these comorbidities from trials, whenever possible we should stratify the trial population according to the most important prognostic variables before randomization, rerandomize upon finding an imbalance in any of these variables after

randomization (when possible), or make statistical judgements to the data to correct for imbalances. It might be impractical to ensure that *all* combinations of comorbidities are balanced - but thankfully sound causal inference is not so demanding. Therefore, the desire to achieve trial population homogeneity is no defense of excluding patients with multimorbidity from randomized trials.

Of course, there are often other reasons for exclusion criteria, as I have noted. Notwithstanding other plausible reasons, we should not exclude multimorbid patients simply to decrease the number of unique potential confounders in the study. Understanding the proper role of confounders (not as causes, but as correlates of *C*) helps us to evaluate the soundness of routine clinical trial practices like this one.

6.7 Conclusion

It is unlikely that a randomized trial – even if well designed and conducted – will achieve a balanced distribution for each potential confounding variable. Yet balance in all confounders, construed as confounding causes of the study outcome, is sometimes held up as a logical ideal, our confidence in our causal inference increasing as our comparative group study approaches the ideal. It turns out that the balance assumption is a false idol. Instead, we must worry about the distribution of complex causes of the outcome that do not involve the study exposure (those included in C). In the ideal comparative study, C is distributed equally among study groups. In any real study, if C is less unequally distributed than the outcome then we can conclude that the exposure caused the outcome. Confounders are primarily important not as causes of the outcome but as correlates of *C* (prognostic factors/covariates) that may alert us to an imbalance in *C*. Randomization prevents systematic imbalances in prognostic factors at baseline, which may improve the comparability of our study groups. But it is the comparability of our study groups with respect to complex causes in C that is of direct relevance to comparative group study causal inference. This recognition might temper unwarranted RCT worship.

Applying these ideas to the homogeneity rationale for excluding patients with multimorbidity from RCTs shows the rationale to be wanting. Instead of excluding chronic comorbidities because they are potential confounding causes, triallists should identify

comorbidities (or combinations of comorbidities) that are strong correlates of C and use other techniques to safeguard their balance in a randomized trial. Utilizing less restrictive eligibility criteria would help with the prevailing problem of unrepresentative trials in medicine.

Chapter 7 - Myths and Fallacies of Simple Extrapolation in Medicine

7.1 Simple Extrapolation

My concern in this chapter is with an influential approach to therapeutic prediction that I will call *simple extrapolation*. It is an approach that has been advocated by EBM from nearly the beginning of the movement. Philosophers of science have criticized EBM – also from the beginning. Most resonant among the criticisms have been those objecting to EBM's privileging of RCTs above other kinds of evidence, discussed in the previous chapter. Simple extrapolation is equally consequential and deeply problematic, yet its faults have been mostly overlooked in the philosophy of science, as well as in medicine, public health and health policy. Simple extrapolation is anathematic to EBM's own aspirations of a medicine well-grounded in evidence, despite the support this approach has received from EBM.

Recall that RCTs are the clinical research studies most trusted in EBM and evidence-based policy because it is believed that RCTs are most likely to have internal validity, offering up true conclusions about the effectiveness of the medical or public health intervention for the study population. Simple extrapolation is a proposed solution to the vexing problem of *external* validity or generalizability (how do we go about establishing that the conclusions we drew about the study population are also true of a target population?).

The problem of external validity is viewed as a serious one within medicine, particularly because of the preponderance of unrepresentative RCTs. Peter Rothwell claims that "[l]ack of consideration of external validity is the most frequent criticism by clinicians of RCTs, systematic reviews, and guidelines" (2005, 82). It is also a general problem, a concern in any practical or policy domain – including evidence-based health policy - in which study results are projected to a different context. Michael Woolcock at the World Bank argues that although it has been neglected in evidence-based social development, "the external validity problem is widespread and vastly consequential" (2013, 3).

There are various ways of communicating the effectiveness of a treatment, including qualitative descriptions like 'effective', 'ineffective' or 'harmful'. In chapter 5, we saw that the standard model of therapeutic prediction involves extrapolating or 'generalizing' the effect size of the intervention, as measured in an intervention study. I argued that this

inference scheme requires the assumption that the study and target are sufficiently similar in relevant ways, and that sufficient similarity can be established methodologically or empirically. The sufficient similarity route is also the one considered in recent philosophical treatments of extrapolation in the health sciences (Steel 2008; Broadbent 2013). However, it is not the route traveled in simple extrapolation.

Piet Post and colleagues (2013) did a systematic review of the medical literature to identify approaches to generalizing efficacy results from RCTs. Of the 15 sources that met their inclusion criteria, they categorized 9 of them as recommending an approach that Post and colleagues then go on to defend, the approach that I call simple extrapolation.[97] They summarize the method as follows: "Checking in- and exclusion criteria and evaluating whether there are compelling reasons why the relative effect found in trial results should not be applied to the patient group" (639). Or put differently: "accepting that results of randomized trials apply to wide populations unless there is a compelling reason to believe the results would differ substantially as a function of particular characteristics of those patients" (641-642). Post and colleagues recommend this approach to clinical guideline developers and individual practitioners alike. Their approach favours a presumption of generalizability of the trial results to the target population, which is invalidated only by a strong reason to presume otherwise. A simple method of extrapolation indeed.[98]

Post and colleagues reference several EBM sources that endorse simple extrapolation. For instance, Antonio Dans and colleagues offer the following advice:

> A better approach than rigidly applying the study's inclusion and exclusion criteria is to ask whether there are compelling reasons why the results should not be applied to the patient. A compelling reason usually won't be found, and most often you can generalize the results to your patient with confidence (1998, 545).

This credo is repeated practically verbatim in two clinical epidemiology/EBM textbooks (Weiss 2006; Guyatt et al. 2008), and is recommended in other terms within other EBM

---

[97] Having reviewed these sources, I believe that many of them are miscategorised and support an approach much more akin to the sufficient similarity strategy.

[98] An even less complicated strategy – simplest extrapolation – would let the assumption of generalizability go completely unchecked. Alex Broadbent (2013) uses 'simple extrapolation' to connote this approach, while Nancy Cartwright (2012) uses the term 'simple induction'.

guides (Sheldon et al. 1998; Straus et al. 2011).[99] The advice that one should 'extrapolate the results unless a compelling reason is known' more closely resembles a medical aphorism than a rigorous scientific procedure. Nonetheless, I have found that clinical guidelines often extrapolate from clinical trial evidence to target patients without explicitly and rigorously establishing external validity (Fuller 2013a). This approach is another example of *cookbook medicine*; evidence users are told to the same simple recipe to solve the problem of extrapolation in all cases.

There is no doubt that this advice diverges from the strategy of providing positive justification for the extrapolation by establishing sufficient similarity between the study and the target. In a *Lancet* article, Nancy Cartwright (2011b) argues that extrapolators need such a justification to support the claim that the intervention 'will work for us' in our target context. In response to Cartwright, EBM authorities Mark Petticrew and Iain Chalmers write:

> Cartwright, like many others, conceptualises the challenge as being to demonstrate that the characteristics and circumstances of the research are sufficiently similar to those to which extrapolation is being contemplated. But why should the challenge be conceptualised that way round? Why not instead ask "Are there any good reasons to believe that the research is not relevant to us, that 'it won't work for us'?" If there are not, and considering the undesirable alternative ways of reaching a decision, the default position should be that the result should be regarded as applicable (2011, 1696).

In contrast to so-called 'rigid alternatives' like extrapolating the effect size only if the target population meets the eligibility criteria of the study, or placing the burden of proof on the extrapolator to demonstrate sufficient similarity, simple extrapolation treats generalization as a default. Compelling reasons *not* to extrapolate the effect size can be biologic, social/economic, or epidemiologic; they might concern, for instance, differences in the disease or patient physiology, differences in patient or provider compliance, or differences in comorbid conditions or patient risk of adverse outcomes (Dans et al. 1998; Post et al. 2013).

As my first quote from Post and colleagues (2013) illustrates, simple extrapolation more precisely recommends extrapolating the *relative* effect size, the effect size quantified using a measure such as the relative risk (RR). Recall that the relative risk can be defined as the ratio of the treated outcome rate to the control outcome rate (RR = treated outcome

---

[99] Not all members of the EBM community recommend simple extrapolation; Jeremy Howick and colleagues (2013) argue that several notable examples of failed extrapolations are enough to reject it as a robust strategy.

rate/control outcome rate). Although in principle it could be used to extrapolate from non-randomized group studies, in practice simple extrapolation is most often discussed in the context of extrapolating from RCTs; and though it could be used to extrapolate side effects, it is typically recommended as an approach to extrapolating effectiveness.

What justification is put forth for treating generalizability as a default conclusion in the individual case? Seemingly what does the work is the assumption, articulated by Dans and colleagues in their quote above, that *most often* – in most cases - trial results are generalizable (Dans et al. 1998; Weiss 2006; Guyatt et al. 2008). Along these lines, Post and colleagues claim that "in general, relative effect of treatment on benefit outcomes seldom differs to an important extent across subgroups of patients" (2013, 638). I call this assumption *general generalizability*: in general – in most instances – the relative effect size measured in a trial is generalizable to the target population under consideration. As we will see, a few authors cite empirical studies in support of general generalizability. Other authors seem to take this assumption, as well as the overall approach, to be intuitively reasonable, as if suggested by some special feature of relative effect measures like the relative risk.

In this article, I will argue that simple extrapolation in medicine is *not* reasonable; it is a deeply problematic solution to the problem of extrapolation. As a prelude to what is to come, simple extrapolation implies a particular inference scheme that we can represent informally as follows. 'We don't know of a compelling reason why the relative effect size would be different in this target population; therefore, the relative effect size measured in the trial is generalizable to this target population.' General generalizability ('most often a trial's relative effect size is generalizable') is a crucial (yet often suppressed) assumption of the inference. The thought that there is something unique about the relative risk that makes it transportable to most populations is mythical thinking, the 'myth of the golden risk ratio'. When we strip away that assumption, we expose an argument that commits the fallacy known as the argument from ignorance; absence of evidence (of non-generalizability) is not necessarily evidence of absence (of non-generalizability). Simple extrapolation may lead physicians to extrapolate the effect size more often than they should, which could result in over-treatment. As I suggested in the Introduction, this problem is a special one for patients with *multimorbidity*, who are often treated with multiple interventions after extrapolating from the results of multiple trials.

Daniel Steel (2008) discusses a similar approach to the problem of extrapolation that he calls 'simple induction' and that is used in toxicology to extrapolate the toxic effects of exposures from animal models to humans. Steel suggests that simple induction's chief weakness is that it produces many mistaken extrapolations, which I will argue is also true of simple extrapolation. Jacob Stegenga (2015) also criticizes simple extrapolation as represented by some of the quotes I provided above; he calls it 'simple extrapolation, unless' in recognition of the role of exceptions to the extrapolation rule ('compelling reasons'). Stegenga objects that simple extrapolation ignores three crucial considerations: systematic differences between trial and target patients that may modulate the effect size; the fact that trials suffer from rampant publication bias and provide misleading effect size estimates to extrapolate; and knowledge of disease and treatment mechanisms. Each of these thorny issues will rear its head in what follows, but the structure of my critique will differ from Stegenga's analysis.

In section 7.2, I reject the assumption of general generalizability of the relative effect size, arguing that it is empirically unjustified (7.2.1), mathematically unjustified (7.2.2), and theoretically unjustified (7.2.3), and that it neglects two important reference class problems (7.2.4). In section 7.3, I consider possible defenses of simple extrapolation beyond general generalizability but conclude that these defenses are poor armour: they underappreciate wide gaps in our biomedical knowledge base (7.3.1) as well as the practical consequences of extrapolating (7.3.2). In section 7.4, I conclude that simple extrapolation is a flawed approach, and suggest alternate solutions to the problem of extrapolation that are often less simple but more sound.

7.2 Against the Assumption of General Generalizability

Simple extrapolation leans heavily on the assumption that the relative effect size is usually transportable to the target population. Although the assumption is only implicit in most sources advocating simple extrapolation, as we just saw Post and colleagues (2013) explicitly acknowledge that their approach relies on it. So do Dans and colleagues: "This reasoning, and much of what follows, assumes that relative risk reduction remains constant across

[target patient] subgroups" (1998, 548).[100] Although exponents do not specify how exact this constancy is, it must be exact enough to justify using the trial's relative risk as an approximation for the target population's relative risk in health policy or clinical decision-making.

General generalizability seems to function as a probabilistic warrant for extrapolation. The assumption that most effect sizes are generalizable to most target populations is meant to confer high probability on the conclusion that *this* effect size is generalizable to *this* target population. Asking oneself whether there are any compelling reasons to doubt generalizability for the particular case is a minor check to make sure that this high prior probability is not grossly misguiding us. Presumably, general generalizability holds that effect sizes are transportable so often and thus the prior probability of transportability is so high that unless there are is very strong ('compelling') evidence against transportability we should accept that the effect size is transportable in this case. This thinking might explain why simple extrapolation feels no need to check for further reasons *in favour* of extrapolating the effect size (such as similarity between the study and target) – it feels that we already have reason enough in the form of a strong probabilistic warrant. I could question this line of reasoning from multiple angles. In this section, I will argue that the general generalizability assumption lacks justification, and that simple extrapolation faces two serious reference class problems.

7.2.1 Empirically Unjustified

In math, two numbers (A, B) are 'in the golden ratio' if the ratio of the larger number (A) to the smaller number (B) is equal to the ratio of their sum to the larger number, or $A/B = (A+B)/A = \varphi$. The golden ratio (φ) – variously known as the 'golden section' or 'divine proportion' – is a constant, approximately equal to 1.618. First defined by Euclid, it is an important ratio in mathematics, but also creeps up frequently in art, architecture and nature.

---

[100] In comparison, some clinical researchers assume that quantitative effects typically *differ* across patient subgroups inside or outside of a trial (Yusuf et al. 1984; Lubsen and Tijssen 1989; Bailey 1994). Kent Bailey argues that "if overall treatment effects in terms of relative risk reduction [in a trial] are presented, it should be stressed that such a percent reduction does not necessarily apply to any given individual or to any real population" (1994, 22).

Few were as transfixed with the golden ratio as the Nineteenth Century polymath Adolf Zeising, who described its appearance in botany, zoology, astronomy, chemistry, physics and human anatomy (Padovan 1999). For example, Zeising reported that the ratio of the distance from a person's feet to their navel (A) to the distance from their navel to the top of their head (B) is equal to the golden section, as is the ratio of their total height (A) to their feet-to-navel distance (B).

In medicine, we sometimes proceed as though a treatment's relative risk is similarly golden, that it has some special property by virtue of which it is constant or nearly constant across human populations. It is not intuitively obvious that relative risks should be relatively constant; such an assumption requires justification. One way to justify belief in the general phenomenon of golden risk ratios is through induction from empirical data, by measuring the ratio in samples like Zeising did with φ. Proponents of general generalizability often do appeal to empirical evidence in order to prop up this assumption.

As Post and colleagues (2013) point out, several studies have shown that a particular treatment's relative effect size is similar across two groups that differ systematically with respect to a variable like age (Turnbull et al. 2008a) or sex (Turnbull et al. 2008b). However, there are also many counterexamples in which relative effectiveness has varied according to age (Hlatky 2009; Pignon 2009), region (Fine 1995), disease severity (Rothwell 1995), level of comorbidity (Greenfield et al. 2009), or treatment site (Horwitz et al. 1996). Rothwell (2005) lists many other examples in which the effects of treatment have varied according to patient characteristics or the local context. Claiming particular cases in which the relative effect was similar or dissimilar will not do; we require a systematic review that includes various treatments and various target populations.

Post and colleagues (2013) cite systematic research (Furukawa, 2002; Deeks, 2002) showing that the relative risk of a beneficial outcome is typically similar among RCTs included in the same meta-analysis (a meta-analysis pools the results of several studies). Assuming that their interpretation of the data is correct, it provides an enumerative induction from previous instances in which the RR was similar among (trial) populations to the conclusion that in *most instances* an intervention's RR is generalizable to the (target) population.

Unfortunately, as I've argued RCTs are a poor sample for this induction because of the fact that they are often highly atypical with respect to their patients and care settings (Fuller 2013b). Recall that trials often exclude older patients, patients with multiple diseases, and patients taking concurrent medications (Van Spall 2007). Moreover, patients in trials often receive treatment and care that differ compared to routine practice (Weiss et al. 2008; Dekkers et al. 2010). These peculiarities of trial populations are the reason that many commentators worry about the generalizability of RCT results to typical target populations in the first place. Further, RCTs are usually only included in the same meta-analysis if they are similar in important respects. In fact, a meta-analysis is typically only done if the RCTs exhibit low heterogeneity in terms of the effect size (McAlister 2002)! Finally, several reviews suggest that published trials – oftentimes the only trials that meta-analysts have access to – are more likely to agree on the effectiveness of a drug than all trials, published and unpublished (Song 2010). All of these points are reasons to worry that trials included in a meta-analysis are unusually comparable with respect to the net treatment effect. Additionally, as Stegenga argues (2015), because publication bias systematically overestimates the effectiveness of many treatments it weakens the generalizability of our RCT evidence base. Studies revealing widespread publication bias provide empirical evidence *against* general generalizability.

In order to adequately support as grand an assumption as general generalizability, we would need numerous intervention-specific systematic reviews, collectively constituting a representative sample of interventions. Each review would have to include numerous studies or sub-group analyses, collectively constituting a representative sample of target populations. Unfortunately, comparisons involving various patient subgroups are rarely had, so we would first have to do the primary analyses, and would likely have to run many more studies. Such a massive and unusual research program seems unlikely. As it stands, the empirical evidence put forth in support of general generalizability falls far short.

7.2.2 Mathematically Unjustified

While the constancy of the golden ratio φ within various natural contexts can be shown empirically, its constancy in several mathematical contexts is demonstrated a priori,

mathematically. We can prove that for any rectangle, if its length and width are in the golden ratio then the ratio of its length to its width is equal to 1.618. I will now consider whether we can similarly come up with mathematical justification for the belief in golden risk ratios.

Recall that the relative risk is the ratio of the outcome rates, treated and controlled. We can compare the relative risk to the absolute risk reduction (ARR). Recall that the absolute risk reduction subtracts the treated outcome rate from the control outcome rate (ARR = control outcome rate − treated outcome rate). Meanwhile, the relative risk reduction (RRR) is the ratio of the ARR to the control outcome rate (RRR = ARR/control outcome rate). Proponents of simple extrapolation assume that the RR and RRR ('relative measures') are transportable rather than the ARR (an 'absolute measure').

Paul Glasziou and Les Irwig (1995) suggest one commonly cited reason for making different assumptions about these different measures. Say that the (unbiased) control rate of an undesirable outcome in a trial is 12% and the (unbiased) treated rate is 9% - the treatment prevents the outcome. Then the RR is 0.75 and the ARR is 3%. What if the control rate or 'base rate' of the outcome in our *target* population is 2%? The trial's RR might be generalizable to this target population; if it is, then the treated rate in the target population is 1.5% (treated rate = RR x control rate = 0.75(2%)). However, the trial's ARR cannot be generalizable to the target population because then the treated rate in the target population would be -1% (treated rate = control rate − ARR = 2% - 3%) and negate outcome rates are impossible. Thus, Glasziou and Irwig conclude: "It is usually more reasonable to assume that the reduction in relative risk stays constant" (1356).

From the mathematics, we see that it is impossible that the absolute risk reduction is universally transportable – transportable to all populations - because we would sometimes wind up with absurd negative outcome rates. On the other hand, the formula for RR will never generate negative outcome rates (so long as we input positive numbers). However, it can still calculate rates that are absurd for a different reason. Let's say that the RR in the trial is 5.0 – the treatment produces a beneficial outcome. Let's also say that the control rate of the outcome in the target population is 25%. The trial's RR cannot be generalizable to the target because then the treated rate would be 125% (treated rate = 5.0 x 25%) and outcome

rates greater than 100% are just as absurd as negative outcome rates. Thus, the universal transportability of relative risks greater than 1.0 is impossible.[101]

In their quote, Glasziou and Irwig assume a *reduction* in relative risk, that the effective intervention prevents an undesirable outcome. However, many therapies produce beneficial outcomes like recovery from depression, cancer or even headache. While it is true that when we extrapolate relative risks less than 1.0 we never calculate impossible rates, as Alex Broadbent (2015) argues the mere mathematical possibility that certain relative risks are universally transportable does not imply that they are *in fact* universally transportable.

A positive mathematical or a priori argument for universal transportability would have to show a conceptual dependency between treated rates and control rates strong enough to ensure a universal functional dependency between these rates; in other words, the conceptual link would have to guarantee that the treated rate always varies linearly with the control rate as a matter of necessity. To explore this possibility, I will borrow the classic 'potential outcomes' conceptual framework from Sander Greenland and James Robins (1986).

In quantifying the effect size of a treatment X on a binary outcome O in a target population, the ideal comparison is a counterfactual one: the outcome rate in the entire population if treated versus the outcome rate in the entire population if untreated. Patients differ in terms of their 'potential outcomes' under the two hypothetical treatment conditions. We can sort the target patients into four potential outcome types: individuals who will get the outcome if and only if they are treated (O_X), individuals who will get the outcome if and only if they are not treated ($O_{\neg X}$), individuals who will get the outcome either way ($O_{X/\neg X}$), and individuals who will *not* get the outcome either way ($\neg O$). If we let O_X, $O_{\neg X}$, $O_{X/\neg X}$ and $\neg O$ represent the frequency of individuals of each potential outcome type, we can calculate the relative risk as follows: $RR = (O_X + O_{X/\neg X})/(O_{\neg X} + O_{X/\neg X})$. While the framework assumes determinism, we could adapt this formula to allow for indeterminism. To do so, we could place a coefficient in front of each potential outcome variable representing the probability of the outcome given the treatment condition described in the subscript. The RR would then represent the effect size as a change in probability of the outcome.

---

[101] If we construe relative risks and absolute risk reductions as quantifying a change in *probability* of the outcome (as a ratio or as a difference, respectively), the problem does not go away because negative probabilities and probabilities greater than 100% are incoherent.

When we conceptually decompose the treatment and control rates into potential outcome rates (as we have done), we see that the numerator and denominator do indeed share the variable $O_{X/\neg X}$. However, this does not ensure that if the control rate increases then the treated rate increases proportionately because the treated and control rates each depend on one other variable (the treated rate depends on O_X, while the control rate depends on $O_{\neg X}$). All four variables represent frequencies that can vary such that when we redefine the target population we might capture different relative proportions. Therefore, it is mathematically possible that two target populations will have different relative risks. There is no a priori argument for the universal transportability of the relative risk.

In fact, if any of the target populations contain multiple potential outcome types, we can show that universal transportability of the RR is *im*possible – even for a RR less than 1.0. As an illustration, consider a simple four person target population in which the RR due to intervention is 0.5 (Figure 7-1, target population a). The population is composed of the following potential outcome types: one individual of type $O_{\neg X}$ (25%), one individual of type $O_{X/\neg X}$ (25%), two individuals of type $\neg O$ (50%), and no individuals of type O_X (RR = (0 + 0.25)/(0.25 + 0.25) = 0.5). We could substitute 100 million people for each individual I have described - it makes no difference to the RR or the argument that will follow.
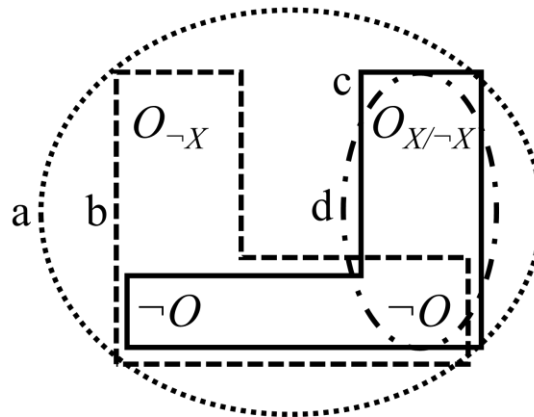
Figure 7-1. Target populations as sets of potential outcomes. a: widest potential target population (RR = 0.5). b: subpopulation/potential target population (RR = 0). c: subpopulation/potential target population (RR = 1.0). d: subpopulation/potential target population (RR = 1.0). $O_{\neg X}$: individual who will get the outcome if and only if they are not treated. $O_{X/\neg X}$: individual who will get the outcome either way. $\neg O$: individuals who will not get the outcome either way.

206

There are fourteen overlapping subpopulations of a four-person target population (excluding an empty subpopulation and a four-person subpopulation), each of which is a potential target population in its own right. Some of these target populations have a RR of 0.5. Yet others have a RR of 0 (e.g. subpopulation b in Figure 7-1); and others have a RR of 1.0 (e.g. subpopulations c and d in Figure 7-1). Thus, no matter what the RR is in the trial, it cannot be universally transportable to all of the potential target populations.

In this example, of the 15 target populations (including the four-person population), the RR is 0.5 in four populations, 0 in four populations, 1.0 in four populations, and indeterminate or 0/0 in the remaining three populations. In the best case scenario in which the RR is 0.5 or 0 in the trial (the preventive intervention showed efficacy), it is still generalizable to less than half of the target populations in this setting. Even more worryingly, the effectiveness of the intervention is qualitatively different in almost half of the target populations (it is totally 'ineffective' in the populations with either an indeterminate RR or a RR of 1.0).

A defender of general generalizability might argue that we run into trouble only when we consider small subpopulations to be legitimate target populations because small populations are more likely to contain skewed proportions of potential outcomes. We could instead choose as *the* target population the widest population consisting of all individuals for whom the intervention is indicated. This defender might even argue that the trial population is sampled from this widest population, and thus the trial's RR is unlikely to differ dramatically compared to the widest population's RR. However, the trial population is not usually a *random* sample from the widest population, it is a highly filtered convenience sample. Thus, the trial's RR might not be generalizable to the widest population. To further complicate matters, as we have just seen the trial's RR is generalizable to some but not all of the subpopulations of this widest population (other potential target populations) because the RR presumably depends on which kinds of individuals a subpopulation includes.

Whenever there is a low frequency of treatment responders in the widest potential target population, the majority of subpopulations will not contain any treatment responders. Thus, a preventive intervention that has low or modest efficacy in a trial – a treatment that is effective for a minority of trial participants – might not be effective *at all* in the majority of target populations. Unfortunately, many modern preventive medications are associated with a

low or modest ARR in clinical trials (Moynihan and Cassels 2005), which may suggest a low frequency of treatment responders (type $O_{\neg X}$ individuals) in the trial and potentially in the general population.[102] In that case, even if these effect sizes are generalizable to some target populations, they are not generalizable to *most* potential target populations. Not only does general generalizability lack mathematical justification, but if we understand it as an assumption about all of the populations of which we could conceive, then basic accounting suggests that it may be false.

7.2.3 Theoretically Unjustified

Returning to our running analogy, Zeising saw the golden ratio φ as an ideal that permeated all the forms of nature he investigated. According to Richard Padovan (1999), in the constancy of the golden ratio within the physical world, Zeising saw the working of a universal law.

      Broadbent (2013, 2015) discusses the contemporary tendency of epidemiologists to privilege the relative risk over other measures of association such as the absolute risk reduction, a tendency that has been dubbed 'risk relativism'. He speculates that part of the appeal of risk relativism is the assumption that relative risks are more often transportable than other measures, and asks whether they might figure as constants in epidemiological laws. For example (2015), the law relating lifetime rate of lung cancer among male smokers (Rs) to the lifetime rate among male non-smokers (Rns), would be: Rs = 20 x Rns; the constant in this law, the number 20, is the relative risk of lung cancer in males due to smoking across male populations. To come up with the epidemiological laws that Broadbent

---

[102]   ARR = outcome rate if untreated – outcome rate if treated
           ARR = $(O_{\neg X} + O_{X/\neg X}) - (O_X + O_{X/\neg X})$
           ARR = $O_{\neg X} - O_X$
A low ARR can be explained by: a low frequency of $O_{\neg X}$ individuals and very low frequency of O_X individuals; or a moderate-high frequency of O_X individuals and a slightly higher frequency of $O_{\neg X}$ individuals. The first explanation is more plausible than the second explanation in most cases because the second explanation requires that the intervention has a surprising opposite effect in a large proportion of people. A third explanation – more difficult to rule out - is that the intervention acts indeterministically and achieves a small lowering of the chance of the outcome in a large number of treatment responders.

describes, we simply rearrange the formula for RR and replace the RR variable with the value we measured in the study, turning the relative risk into a constant specific to that exposure.

After extrapolating the relative effect size from a trial, the EBM approach to individualizing treatment involves calculating the absolute effect size in the target population from the relative effect size and the untreated outcome rate in the target population (Glasziou and Irwig 1995; Straus et al. 2011; Scott and Guyatt 2011; Post et al. 2013). For instance, we calculate the target's absolute risk reduction by multiplying the trial's relative risk reduction by the target's untreated outcome rate (Scott and Guyatt 2011): ARR = RRR x untreated rate. Once again, we simply rearrange the formula for RRR and assume the trial's RRR to be a constant for that intervention. In epidemiology and in EBM, we often treat RRs and RRRs as if they are constants in universal laws. Perhaps we can locate theoretical (versus a priori) justification for golden risk ratios in this peculiar way of thinking about population health.

It is first instructive to recall that the quantities related by the epidemiological laws above are outcome rates (the ARR can be rewritten as a difference in outcome rates). Outcome rates are simply population aggregates. If target populations consist of individuals of different potential outcome types, it is strange – even incoherent, given the mathematical considerations presented in the previous section - to think that for each target population we define we capture just the right types of individuals in just the right proportions so that the relative treatment effect stays constant. A more coherent explanation for why the RR is a constant could posit that target populations *do not* consist of individuals of different potential outcome types; in other words, the treatment effect is constant across individuals. The aggregate treatment effect would then always equal the individual treatment effect.

The assumption that the effect of the treatment is constant across individuals is the strongest imaginable statement of universal transportability. It suggests that the effect size is transportable to each and every individual, and thus to all populations we could form from those individuals.[103] It implies that conceptual frameworks like potential outcomes contain

---

[103] Steel (2008) suggests that the problem of extrapolation only faces extrapolators dealing with populations that are heterogeneous in terms of characteristics that affect the causal relationship being studied; for him, the problem of extrapolation is short for "*the problem of extrapolation in heterogeneous populations*" (3). Steel argues that homogeneity of treatment effects is an unreasonable assumption in common contexts.

metaphysical baggage. These frameworks may be conceptually thorough, accounting for all types of individuals of which we could conceive, but for effective treatments we need only one potential outcome type: individuals for whom the exposure makes a difference. The fact that most relative risks for effective preventive interventions are between 0 and 1.0 must then mean that those treatments act indeterministically, lowering the probability of the outcome by the same amount in each individual.[104]

Yet sometimes we do find different effect sizes in different subgroups, as illustrated by the examples in section 7.2.1. When we fail to look for subgroup differences in a trial, it is often not because we believe that true subgroup differences are nonexistent but because we lack statistical power to estimate them with reasonable precision. *Treatment effect heterogeneity* is a problem for the extrapolator because it implies that the effects measured in trials may fail to generalize to the target populations in which they are interested.

Our most general biomedical theory should lead us to expect differences in responsiveness to intervention among individuals. Biomedical theory may not be powerful enough to predict exactly which individuals will respond to an intervention and with what probability (at least at present), but it does predict biological heterogeneity within populations, especially for the kinds of diseases we are trying to prevent (such as cardiovascular disease). Recall that these diseases are often described as multifactorial, caused by the interaction of genetic and environmental factors.[105] Genetics and environmental exposures vary among individuals, producing phenotypic outcome variation. Since an intervention is simply one of these exposures, we should expect that the effects produced by its interaction with these inconstant causes will also be inconstant.[106] Thus, the universal transportability assumption flies in the face of established theory.

---

[104] This statement is stronger than what is called 'contextual unanimity' in the probabilistic causality literature. Contextual unanimity holds that a cause raises (lowers) the probability of its effect in every context (Dupre 1984; Cartwright 2010); it does not require that the cause raise (lower) the probability by the same fixed amount.

[105] Rothman and Greenland (2005) convincingly argue that *all* human outcomes are attributable to the interaction of genetic and environmental factors.
[106] Kravitz et al. (2004) provide several examples of treatment effect heterogeneity produced by particular genetic, behavioural and environmental variables.

Simple extrapolation is not out of ammunition just yet. Recall that simple extrapolation need not rely on universal transportability but on something weaker: general generalizability. Even though universal transportability is not supported by theory, perhaps general generalizability is. The following argument is best couched in Bayesian terms. In section 7.2.2, I argued that whenever there is variation in responsiveness to the intervention and a low frequency of responders, we should worry that the effect size is not generalizable to most *potential* target populations, most of the populations of which we could conceive. This suggests that the *prior probability* of generalizability to actual target populations – those that we choose in practice - might be low. The prior probability in this case reflects our initial ignorance as to the RR in each conceivable target population (thus, the prior probability of generalizability to the chosen target population is equal to the proportion of potential populations in which the effect size is generalizable – in this case a low proportion).[107]

Guideline developers, health policy makers and physicians can partly overcome their initial ignorance by comparing the study population/context to the chosen target population/context, thus forming a *posterior probability* of generalizability that is based on evidence of comparability and non-comparability. In this example, for the posterior probability of generalizability to be high, the extrapolator would have had to have chosen a target population from among the set of potential populations for which the evidence favouring generalizability is strong enough to overcome the low prior probability.

Yet given the changing demographics and characteristics of patients in community and hospital practice (Goulding 2003, Barnett 2012, Qato 2008), target populations increasingly contain many older patients, patients with multiple diseases, and patients taking multiple medications. As mentioned earlier, RCTs often exclude these very patients. Age-

---

[107] One might argue that in practice some conceivable populations should not be considered potential target populations, and that ruling out some of the conceivable populations might alter the prior probability so that it is no longer low. In order for our prior probability to be high, we would have to rule out as potential target populations some 'wrong' conceivable target populations to which the RR is not generalizable; but by assumption we are initially ignorant as to the RR in each conceivable target population, so have no basis for deciding which conceivable populations are the 'wrong' ones. If we do rule out a conceivable target population (e.g. on the basis that it is arbitrary or awkward to define), we do not know whether it is a wrong population or not, and thus the prior probability should remain unchanged.

related physiological changes, pathophysiological changes, and medications can all influence the way that patients metabolize and respond to treatment, potentially modifying the effect size. Furthermore, patient care often differs systematically in trials compared to routine practice, sometimes in ways that influence treatment effectiveness. Therefore, the posterior probability of generalizability to the chosen target population might actually be lower than the prior probability – not higher.

In summary, we have no reason to believe that the RR figures in universal epidemiological laws as a constant. In fact, we have found no good reason to suppose that there is *anything* special about relative risks that should give us faith that they are universally transportable or even generally generalizable. Golden risk ratios are mythical; they are not a golden ticket for extrapolation. Rather, believing that risk ratios exhibit relative constancy across populations might bias us towards extrapolating them more often than we should, which could lead to many unwarranted or mistaken extrapolations.

7.2.4 Reference Class Problems

So far, I have shown that the general generalizability assumption is empirically, mathematically, and theoretically unjustified; it is a shaky support for simple extrapolation to lean on. I will now argue that even if the assumption were justified, simple extrapolation ignores two weighty and distinct manifestations of the reference class problem. One problem arises when we try to locate a patient in a target population and find that we have several target populations to choose from; it is a problem with the approach of simple extrapolation more generally. A distinct problem arises if we try to locate a target population in a class of target populations and discover that there are several possible classes; it is a problem specifically with the probabilistic warrant that general generalizability is supposed to provide. I will explain the latter problem first.

General generalizability claims that the relative effect size is generalizable in a (great) majority of instances of a particular target population and a particular intervention. If true, it provides probabilistic rationale for concluding that the effect size is generalizable in *this* instance: the probability of generalizability is high. The reference class for this probability is suppressed; the probability of generalizability is high within the most *general*

class of all target populations and interventions. Even if the probability of generalizability is high given this broadest reference class, there are many relevant subclasses. The intervention in question might be a surgical intervention; and the probability of generalizability might be low for surgical interventions because the effectiveness of surgery depends greatly on characteristics of the surgical site. Or the target population in question might have concurrent diseases; and the probability of generalizability might be low for target populations with concurrent diseases because diseases and the medications we prescribe for them interact with the intervention. It is not always rational to choose the widest reference class; in fact, as I suggested in Chapter 5 there are strong arguments for choosing a narrower reference class whenever the probability in the narrower reference population is known. Thus, even if general generalizability is true it would sometimes be unreasonable to rely on its probabilistic warrant for extrapolation.

The other reference class problem I mentioned is not specifically a problem with general generalizability, but this point in the discussion is a natural place to raise it. It is a problem that confronts any approach to extrapolation in a discipline like clinical medicine that deals with individuals. We encountered this reference class problem in Chapter 5: a patient belongs to many target populations in which the effect size or change in probability of the outcome due to intervention might differ. The arguments of the preceding sections established that a trial's effect size will likely be generalizable to some target populations but not to others. Simple extrapolation might conclude (perhaps correctly) that the effect size is generalizable to one target population that includes the patient, and also conclude - again, correctly - that the effect size is *not* generalizable to another population that includes the same patient. Simple extrapolation gives no guidance on what should be done in this highly plausible and paradoxical scenario. Again, this problem arises for any approach to extrapolating the effect size in clinical medicine, but simple extrapolation ignores the problem by providing no guidance on the appropriate choice of target population. When the target population is not specified by the approach, there is no metaphysically privileged population that we can assume.

General generalizability is a problematic assumption. It lacks justification, and even if it were defensible it would often be unreasonable to rely on the probabilistic warrant it provides. Further, as the second reference class problem above shows, the trouble with

simple extrapolation goes beyond general generalizability. In section 7.3, I will argue that –
once stripped of general generalizability - simple extrapolation rests on a fallacy from which
it has no respite.

7.3 Argumentum ad Ignorantium

Could simple extrapolation survive without its highly suspect general generalizability
assumption? Without the assumption, the argument in favour of extrapolating the effect size
in simple extrapolation would look like this: 'We don't know of a compelling reason why the
relative effect size would be different in this target population; therefore, the relative effect
size measured in the trial is generalizable to this target population.' Readers familiar with
informal logic might recognize this inference as an argument from ignorance. We don't
know that the effect size is non-generalizable; therefore, the effect size is *not* non-
generalizable (or the effect size *is* generalizable). As 'absence of evidence is not necessarily
evidence of absence', we conventionally charge this inference with committing an informal
fallacy.

However, as Douglas Walton shows (2006), the argument from ignorance is not
always fallacious. Epistemic considerations might rescue the inference if we usually have
evidence of non-generalizability whenever an effect size is not generalizable. Alternatively,
practical considerations could salvage the argument if there is little to be lost by
extrapolating but much to be gained. Unfortunately, we will see that knowledge
considerations instead dictate that we should balance reasons against generalizability with
reasons in favour of it; and pragmatic considerations dictate that reasons need not always be
'compelling' like simple extrapolation demands. Thus, I cannot acquit simple extrapolation
of the charge of fallacy.

7.3.1 Knowledge Considerations

Although absence of evidence is not *necessarily* evidence of absence, sometimes absence of
evidence does provide evidence of absence. Walton (2006) argues that when a knowledge

base K is complete or complete enough, if a proposition *P* is absent from K we can reasonably take *P* to be false.

Here is the thinking. If K is complete, then it is true that: 'if *P*, then we know that *P* (K(*P*))'; and, by implication: 'if not K(*P*), then not *P*'. We can let *P* stand for the proposition 'the effect size is non-generalizable', and the second conditional then reads: 'if we do not know that the effect size is non-generalizable, then the effect size is not non-generalizable.' Simple extrapolation's method of asking whether there is a compelling reason to believe that the effect size would be different in the target population can be seen as questioning whether we know that *P*.

When our knowledge base is incomplete, these strict conditionals are strictly false because *P* might be one of the true propositions that is missing from our knowledge base. However, if our knowledge base is *reasonably* complete then the inference from 'not K(*P*)' to 'not *P*' might still be reasonable. As an illustration, I might argue that if Bigfoot (as a species) exists, we would probably know it (after all, how could a race of giant bipedal hominoids escape our notice?). Based on the evidence, we don't know that Bigfoot exists (not K(*P*)); therefore, Bigfoot does not exist (not *P*). Whether or not you are persuaded by this argument, it would be wrong to label it fallacious. The argument is at least admissible because if our knowledge of large apelike mammals in the North American wilderness is reasonably complete, then lack of knowledge of Bigfoot's existence can reasonably be taken to support the claim that Bigfoot does not exist.[108]

The problem with using this argument scheme to extrapolate the effect size in medicine is that in many cases the relevant knowledge base is neither complete nor reasonably complete but *radically incomplete*. Compelling reasons for declaring that the effect size is not generalizable to the target population include physiological or pathophysiological differences between the trial and target population (Dans et al. 1998; Post et al. 2013). Physiological and pathophysiological differences describe differences in the mechanisms of drug metabolism and drug action, as well as in the disease mechanisms targeted by an intervention. For example, different ethnic groups sometimes differ systematically with respect to drug-metabolizing liver enzymes, and also in terms of the

---

[108] Elliott Sober (2009) provides a formal probabilistic justification for these kinds of 'absence of evidence' arguments.

underlying pathogenesis of common diseases (Dans et al. 1998). Unfortunately, as Jeremy Howick and colleagues (2013) argue, our knowledge of therapeutic mechanisms is often very incomplete. To further complicate matters, authors publishing the results of clinical trials often include insufficient detail about the recruited patients or the recruitment process (Gross et al. 2002).

Crucially, our knowledge of all of the relevant causal differences among patients is limited. I mentioned earlier that many modern therapeutics are associated with a low absolute effect size in clinical trials, and thus ostensibly a low frequency of treatment responders. This observation provides indirect evidence that our knowledge of relevant causal differences is often limited. If researchers knew before the trial who the non-responders would be, they would likely have excluded them from the study rather than dilute the effect size and give these patients an intervention that will not help them but might harm them; thus, the researchers likely had limited knowledge of relevant differences before the trial. Further, trial results usually reveal the overall effect size and not which participant characteristics explain heterogeneity in responsiveness; thus, the researchers could have no less limited knowledge of the relevant differences after the trial. If researchers have limited knowledge of relevant differences within populations, they also have limited knowledge of relevant differences between populations.

Our incomplete knowledge concerning relevant causal differences is a central reason why randomized trials are often preferred over non-randomized studies in evaluating the efficacy of an intervention. Causal variables are called 'confounders' in a study when they are imbalanced among study groups and when their imbalance biases the causal interpretation of the results. We could employ a principle analogous to simple extrapolation with respect to confounders: presume that the study groups are comparable with respect to potential confounders unless we know of a compelling reason to believe they are not comparable (i.e. we find an imbalance in a potential confounder). However, as with simple extrapolation, a search for compelling reasons will only catch differences that are known to be relevant. Those advocating for randomized trials are often worried about the realistic possibility of *unknown* confounders, which randomization is thought to neutralize (Byar et al. 1976; Hill & Hill 1991; Schmoor et al. 1996). Advocates of randomization (which include advocates of simple extrapolation (Guyatt et al. 2008; Straus et al. 2011; Post et al. 2013))

216

should be *even more* worried about differences between study populations and target populations of unknown relevance because studies and targets usually differ in terms of factors like comorbidities and care monitoring that are in many cases relevant.[109]

We cannot rely on a reasonably complete knowledge base to validate simple extrapolation's argument from ignorance. If we are to use knowledge of underlying mechanisms to extrapolate the effect size, we must instead search for evidence of similarity in addition to evidence of dissimilarity between trial and target. Rather than a default inference that presumes the effect size to be the same in both contexts (barring evidence to the contrary), this approach would involve the procedure – more traditional in scientific reasoning – in which a conclusion is reached on the balance of evidence for and against it.

7.3.2 Pragmatic Considerations

In the previous section, I questioned whether there was something special about our scientific knowledge base that would allow us to arrive at knowledge of generalizability via an argument from ignorance. But maybe *knowledge* of generalizability is the wrong idea. Maybe knowledge is too strict of an epistemic standard to enforce on the conclusion. In science, a hypothesis is only declared knowledge once it is highly confirmed. Rather than a method for deciding when generalizability of the effect size is known, simple extrapolation is plausibly a method for deciding when we should *accept* that the effect size is generalizable for the practical purpose of acting rationally.

Evidence for this interpretation of simple extrapolation can be found in the very wording chosen by its advocates. Post and colleagues recommend "*accepting* that results of randomized trials apply to wide populations unless there is a compelling reason" (2013, 641-642; my emphasis). Simple extrapolation helps clinicians judge whether we can accept that the trial's effect size is generalizable to the target population so that we can decide whether or not to intervene. The question of generalizability is only relevant to this decision if the decision hinges on it. Physicians and policy-makers should only bother using simple extrapolation if upon accepting that the intervention's effect size is generalizable they would

---

[109] I thank Mathew Mercuri for suggesting the analogy between unknown differences in extrapolation and unknown confounders in causal inference.

recommend/implement - or at least seriously consider recommending/implementing – the intervention. Thus, perhaps simple extrapolation's default of generalizability is justified in case there is little to be lost by extrapolating and intervening but much to be gained.

Yet clearly the decision of whether or not to use the intervention should depend on more than the probability that the relative risk of the beneficial outcome is generalizable; it should depend also on the benefits and costs we should expect if we are right and if we are wrong. The decision to extrapolate must involve considerations beyond reasons suggesting that the effect size will differ; it must account for the consequences of our actions.

On this interpretation, simple extrapolation constitutes what Heather Douglas (2000), following Carl Hempel (1965), calls an 'acceptance rule' in science. An acceptance rule is a method for deciding when to accept a hypothesis. As Douglas argues, when accepting the hypothesis has some practical implications (e.g. the use or regulation of an exposure), and these practical implications will lead to outcomes on which we attach (non-epistemic) values, the acceptance rule must be sensitive to these outcomes and their value. Thus the threshold for accepting the hypothesis – what probability of error or 'inductive risk' we are willing to tolerate – must track the consequences.[110] Whatever your opinion about whether acceptance rules should consider practical consequences in science, surely acceptance rules should be sensitive to *practical* consequences in clinical *practice*.

Yet simple extrapolation is *insensitive* to the consequences because the evidential standard for reasons that would forestall generalizing is invariant: it is set to 'compelling' in all cases. This suggests that we always require reasons so strong that they would raise the probability of non-generalizability substantially. This use of a rigid standard might be due to general generalizability and the assumption that the prior probability of non-generalizability is very low. If we set aside general generalizability, then the 'compelling' standard, as a universal bar, is already in trouble. However, regardless of general generalizability, the 'maximum tolerable probability of non-generalizability', a threshold above which we should not extrapolate, must depend on the consequences. In particular, it must depend on the

---

[110] In this spirit, Ross Upshur (2001) argues that what is at stake in accepting or rejecting a hypothesis in clinical research should influence the significance level, the statistical threshold.

important outcomes we can expect if we are right about generalizability and if we are wrong, the effect size for each outcome, and the desirability of each outcome.[111]

To illustrate, if the side effects of a drug are common or serious, and the potential benefits are uncommon or minor, then even *weak* evidence that the benefits will not obtain in the target population might justify not extrapolating the beneficial effect size. After all, if we extrapolate, treat with the drug and are correct about generalizability, then the target population may only enjoy marginally greater benefit than harm. But if we extrapolate, treat with the drug and are wrong about generalizability, then the target population may suffer significant net harm. Sometimes, less-than-compelling reasons to believe that the effect size will differ are enough to accept that the effect size is not generalizable. Insisting on compelling evidence in every case could lead us to extrapolate and treat more often than we should.

---

[111] We can model this decision using decision theory by comparing the expected utility of extrapolating and intervening (EU_I) with the expected utility of not extrapolating and not intervening ($EU_{\neg I}$). If we intervene, there are two possible scenarios: the intervention benefit is generalizable (G), or the intervention benefit is not generalizable ($\neg$G). There is a partial expected utility associated with G (EU_G), as well as a partial expected utility associated with $\neg$G ($EU_{\neg G}$). The EU_G depends on the benefits and harms of intervention, as well as their probabilities. We might measure the probability of a benefit or a harm for the patient using the effect size (but recall from chapter 5 that this requires a tricky particularization inference). If we assume one major benefit B (the primary outcome targeted by the intervention) and one major harm H (a major side effect or adverse event), then $EU_G = p(B)u(B) + p(H)u(H)$, where $p(B)$ and $p(H)$ are the probabilities and $u(B)$ and $u(H)$ are the utilities ($u(B)$ is positive and $u(H)$ is negative). We can further assume that $EU_{\neg G} < EU_G$; otherwise, there is no point in even considering intervention. Meanwhile, the total expected utility of intervening is $EU_I = p(G)EU_G + p(\neg G)EU_{\neg G}$. According to decision theory, we should implement the intervention if $EU_I > EU_{\neg I}$. Simple extrapolation implies that this will usually be the case *unless* compelling evidence substantially raises the $p(\neg G)$ (and thus substantially lowers EU_I). However, EU_I also depends on EU_G and $EU_{\neg G}$, and thus on the benefits and harms of intervention and their effect sizes. Thus, even if $p(\neg G)$ is low to begin with (a risky assumption), EU_I might only be slightly greater than $EU_{\neg I}$ to begin with, and raising $p(\neg G)$ by just a little might tip the balance in favour of $EU_{\neg I}$. John Worrall (2010) and Stegenga (2015) also model treatment decisions using decision theory.

7.4 Less Simple, More Sound

Bradford Hill famously wrote: "At its best…a trial shows what can be accomplished with a medicine under careful observation and certain restricted conditions. The same results will not invariably or necessarily be observed when the medicine passes into general use" (quoted in Horton 2000, 3152). Advocates of simple extrapolation often less cautiously assume that relative treatment effect sizes are generally generalizable to everyday target contexts, as if relative risks are golden like φ. This assumption is empirically, mathematically and theoretically unjustified; we should not subscribe to the myth of the golden risk ratio. Yet without general generalizability, simple extrapolation is in equal trouble. From the absence of compelling reasons to think that the effect size will differ, simple extrapolation concludes that the effect size will not differ. Neither epistemic nor practical considerations can rescue this fallacious inference. Simple extrapolation is just too simple, and as a result of its default of generalizability, extrapolators might extrapolate the effect size from a trial more often than they should. For patients with multimorbidity, this can easily result in over-treatment

The irony is that simple extrapolation's chief proponent, EBM, demands utmost rigour in evaluating comparability among study groups – the guarantor of internal validity. EBM has worked to advance the clinical science of internal validity, especially through hierarchies of evidence for treatment (which, as Adam La Caze argues (2009), are essentially hierarchies of internal validity). The RCT or systematic review of RCTs sits atop the hierarchy because it is believed to be the most rigorous test of an intervention's efficacy. Why should EBM relax the well-motivated concern for rigour when it comes to evaluating comparability between study populations/settings and target populations/settings – the chief determinant of *external* validity? In lieu of a simple aphorism, a cookbook approach, EBM and the health sciences need rigorous approaches to assessing external validity that are informed by a science of extrapolation.

We can still practice evidence-based decision-making without subscribing to such a simplistic solution to the problem of extrapolation. Despite simple extrapolation's shortcomings, a restricted or modified version of the approach might sometimes be warranted. We might assume that in a restricted domain such as treatment with a particular therapeutic, the trial's effect size is generalizable to most target populations. Such an

assumption requires justification; we might have empirical evidence comparing the effect size in a wide range of typical target populations, or we might have theoretical rationale for believing that responsiveness varies little among individuals (low treatment effect heterogeneity). Alternatively, in a restricted domain our biomedical knowledge base might be reasonably complete, giving us confidence that the effect size is generalizable whenever we lack knowledge to the contrary; or the potential benefits of intervening might be so great that we should accept that the effect size is generalizable unless the evidence suggesting otherwise is compelling.

For extrapolating the effect size, the main alternative to simple extrapolation is to establish sufficient similarity between the study and target. One approach to securing similarity is methodological; we can match the study to the target, or vice versa. The methodological strategy is exercised by pragmatic RCTs, in which the trial population and trial setting are deliberately chosen so as to be representative of some target population. We can also design alternative trials or carefully controlled cohort studies that are more generalizable to everyday clinical contexts. The reverse strategy - dismissed by simple extrapolation proponents (Post et al. 2013; Dans et al. 1998) - is to match the target to the study by choosing a target population that meets the eligibility criteria of the study and/or mirrors the baseline characteristics of the study population.

In contrast to the methodological strategy, a theoretical approach to establishing adequate comparability uses causal-empirical evidence: knowledge of the relevant causal variables, and empirical evidence of their distribution in the target and the study. This strategy is evident in mechanisms approaches to extrapolation (Steel 2008; Clarke et al. 2013), in which the theoretical mechanism describing the intervention's working provides knowledge of the relevant causes.

Each of these approaches faces its own difficulties. Though they are on the rise, pragmatic trials are still uncommon (Vallvé 2003). On the other hand, extrapolating only to populations that meet the trial's eligibility criteria limits the predictions we can make about the effectiveness of our interventions. Finally, as I mentioned earlier our knowledge of mechanisms needed for a theoretical approach to extrapolation is often radically incomplete. While simple extrapolation might lead to a high rate of false positive or unwarranted extrapolations, theoretical approaches to demonstrating sufficient similarity might lead to a

high rate of false negative extrapolations (those in which we do not extrapolate when we should), particularly if the standard of evidence for sufficient similarity is demanding.

Another way to deal with the problem of extrapolating the effect size is to side-step it; for instance, by extrapolating something weaker from the study, such as the qualitative treatment effect. We can give the problem of extrapolation even wider berth by predicting without extrapolating from a study; for example, through mechanistic prediction, or by predicting from clinical experience (which, as we have seen, have problems of their own). In healthcare, predicting the intervention's effect size is desirable; it allows us to estimate how much of a difference the intervention makes. Side-stepping the difficulty of extrapolating the effect size may thus seem like a poor solution. But some of the arguments of section 7.2 should lead us to worry that an effect size is not generalizable much or even most of the time – general non-generalizability! An approach that makes no promises about extrapolating the effect size is superior to one that makes promises and then routinely breaks them, leading us astray.

No solution to the problem of extrapolation is unproblematic, and thus it may be necessary to have multiple strategies in our repertoire, to be brought out and applied as the circumstances dictate. Whatever strategy we execute on a given occasion, it seems clear to me that our approach must satisfy at least two desiderata. First, it must be consistent with some more general theory of the conditions under which the effect size travels between two populations. What I am looking for in a general theory of extrapolation is an analogue to general theories of causal inference in epidemiology like those that utilize structural models or the potential outcomes framework I introduced in section 7.2.[112] Second, an extrapolation approach must include an acceptance rule – a guide advising us when to extrapolate – that is sensitive to the practical consequences of extrapolating, or the benefits and costs we can expect if we are right and if we are wrong. These two desiderata will often conflict with a third desideratum that might partly motivate EBM's embracement of simple extrapolation, namely simplicity. In tackling the problem of extrapolation in medicine, simplicity is certainly to be valued - but not at the expense of soundness.

---

[112] Daniel Steel (2008) as well as Elias Bareinboim & Judea Pearl (2013) offer theories of extrapolation using structural models.

Chapter 8 - Discussion

8.1 Main Conclusions

The Central Aim of this dissertation was a philosophical analysis of chronic disease and of evidence-based medicine, especially concerning treatment and prevention. In the previous six chapters, we explored the nature of chronic disease, models of disease classification, mechanistic prediction, epidemiological prediction, comparative group study causal inference, and the problem of extrapolation. In my analysis I focussed on the following six research questions. What are chronic diseases? What is our current model of disease classification, and what classificatory ideal should we seek in order to promote effective disease prevention? Why do mechanistic predictions miss in medicine, especially when predicting the results of intervention? What is the standard model of prediction in medicine? What roles do confounders and causes serve in comparative group study causal inference? And why is simple extrapolation a poor solution to the problem of extrapolation? I will conclude by looking back at where we come from, and looking around to see where we have arrived.

I began by introducing the story of the new medical model. The old medical model, the biomedical model described by Engel, aims to cure acute diseases using knowledge from the basic biomedical sciences. The second half of the Twentieth Century saw the ascent of chronic disease and evidence-based medicine, which transformed the traditional medical model into what I call 'the new medical model'. The new model aims to prevent and manage chronic diseases using the principles of EBM. Chronic diseases and EBM pose many daunting challenges for modern medicine, many of which surfaced and resurfaced throughout our ongoing discussion.

We can now continue the narrative of the new medical model, which runs an (admittedly tortuous) course through the preceding discussion. To summarize, chronic diseases constitute the model's ontology (Chapter 2). To prevent a chronic disease, physicians target its multifactorial risk factors (Chapter 3). In the medical treatment and prevention of a chronic disease, they use interventions developed from an understanding of

the underlying mechanisms (Chapter 4) and tested in comparative group studies. They predict an intervention's effect size for their patients based on the results of these studies (Chapter 5) by first assessing a study's internal validity (Chapter 6), and then using the principle of simple extrapolation (Chapter 7). They then repeat this procedure for any additional chronic diseases.

In Part One, I argued for a particular ontology: chronic diseases as bodily properties. The pathogenesis of a disease is the process or mechanism that produces the disease, and the manifestations of a disease are effects of the disease. I also argued that all types of disease – whether acute or chronic - are defined constitutively, or in terms of what constitutes the disease. What distinguishes one type of disease from another type is the kind of property that is picked out. Because many chronic diseases are defined dispositionally, different types of chronic disease are often distinguished according to the characteristic manifestation towards which they are disposed. To illustrate, diabetes mellitus is defined as a state disposed towards hyperglycemia; it is distinguished from chronic obstructive pulmonary disease (COPD), which is defined as a state disposed towards airflow limitation. Each diagnosis refers to a different type of bodily property.

One might think that acute infectious diseases are easier to prevent than chronic and noncommunicable diseases because while infectious diseases have a single etiologic agent, chronic and noncommunicable diseases are multifactorial in their etiology. Chronic diseases are caused by variable risk factors like smoking and sedentary lifestyle. However, acute infectious diseases are also multifactorial in their etiology, they are caused not only through exposure to a particular microbe but also through environmental and host factors. With respect to prevention, a core difference between acute infectious diseases and chronic diseases is that while most acute infectious diseases achieve the monomechanism ideal, most chronic diseases do not. In the monomechanism ideal, each type of disease has one etiologic mechanism represented by one mechanistic model. In order to achieve the ideal for a chronic condition, we may need to define new chronic diseases according to their etiologic mechanisms.

If we achieve the monomechanism ideal for a particular disease, we might then be able to rationally design an intervention that would disrupt the etiological mechanism for any case of the disease. Unfortunately our mechanistic models often fail to predict the results of

medical interventions. There are several general explanations for mechanistic prediction failure. Models can be incorrect, which might lead us to predict that the intervention is effective when in fact it is ineffective. Models can also be incomplete, which might lead us to predict vaguely, or to make no prediction. Lastly, models can be overly abstract, which might lead us to predict that the intervention works in general when it works only in certain circumstances. Interventions discovered using mechanistic models are typically tested in clinical trials, and those that clear trials are marketed for clinical use.

Part Two was largely concerned with epidemiological inferences. In order to infer that an intervention would be effective and safe in clinical practice, physicians could make mechanistic predictions with the same models used in rational drug discovery. However, EBM cautions that mechanistic predictions are generally unreliable, and advocates instead for predictions based on the results of population studies. EBM's preferred intervention study is the RCT because EBM proponents believe it has high internal validity. I examined one justification for this belief, the idea that randomization balances all confounding causes, and rejected the belief as neither true nor directly relevant for causal inference. I proposed a new theory of causal inference in which it is the balance of a condition C, representing all of the complex causes, that determines sound causal inference. My account explains how randomization prevents selection bias (by preventing systematic imbalances in correlates of C), but it suggests that we can still draw causal conclusions in non-randomized studies so long as the study groups are sufficiently comparable with respect to C.

Once they have appraised the internal validity of an intervention study, physicians can then use the standard model of prediction, which I call the Risk Generalization-Particularization Model. In generalization, we extrapolate the effect size from the study to the target population. In particularization, we convert the effect size into a change in probability of the outcome for a patient in that target population. The most obvious way to warrant a generalization is to show that the study and target are sufficiently similar. However, in EBM's approach to generalization, simple extrapolation, we instead look for compelling *dissimilarities* between the two populations/settings. Finding none, we extrapolate the relative effect size. Yet there is nothing about relative effect sizes that should lead us to believe that they are generally generalizable. Unfortunately, without assuming

general generalizability, simple extrapolation rests on a fallacious argument for ignorance. In short, simple extrapolation is a deeply problematic solution to the problem of extrapolation.

After extrapolating the effect size in the Risk GP Model, we particularize, or transform the effect size into a change in probability for a patient. This inference assumes that the patient is a member of the target population, and also that each member of the population is equally likely to be the patient whose risk is being assessed. These assumptions are often problematic. Moreover, Risk GP ignores the ominous reference class problem. When any of Risk GP's assumptions fail, we could invoke an alternate model of prediction such as mechanistic prediction or predicting from personal clinical experience. To benefit from the use of alternate models, medicine should embrace prediction model pluralism rather than model inflexibility.

The standard approaches to chronic disease care and evidence-based reasoning that I presented in this dissertation are normative ideals: WHO and other public health bodies advocate targeting chronic disease risk factors; the principles of rational drug discovery dictate that we should design new interventions using our understanding of mechanisms; and EBM recommends generalizing/particularizing study results, relying on RCTs, and using simple extrapolation. Thus, it seems that the new medical model is a *normative* model of practice. Whether or not it is descriptive of the practice of medicine in the year 2016 - whether or not physicians or the guidelines upon which they rely enact the new medical model - is an empirical question. Regardless of whether or not it fully describes modern medicine, it exerts a powerful influence on medicine and deserves further careful study.

Many of the puzzles that confront the new model, including the ones I just summarized, are at their core philosophical problems. Philosophers often offer their services to the world by remarking that philosophy can help with problems by clarifying them - by cleaning up our concepts and making explicit our assumptions. This remark is accurate, if perhaps a bit conservative. I would add that philosophers can help with philosophical problems by *solving them* (or, as the case may be, by *dissolving* them). If many problems facing medicine are at their core philosophical, and if philosophers can help to clarify and

solve philosophical problems, it follows that philosophers can help with many of the problems facing medicine.[113]

At the outset, I proposed a Central Thesis: many practical problems in chronic disease care and in EBM are intimately connected to conceptual, metaphysical and epistemic problems (including the philosophical questions I addressed in this dissertation). Throughout the previous chapters, I gestured towards some of these connections. In order to defend my thesis, I will now briefly revisit the ten problems of modern medicine that we started with to see how they are illuminated by the results of my philosophical project. I will also propose a few philosopher's prescriptions, drawing mostly on the preceding discussion. Rather than quick fixes ('take these philosophical theories and call me in the morning'), these prescriptions are hefty challenges calling for more hard thinking.

8.2 Problems of Modern Medicine: Philosophical Diagnoses and Prescriptions

8.2.1 Reductionism

Critics often charge medicine with being overly 'reductionistic', yet there are different ways of understanding reductionism in philosophy. For Leen De Vreese and colleagues, physicians and medical researchers committed to "methodological reductionism" always explain diseases "by reference to constituent components of their bearers" (2010, 372); in other

---

[113] In 2014, I attended a panel discussion on big data versus hypothesis-driven science at a large annual conference, the 64th Lindau Nobel Laureate Meeting. Four science Nobel Laureates sat on the panel. During the question and answer period, an audience member asked the panel what they thought a hypothesis was (a question that seemed to me to be quite relevant, given that the session explored hypothesis-driven research). In front of an audience of over six hundred junior medical researchers, the panel moderator dismissed the question as a 'philosophical question', and proceeded to the next audience member in the queue. At the same meeting, I presented a talk titled "Philosophy in the Service of Biology in the Service of Medicine" at a breakout session. As far as I could tell, my main message, captured in the title of the talk, fell on mostly deaf ears. If I am right that philosophy can assist medicine, there is work to be done in convincing medical scientists and physicians (and philosophers) of this point. Thankfully, I have encountered many scientists, physicians and philosophers who are already convinced.

words, doctors explain diseases by referring to parts of the body - genes, cells and so on. Diagnosing or identifying chronic diseases is reductive in a related sense. I argued in Chapter 2 that chronic diseases are bodily properties, most commonly properties of body *parts*. For instance, coronary artery disease (CAD) is a property of the patient's coronary arteries. Thus, diagnosing someone with a chronic disease typically involves reference to diseased parts of the patient's body. If instead chronic diseases referred to properties of whole persons (e.g. behaviours) or populations (e.g. average outcomes), no reductionism would be implied. Furthermore, I suggested in Chapter 3 that most etiologic mechanisms in medicine are represented at reductive levels. Reductive mechanistic explanations influence medicine's prevention efforts by directing the medical gaze to lower levels in the decomposition hierarchy. In diagnosing and explaining chronic diseases, reductive explanations are certainly privileged.

De Vreese and colleagues (2010) suggest an alternate strategy for medicine: "explanatory pluralism". In explanatory pluralism, different levels of explanation are chosen in different contexts in order to best suit different explanatory tasks. For instance, in preventing chronic diseases, interventions at the social-behavioural level might often be the best medicine; thus, explaining chronic diseases at this non-reductive level might help us to identify good targets on which to intervene. Similarly, adopting a pluralist strategy in the diagnosis of chronic health problems – for instance, chronic pain or chronic shortness of breath – might sometimes require us to seek non-disease explanations for these conditions, given that our current chronic diseases are reductive in the sense I just described. This would also require a research agenda dedicated to uncovering the non-biological causes of chronic illness. When the ultimate aim is to intervene and ameliorate suffering, the optimal level of explanation and diagnosis should depend on the level at which intervention is best directed.

8.2.2 Fractured Care

A patient's care becomes fractured when it is parcelled into pieces without continuity among the pieces, like a jigsaw puzzle that is taken apart and never put back together. For patients with multimorbidity, care is often divided into multiple pieces corresponding to their multiple chronic diseases. In Chapter 2, I distinguished between token multimorbidity and

type multimorbidity. Token multimorbidity refers to the co-occurrence of multiple numerically distinct chronic disease properties, while type multimorbidity refers to the co-occurrence of multiple conceptually distinct chronic disease categories. Reductive explanations that refer to chronic disease properties thus provide a basis for fractured care. A primary care physician might send their patient with multiple disease properties to several organ system specialists - a cardiologist for a disease property of the heart, a respirologist for a disease property of the lungs, and so on.[114]

Diagnosing a patient's chronic diseases as belonging to multiple distinct types (type multimorbidity) can also lead to a fracturing of care when the Risk GP Model is used to make multiple therapeutic predictions. When multiple diagnoses apply to a single patient, that patient belongs to multiple target populations, each defined according to a different disease. Best practice – codified in practice guidelines – often requires the use of interventions to prevent or manage outcomes associated with particular diseases. When a patient with multiple diagnoses belongs to multiple target populations, applying the Risk GP Model for each of these target populations results in multiple isolated therapeutic predictions. If the predicted effects are favourable, the physician and patient might then implement multiple interventions. Unfortunately, as the studies from which we extrapolate typically investigate the effects of only one intervention, the Risk GP Model fails to consider how the use of other interventions should modify our predictions. Care can thus become fractured or uncoordinated even when it is managed by a single healthcare provider.

In Chapter 2, I suggested that recognizing convergence might help to mitigate the problem of fractured care. Convergence occurs when multiple disease tokens belong to the same disease type, or when multiple disease types refer to the same disease token. When care is parceled out according to disease types or disease tokens, identifying convergence – in research and in practice – might help to reduce the number of 'care parcels'. Convergence also refers to special causal relations among diseases. Two diseases might have a common cause (pathogenesis) or a common effect (manifestation), or one disease might cause the

---

[114] Systemic lupus erythematosus provides a counterexample to fractured care. Lupus typically manifests pathological states in multiple organ systems, yet many patients are managed mostly by a single specialist, a rheumatologist, in part by treating the autoimmune process common to the various manifestations.

other. In these cases, even if the diseases are distinct, care for each of these diseases need not be discrete; we can organize care around the common pathogenesis or the common manifestation, or around the one disease that causes the other. If research and practice are to better take advantage of convergence, it will require a research agenda dedicated to investigating the causal relations among diseases rather than single isolated diseases.[115]

8.2.3 Multifactorialism

One worry about the 'multifactorial' etiology of chronic diseases is that because there are several etiologic factors for each type of disease (none of them necessary), preventing chronic diseases is not straightforward. In Chapter 3, I argued that all diseases, including acute infectious diseases, are multifactorial, and that chronic diseases do not have necessary causes only because we do not define them monocausally. Multifactorialism is the wrong way to think about chronic disease causation and prevention. With respect to prevention, the problem is not that chronic disease types have multiple causes but that they have multiple etiologic mechanisms. They have multiple etiologic mechanisms because we define them constitutively, as particular types of properties, and the same type of property can arise from multiple types of etiologic mechanism. Consequently, there is no one type of mechanism on which we can intervene to prevent all cases of the disease.

My prescription for disease classification and prevention was the monomechanism ideal. A particular disease type satisfies the monomechanism ideal when it has one etiologic mechanism represented by one mechanistic model. Once we have a mechanistic model, we can then represent an intervention on the mechanism, and attempt to design a real world intervention corresponding to the hypothetical intervention. I argued in Chapter 4 that we

---

[115] Physicians do sometimes target the convergent pathogenesis or manifestations of chronic diseases. They aim to prevent cardiovascular disease (CVD) events that can result from multiple comorbid diseases (e.g. hypertension, type II diabetes), and they do so by constructing disease-nonspecific target populations (e.g. 'patients at high risk for CVD'). To prevent CVD events, they then use interventions (e.g. statins) that disrupt the common pathogenesis of CVD events (atherosclerosis). Notwithstanding exceptions, the dominant approach in science is to investigate interventions tailored to individual diseases or conditions, interventions that are born in research labs using single disease causal models and that are tested in single disease human studies.

might fail to predict the results of intervening on a mechanism if our model does not have appropriate scope. If our mechanistic model represents some but not all of the etiologic mechanisms for a particular type of disease, an intervention will at best prevent some cases of the disease. We might incorrectly predict that the intervention will be effective in a case that falls outside of the scope of the model. The monomechanism ideal helps us to avoid these faulty predictions.

For many chronic diseases, the only way to achieve the ideal is to redefine disease categories according to etiologic mechanisms. At present, our mechanistic models representing these mechanisms are often no more than mechanism sketches. Yet we also saw that not all incomplete models are poor predictors. Defining new categories of chronic disease using mechanism sketches might also help to better align research with prevention; because so much research is disease-centered, the monomechanism ideal ensures that research is organized around particular types of etiologic mechanism.

8.2.4 The Growing Burden of Chronic Diseases

The story we often tell about the growing global prevalence of chronic diseases is that populations are aging and chronic diseases accumulate with age. In Chapter 2, I noted that chronic diseases are generally incurable, which explains why cohorts accumulate chronic diseases over time. I argued that this incurability is incurability-at-present rather than incurability-in-principle. Chronic diseases are incurable because we do not presently possess the knowledge or therapeutics required to cure. Thus, incurability reflects the state of medical science and scientific medicine. I also noted in Chapter 3 that one way to curb the growing burden is to reduce the incidence of chronic diseases. Medicine and public health confront several difficulties in preventing chronic diseases. First, our preventive medications are often minimally effective (which again explains the growing burden of chronic diseases in terms of scientific medicine's own limitations).

Moreover, many important causal risk factors for chronic disease are behavioural or social. Individuals often struggle with behaviour modification, and societies often struggle with social change. I argued that one reason for our struggles might be that we often do not represent the pathogenesis of a chronic disease at the 'right level'; the pathogenesis is

understood at reductive levels of organization instead of social-behavioural levels. However even with better prevention, halting completely the growing burden of chronic diseases may be an unrealistic goal. In addition to prevention and cure, medicine and public health must find new ways of managing chronic diseases effectively and affordably. We may require new models of clinical management that depart from our current medical model.

8.2.5 Treatment Effect Heterogeneity, Treatment Futility and Treatment Harm

Why are modern interventions commonly heterogeneous, ineffective or harmful in their effects? The obvious answer would seem to be that there is something wrong with these interventions. Yet heterogeneity, effectiveness and harm are not purely properties of interventions. They depend on the population in which the intervention is used. A treatment that is heterogeneous in its effects within one population is homogenous in its effects within a subset of that population, and an intervention that is ineffective or harmful in one group might be helpful in another group. A complete explanation for treatment effect heterogeneity, treatment futility and treatment harm must then address why we often fail to predict which patients will benefit from treatment and which will not. In Chapter 4, I argued that our mechanistic models of interventions often fail to adequately predict heterogeneity, futility and harm for several general reasons. Sometimes our model is incorrect; it might thus predict that an intervention is beneficial when in fact the intervention is futile or harmful. Other times our model is incomplete or too abstract, which makes it difficult to determine the model's scope and distinguish treatment responders from non-responders.

Ineffective and harmful interventions are generally weeded out during premarketing clinical trials, but many interventions that are heterogeneous in their effects are approved for marketing. In Chapter 5, I argued that the standard model of prediction is to generalize these treatment effects to a target population, and then particularize them for a patient. When treatment effects are heterogeneous (the effect size is neither insignificant nor enormous), we cannot be certain of individual patient responses, and we quantify our uncertainty probabilistically. As I argued in Chapter 7, because treatment effects are heterogeneous in clinical trials we should also worry that there are many target populations to which they are

not transportable, which complicates generalization (generalizability will be discussed in section 8.2.9).

To better predict heterogeneity, futility and harm, we must construct better mechanistic models of interventions, and also learn to make better inferences from the results of intervention studies. But given that our inferences will likely remain highly uncertain, we must also learn to better incorporate all sources of uncertainty in our predictions and decision-making.

8.2.6 Cookbook Medicine

EBM promotes several standard approaches to interpreting the results of epidemiological studies. Among them are the Risk GP Model of prediction and the simple extrapolation approach to the problem of extrapolation. The very use of a single solution to each problem is inflexible. Moreover, in their own ways the Risk GP Model and simple extrapolation are simple recipes. Simple extrapolation relies upon the same unsophisticated aphorism every time ('extrapolate the relative effect size unless a compelling reason is had'). It assumes that relative risks are generally generalizable, or generalizable in most instances. On the other hand, particularization in the Risk GP Model predicts the same effect for each patient in the target population, it does not discriminate among unique individuals. Both simple extrapolation and Risk GP assume that the clinical world is uniform, that all patients are cut from the same mold. It is exactly this stance towards which critics object when they accuse modern evidence-based medicine of using clinical cookbooks.

In Chapter 5, I argued that prediction model pluralism is a defensible alternative to having a standard model of prediction. In model pluralism, the physician pulls out whichever prediction model is best suited to the circumstances. Model pluralism is a healthy antidote to cookbook medicine because it forces us to consider contextual factors that should influence the choice of prediction model, as well as whether the assumptions upon which each model relies obtain in each unique case.[116]

---

[116] Prediction model pluralism calls to mind an older tradition in epidemiology reflected in Bradford Hill's (1965) 'viewpoints' on causation. Rather than a rigid approach to inferring causation based on the results of a single study, Hill promoted a pluralistic approach in which

8.2.7 The Tyranny of Aggregate Outcomes

Because our standard model of prediction relies centrally on aggregate outcomes in populations, there is a sense in which clinical medicine is increasingly ruled by population-level thinking, and patients are seen merely as members of target populations. However, the consideration of aggregate outcomes in clinical reasoning does not ipso facto jeopardize individualized patient care. Population study results can provide evidence for conclusions about individuals. The Risk GP Model makes the inferential move from target populations to patients by transforming risk measures into probabilities. Here again, probabilistic reasoning is not necessarily impersonal. The issue with the Risk GP approach is that the probabilities are typically conditional on very few clinical features. Our target populations are usually broadly defined, and so probabilities derived from frequencies in those populations will often spare relevant patient information.

In Chapter 5 I suggested that Risk GP must confront the reference class problem, and that a 'narrowest reference class'-style solution to the problem would plausibly help make particularization more particular. I also suggested a few ways of modifying our probabilities to account for more patient-relevant evidence: by using the total probability equation or Bayes' Theorem, or by updating our probabilities qualitatively. These strategies adopt a Bayesian approach to prognosis and therapy that is much lacking in a medicine largely ruled by frequentist statistics. In a Bayesian approach to prediction, individual risks represent credences that can be modified by any and all evidence that is relevant to the prediction. While a frequentist interpretation of risks view risks as frequencies in a population, a Bayesian interpretation of individual risks sees them as degrees of belief bearing on our predictions about the individual patient's outcome(s).

---

multiple kinds of evidence - including experimental and non-experimental studies as well as theory - are brought to bear on the inference.

8.2.8 RCT Worship

Philosophers, epidemiologists and clinicians often worry that EBM overemphasizes the reliability of RCT evidence relative to other kinds of evidence, as well as the importance of randomizing in comparative group studies. In Chapter 6, I examined one argument in favour of randomization: the assumption that randomization balances all confounding causes. I argued that the assumption is false, but also that balancing confounding causes is not directly relevant for causal inference. Therefore, we should not judge the strength of RCT evidence on the basis of this claim. I developed my own theory of causal inference in which the distribution of condition C determines the soundness of our causal inference. Randomization prevents systematic imbalances in correlates of C at baseline, or 'selection bias'. At the same time, my account shows that comparability among study groups with respect to C is what is ultimately important for sound causal inference.

What is the upshot of this discussion for the debate surrounding the epistemic superiority of the RCT? First, randomization need not figure as a premise in our causal inference, it is not a necessary condition for good inference. This recognition alone should quiet the loudest among the cheers for the RCT. On the other hand, when we are genuinely concerned about the possibility of selection bias in an observational group study, an RCT is likely to have greater internal validity; in other words, our causal inference based on the RCT results is more likely to be sound (all other things being equal). However, I also argued that in addition to how closely it approaches the ideal study, we can judge a real study according to whether or not it achieves the *required* conditions for causal inference, which are far less demanding. In any comparative group study, if the difference in outcome is greater than the difference in distribution of C, we can conclude that the intervention was effective. Therefore, when the difference in outcome is large enough and the magnitude of bias is likely to be small enough, we *can* draw certain causal conclusions in an observational group study, despite what the most ardent supporters of clinical experiments claim.

Sometimes, an observational study would be less reliable than an RCT, but nonetheless reliable enough. In deciding which kind of study to run – RCT or observational study – or which study results to extrapolate, investigators and clinicians must always consider other factors alongside internal validity, including external validity, ethical

concerns, and costs. (As we will see in the section 8.2.9, RCTs may frequently lack external validity.) On the balance of considerations, an observational study will sometimes be the best choice.

8.2.9 Unrepresentative Trials

Randomized trials are often unrepresentative of clinical practice; they usually include a narrow subset of patients, and they often manufacture conditions that do not resemble the conditions of the real-world clinic. In Chapter 5, we saw why unrepresentativeness is a potential problem. Generalization presupposes a representativeness assumption, which states that the trial and target population are sufficiently similar in the right ways. Without the assumption, we are not justified in transporting the effect size from the trial to the target. Because many of the ways that trials and targets differ are plausibly relevant to how individuals respond to an intervention, we should worry that the representativeness assumption does not hold for these trials. This argument gives credence to countless worries about the external validity of usual RCTs.

Why are our trials so unrepresentative? One major reason why most trials do not mirror practice is that they typically exclude large subgroups of patients. The argument is sometimes made that exclusions enhance the internal validity of RCTs. Some authors even admit that exclusions threaten external validity, but argue nonetheless that sacrificing external validity for internal validity is a good trade-off. Others argue that exclusions protect internal validity specifically by distilling a more homogeneous trial population. If the trial group is more homogeneous, there are fewer covariates for randomization to balance, and thus it is more likely that all covariates will be balanced. However, I argued in Chapter 6 that balancing all covariates is neither necessary, nor sufficient, nor the correct ideal to which we should aspire. Thus, this homogeneity rationale is no defense of the all too common practice of excluding patients from RCTs. Although there might be other reasons – perhaps good reasons – for excluding certain patients from trials, in general we should strive for representative trials if we wish to extrapolate from trial results. Exclusions from clinical trials should be far fewer, and should be explicitly justified whenever they are made. A much

greater number – if not the vast majority - of our RCTs should be pragmatic RCTs, which are designed to be representative of clinical practice.

How should evidence users extrapolate from RCTs, given that RCTs are usually unrepresentative of clinical practice and that the representativeness assumption is poorly articulated? EBM's solution to the problem is simple extrapolation. In Chapter 7, I argued that simple extrapolation is seriously flawed and might lead us to extrapolate trial results more often than we should. Simple extrapolation relies on the general generalizability assumption, which states that trial results are generalizable to the target population in most cases. I argued that trial unrepresentativeness provides evidence against general generalizability. Instead of simple extrapolation, EBM and the health sciences need rigorous approaches to assessing the external validity of trials informed by a science of extrapolation, analogous to the evolving science of causal inference. A science of extrapolation would help us to formalize the conditions under which aggregate effects travel from one population to another; it would thus help us to more precisely fill in generalization's representativeness assumption. A science of extrapolation would also help us develop tools that could assist clinicians in assessing whether the representativeness assumption is satisfied on a given occasion.

8.2.10 Multimorbidity

Multimorbidity epitomizes many of the philosophical and practical problems we have already described, while raising several unique challenges. The nature of multimorbidity was previously unexplored. On close analysis, there are two types of multimorbidity that may generally map onto one another, but will sometimes come apart, creating problems for researchers measuring multimorbidity and physicians managing multimorbidity. Counting the number of diagnoses that a patient carries might overestimate or underestimate the number of distinct disease tokens. Counting types will overestimate the number of tokens if some types apply to the same token. Meanwhile, counting types will underestimate the number of tokens if some types apply to more than one token or if some tokens are undiagnosed.

As I have already mentioned, care is vulnerable to fracturing in patients with multimorbidity whenever care is organized around chronic disease tokens or types. Fractured care can in turn lead to drug-drug and drug-disease interactions. Many examples of drug-drug and drug-disease interactions are paradoxical mechanisms, in which an intervention both promotes and inhibits the same outcome. In fractured care, when interventions and diseases are managed in isolation from one another, a physician may fail to predict that an intervention will have a counteracting effect through an unconsidered second pathway involving another drug or disease.

Cookbook medicine is especially concerning for patients with multimorbidity because unique patients will often have unique constellations of diseases. Applying the Risk GP Model to the care of a patient with multimorbidity may result in overtreatment when patients with multiple diseases belong in multiple disease-defined target populations. Overtreatment imposes significant costs and treatment burden with diminishing returns. Moreover, comorbidities locate patients in narrower reference classes within disease-defined target populations. Unfortunately, we typically generalize to the wide population, and fail to take account of comorbidities when we subsequently particularize.

Patients with multimorbidity are among those most frequently excluded from clinical trials. The practice perhaps presupposes that comorbidities are confounding causes, but confounding causes are not directly relevant in the account of causal inference that I developed. Some comorbidities might be useful prognostic factors. Rather than excluding these factors from trials, we should build them into the statistical design and analysis. Unfortunately, because trials commonly exclude patients with multimorbidity, the generalizability of trials to the population with multimorbidity is highly problematic. This problem should call into question the purportedly superior reliability of RCTs for grounding predictions in routine practice. We are instructed to cross the large void between trials and populations with multimorbidity using simple extrapolation. As simple extrapolation disposes evidence users towards over-extrapolation it contributes to the problem of overtreatment in the multimorbid population.

Because chronic diseases are generally incurable, not only does a population tend to accumulate chronic diseases with age, but so too do individuals. There is thus a growing burden of multimorbidity in modern medicine. The problems of multimorbidity, chronic

disease and evidence-based medicine should call into question our disease-centered medical model. The telos of the new medical model is the treatment of individual disease, which leads to the treatment of individual diseases when patients have multimorbidity. Various authors have advanced rival approaches to our present medical model, including person-centered medicine (Miles & Mezzich 2011). In person-centered medicine, the unit of analysis is the whole patient rather than their diseases. The ethics, epistemology and metaphysics (as well as the practice) of person-centered medicine are yet to be worked out in detail, but it offers a promising alternative to disease-centered care (Fuller & Upshur 2014).

In summary, many practical problems in modern medicine are tangled up with philosophical problems. It will require both medical and philosophical wisdom to unravel them.

8.3 Future Directions

There are many questions left unexplored by this dissertation. I did not examine diagnosis, the social context of medicine, informal clinical judgment, illness experience, nor a great many other topics. I put aside ethical questions altogether. Even within the topics I penetrated – the medical model, the problem of extrapolation, medical science – many interesting questions remain. Now that I have looked back at where we have been and around at where we have arrived, I will end by looking ahead at three future research projects.

8.3.1 What is Disease? The Biomedical and Biopsychosocial Models Revisited

We can identify at least two core commitments of the biomedical model according to Engel (1977), the model's *telos* and its *disease concept*. The biomedical model's telos is disease-centeredness: diagnose and treat disease. This commitment of the biomedical model is what its opponents typically criticize when they call for more humane (Marcum 2008a; 2008b) or person-centered medicine (Miles & Mezzich 2011). It is often forgotten that Engel's original critique of the medical model (1977) contained a particular concept of disease.

According to Engel, the biomedical model "assumes disease to be fully accounted for by deviations from the norm of measurable biological (somatic) variables" (2012, 379);

"sufficient deviation from normal represents disease" (2012, 378). Engel describes deviation from the norm in the language of natural functions. It thus seems that the disease concept Engel has in mind is what we might now describe as a 'Boorsean' concept; it resembles Christopher Boorse's biostatistical theory (1977, 2011), in which disease is deviation from natural biological functioning. Regardless of these extra details, minimally speaking Engel's disease concept defines diseases as biomedical entities, entities described in the language of the basic biomedical sciences. I will call this idea the *biomedical disease concept*. As Engel noted, it is a reductionistic concept (relative to the level of the whole person) as the entities described by the basic biomedical sciences (cells, molecules and organs) are parts of the person.

Engel (1977) described an urgent crisis facing psychiatry, a crisis felt at psychiatric conferences and spilt over the pages of medical journals. Psychiatric illnesses did *not* seem to fit the biomedical model of disease. In order to conform to medicine's biomedical disease concept, psychiatry was given a choice. It would either have to reduce mental illnesses to biological diseases (the "reductionist" option), or break away from medicine altogether (the "exclusionist" option). Engel argued that psychiatrists should escape between the horns of this dilemma by rejecting the biomedical model as an adequate model of medicine. Many current advocates of rival medical models (e.g. "humanistic" or "humane medicine" (Marcum 2008b), "person centered medicine" (Miles & Mezzich 2011)) similarly reject the biomedical model. However, they often do so by rejecting its telos, disease-centeredness. Another – perhaps more radical - route is available to the medical model's detractors: they could reject biomedicine's disease concept.

This route is perhaps the one taken by Engel (1977) when he rejects the biomedical model in favour of a biopsychosocial (BPS) model. On one reading of the BPS model, the boundaries of disease are widened to include not only biomedical entities but also entities described by the psychological and social sciences. It is worthwhile considering this broadening of our disease concept in light of several prominent problems in modern medicine, including chronic illness, medically unexplained physical symptoms (MUPS), multimorbidity, and overtreatment. I will explore implications of a *biopsychosocial disease concept* for medicine. One objection is easily dealt with. It might be argued, contra a BPS disease concept, that just as no mammals are plants, no behaviours or social conditions are

diseases. Against this essentialist objection, I will argue that the important question is not whether disease essentially includes the psychosocial, but whether it should.

8.3.2 Two Problems of Extrapolation

The problem of extrapolation, or the problem of external validity, is a venerable one in many fields - including economics, toxicology, evidence-based policy and evidence-based medicine - in which the results of a study are extrapolated to a target context. In medicine, researchers measure the effect size of interventions in population studies, but clinicians are ultimately interested in knowing the effect size in a target population of patients like theirs. The problem of extrapolation asks: when can we extrapolate the study results to the target population?

Marcellesi (2015) argues that as far as predictive inferences like this one go, the problem of extrapolation has been solved (by Cartwright & Hardie (2012), among others). I will argue that there is no '*the* problem of extrapolation'; instead there are two problems of extrapolation, a metaphysical problem and an epistemic problem. The solutions to which Marcellesi refers are solutions to the metaphysical problem: under what sufficient circumstances are the effects in the study and target the same? The epistemic problem of extrapolation instead asks: how can we *know* that the effects in the study and target are the same? Marcellesi argues that a solution to the problem of extrapolation need not confront the 'extrapolator's circle' (Steel 2010), the problem of how we can establish that the study and target are sufficiently similar without establishing the effect in the target (thus rendering its measure in the study irrelevant). However, while the metaphysical problem can ignore the extrapolator's circle, the epistemic problem cannot. Therefore, before we can declare the problem of extrapolation fully resolved we must solve the epistemic problem and show how to break the extrapolator's circle.

I will argue that a satisfactory *theoretical* solution to the epistemic problem must satisfy the following criteria. It must explain what kind of *theoretical knowledge* and *evidence* can establish that sufficient circumstances - described in a particular solution to the metaphysical problem -obtain; as well as why the theory plus evidence alone do not allow us to predict the effect in the target.

I will identify two general solutions to the epistemic problem. The first solution involves theoretical knowledge of all interactive causes relevant to the effect, and evidence that each cause is distributed similarly in the study and target. The second solution involves knowledge of all of the mechanisms relevant to the effect, and evidence that each mechanism is distributed similarly. In both cases we break the extrapolator's circle so long as we do not know the quantitative functional relationship between the causes/mechanisms and the effect. This would often be the case in medicine because the relevant mechanisms are typically understood and represented diagrammatically.

8.3.3 Medicine Lost in Translation

What are the aims of medical science? To generate knowledge? To produce accurate diagnostic tests and effective treatments? In this paper, I will argue that to understand the aims of medical science we must examine how social structures, including the organization of research, regulatory policies and financial incentives influence the outcomes of research.

I will understand the 'aims of medical science' to mean the products that medical science tends to produce, the ends toward which medical science is aimed rather than the goals that scientists, sponsors, doctors or patients have in mind, individually or collectively. The primary aim of medical science is represented by a currently popular buzzword: *translation*. The concept of translation is captured in many slogans: 'translational medicine' (Marincola 2003), 'translational research' (Woolf 2008), 'knowledge translation' (CIHR 2015), 'from bench to bedside' (Pizzo 2002), and 'closing the research-practice gap' (Contopoulos-Ioannidis et al. 2003). 'Translational medicine' often refers to a new movement that began in the 1990s, but this movement is driven towards advancing the established aims of medical science rather than reforming them. As Solomon argues, "Translational medicine is a new term for endeavors that are as old as scientific medicine" (2015, 176).

I will first translate the language of translation in order to clarify the primary aim of medical science. Translation refers to scientific efforts that generate, evaluate and implement new medical technologies (tests and treatments) that have cleared certain bars. Thus*, the primary aim* of medical science is *empirically successful technologies*. Empirical success is

measured through certain tests such as biological plausibility, animal testing, and especially clinical trials.

There are other aims of medical science. The *primary* aim of medical science is determined by the structure of the research enterprise. All major public health science research granting agencies in the US, UK and Canada as well as private industry have embraced the translation paradigm. Many of our standard tests of empirical success are set by regulators, who must negotiate the sometimes conflicting aims of patients, physicians and industry.

Medical science does not primarily aim towards knowledge or truth about its theories or models; nor does it primarily aim towards truly safe and effective treatments because our tests of empirical success do not directly measure effectiveness and safety in the very patients that will use them. The skepticism I am advocating here is analogous to van Fraassen's (1980) antirealism about scientific theories, and does not imply that our best medical theories are false nor that our best treatments are ineffective. But it should lead us to carefully evaluate whether our current measures of empirical success track truth and effectiveness, especially considering widespread concerns that various interests (e.g. industry) are exploiting medical research for their own aims (e.g. financial profit).

8.4 Final Conclusion

To know chronic disease and EBM is to know a great deal about modern medicine. Chronic diseases are bodily properties generated through etiologic mechanisms; representing the mechanisms using mechanistic models provides a guide to intervention, but is fraught with challenges. Evidence-based medicine promotes predicting from population studies; and here too problems arise in causal inference, extrapolation, and predicting for individuals. The new medical model of evidence-based chronic disease care throws up philosophical as well as practical problems. Resolving the problems will require contact and collaboration between philosophers and physicians, recalling that ancient Galenic union of philosophy and medicine.

References

Ackerknecht, Erwin H. 1982. *A Short History of Medicine*. Revised Edition. Baltimore: Johns Hopkins University Press.

Akker, Marjan van den, F. Buntinx, and J. A. Knottnerus. 1996. "Comorbidity or multimorbidity: what's in a name?" *European Journal of General Practice* 2:65-70.

Andersen, Holly. 2012. "Mechanisms: what are they evidence for in evidence-based medicine?" *Journal of Evaluation in Clinical Practice* 18 (5):992-999. doi: 10.1111/j.1365-2753.2012.01906.x.

Armstrong, D.M., C.B. Martin, and U.T. Place. 1996. *Dispositions: A Debate*. London: Routledge.

Arnold, J. M. O., P. Liu, C. Demers, P. Dorian, N. Giannetti, H. Haddad, G. A. Heckman, J. G. Howlett, A. Ignaszewski, D. E. Johnstone, P. Jong, R. S. McKelvie, G. W. Moe, J. D. Parker, V. Rao, H. J. Ross, E. J. Sequeira, A. M. Svendsen, K. Teo, R. T. Tsuyuki, and M. White. 2006. "Canadian Cardiovascular Society consensus conference recommendations on heart failure 2006: Diagnosis and management." *Canadian Journal of Cardiology* 22 (1):23-45.

ATS. 2000. "Diagnostic Standards and Classification of Tuberculosis in Adults and Children." *Am J Respir Crit Care Med* 161 (4):1376-95.

Ayer, A.J. 1963. "Two Notes on Probability." In *The Concept of a Person and other Essays*, 188-208. Macmillan.

Baigent, C., A. Keech, P. M. Kearney, L. Blackwell, G. Buck, C. Pollicino, A. Kirby, T. Sourjina, R. Peto, R. Collins, J. Simes, and Cholesterol Treatment Triallist Collaborators. 2005. "Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins." *Lancet* 366 (9493):1267-1278.

Bailey, K. R. 1994. "Generalizing the results of randomized clinical trials." *Control Clin Trials* 15 (1):15-23.

Bajcar, Jana M., Li Wang, Rahim Moineddin, Jason X. Nie, C. Tracy, and Ross E. G. Upshur. 2010. "From pharmaco-therapy to pharmaco-prevention: trends in prescribing to older adults in Ontario, Canada, 1997-2006." *BMC Family Practice* 11 (1):75. doi: 10.1186/1471-2296-11-75.

Balshem, Howard, Mark Helfand, Holger J. Schünemann, Andrew D. Oxman, Regina Kunz, Jan Brozek, Gunn E. Vist, Yngve Falck-Ytter, Joerg Meerpohl, and Susan Norris. 2011. "GRADE guidelines: 3. Rating the quality of evidence." *Journal of Clinical Epidemiology* 64 (4):401-406.

Barnes, Peter J. 2012. "Chapter 254. Asthma." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Barnett, Karen, Stewart W. Mercer, Michael Norbury, Graham Watt, Sally Wyke, and Bruce Guthrie. 2012. "Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study." *Lancet* 380 (9836):37-43.

Bateman, E. D., S. S. Hurd, P. J. Barnes, J. Bousquet, J. M. Drazen, M. FitzGerald, P. Gibson, K. Ohta, P. O'Byrne, S. E. Pedersen, E. Pizzichini, S. D. Sullivan, S. E. Wenzel, and H. J. Zar. 2008. "Global strategy for asthma management and prevention: GINA executive summary." *Eur Respir J* 31 (1):143-78. doi: 10.1183/09031936.00138707.

Bechtel, William. 2008. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Oxford: Routledge.

Bechtel, William. 2015. "Can mechanistic explanation be reconciled with scale-free constitution and dynamics?" *Studies in history and philosophy of biological and biomedical sciences* 53:84-93.

Bechtel, William, and Adele Abrahamsen. 2005. "Explanation: a mechanist alternative." *Studies in history and philosophy of biological and biomedical sciences* 36 (2):421-41.

Beeson, P. B., and W. McDermott. 1975. *Textbook of Medicine*. 14 ed. ed. Philadelphia: W. B. Saunders Company.

Beresford, M. J. 2010. "Medical reductionism: lessons from the great philosophers." *QJM* 103 (9):721-724. doi: 10.1093/qjmed/hcq057.

Bernstein, AB, E Hing, AJ Moss, KF Allen, AB Siller, and RB Tiggle. 2003. Health Care in America: Trends in Utilization. Hyattsville: National Center for Health Statistics.

Binney, Nicholas. 2015. "Nosology, ontology and promiscuous realism." *Journal of Evaluation in Clinical Practice* 21 (3):391-397. doi: 10.1111/jep.12274.

Black, D. 1998. "The limitations of evidence." *Perspectives in Biology and Medicine* 42 (1):1-7.

Bliss, Michael. 2011. *The Making of Modern Medicine: Turning Points in the Treatment of Disease*. Toronto: University of Toronto Press.

Bluhm, R. 2005. "From hierarchy to network - a richer view of evidence for evidence-based medicine." *Perspectives in Biology and Medicine* 48 (4):535-547. doi: 10.1353/pbm.2005.0082.

Bluhm, R. 2013. "Physiological mechanisms and epidemiological research." *J Eval Clin Pract* 19 (3):422-6. doi: 10.1111/jep.12035.

Bluhm, R., and Kirstin B. 2011. "Evidence-based medicine." In *Handbook of the Philosophy of Science: Volume 16: Philosophy of Medicine*, edited by D. M. Gabbay, P. Thagard and J. Woods, 203-238. Amsterdam: Elsevier.

Boorse, Christopher. 1977. "Health as a Theoretical Concept." *Philosophy of Science* 44:542 - 573.

Boorse, Christopher. 2011. "Concepts of health and disease." In *Handbook of the Philosophy of Science: Volume 16: Philosophy of Medicine*, edited by D. M. Gabbay, P. Thagard and J. Woods, 13-64. Amsterdam: Elsevier.

Borgerson, Kirstin. 2009. "Valuing Evidence: Bias and the Evidence Hierarchy of Evidence-Based Medicine." *Perspectives in Biology and Medicine* 52:218 - 233.

Bornhoft, G., S. Maxion-Bergemann, U. Wolf, G. S. Kienle, A. Michalsen, H. C. Vollmar, S. Gilbertson, and P. F. Matthiessen. 2006. "Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity." *BMC Med Res Methodol* 6:56. doi: 10.1186/1471-2288-6-56.

Boyd, C. M., J. Darer, C. Boult, L. P. Fried, L. Boult, and A. W. Wu. 2005. "Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance." *Journal of the American Medical Association* 294 (6):716-724. doi: 10.1001/jama.294.6.716.

Brennan, T., and W. Shrank. 2014. "New expensive treatments for hepatitis C infection." *JAMA* 312 (6):593-4. doi: 10.1001/jama.2014.8897.

Britton, A., M. McKee, N. Black, K. McPherson, C. Sanderson, and C. Bain. 1999. "Threats to applicability of randomised trials: exclusions and selective participation." *Journal of Health Services Research & Policy* 4 (2):112-21.

Broadbent, Alex. 2009. "Causation and models of disease in epidemiology." *Studies in history and philosophy of biological and biomedical sciences* 40 (4):302-11.

Broadbent, Alex. 2013. *Philosophy of Epidemiology*. Basingstoke: Palgrave Macmillan.

Broadbent, Alex. 2014. "Disease as a theoretical concept: the case of "HPV-itis"." *Studies in history and philosophy of biological and biomedical sciences* 48 Pt B:250-7. doi: 10.1016/j.shpsc.2014.07.010.

Broadbent, Alex. 2015. "Risk relativism and physical law." *Journal of Epidemiology and Community Health* 69 (1):92-94.

Byar, D. P., R. M. Simon, W. T. Friedewald, J. J. Schlesselman, D. L. DeMets, J. H. Ellenberg, M. H. Gail, and J. H. Ware. 1976. "Randomized clinical trials. Perspectives on some recent ideas." *N Engl J Med* 295 (2):74-80.

Campaner, Raffaella. 2010. "Understanding mechanisms in the health sciences." *Theoretical Medicine and Bioethics* 32 (1):5-17. doi: 10.1007/s11017-010-9166-5.

Campbell-Scherer, Denise. 2010. "Multimorbidity: a challenge for evidence-based medicine." *Evidence-Based Medicine* 15 (6):165-166. doi: 10.1136/ebm1154.

Campbell-Scherer, Denise. 2012. "The 11th hour-time for EBM to return to first principles?" *Evidence-based medicine* 17 (4):103-4.

Caplan, A. L. 1992. "Does the philosophy of medicine exist?" *Theor Med* 13 (1):67-77.

Carnap, Rudolf. 1945. "On Inductive Logic." *Philosophy of Science* 12 (2):72-97.

Carrier, M., G. Le Gal, P. S. Wells, and M. A. Rodger. 2010. "Systematic review: case-fatality rates of recurrent venous thromboembolism and major bleeding events among patients treated for venous thromboembolism." *Ann Intern Med* 152 (9):578-89.

Carter, Codell K. 2003. *The Rise of Causal Concepts of Disease*. Aldershot: Ashgate.

Cartwright, Nancy. 1989. *Nature's Capacities and their Measurement*. Oxford: Oxford University Press.

Cartwright, Nancy. 2007. "Are RCTs the Gold Standard?" *Biosocieties* 2:11 - 20.

Cartwright, Nancy. 2009. "Presidential Address." American Philosophical Association Pacific Division, Vancouver.

Cartwright, Nancy. 2010. "What are randomised controlled trials good for?" *Philosophical Studies* 147 (1):59-70. doi: 10.1007/s11098-009-9450-2.

Cartwright, Nancy. 2011a. Predicting 'It will work for us': (Way) beyond statistics. In F. R. Phyllis McKay Illari, and Jon Williamson (Ed.), *Causality in the Sciences*: Oxford Scholarship Online.

Cartwright, Nancy. 2011b. "A philosopher's view of the long road from RCTs to effectiveness." *Lancet* 377 (9775):1400-1401.

Cartwright, Nancy. 2012. "Will This Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps." *Philosophy of Science* 79 (5):973-989.

Cartwright N, and J. Hardie. (2012) *Evidence-Based Policy: A Practical Guide to Doing it Better*. New York: Oxford University Press.

CAST Investigators. 1989. "Preliminary report: effect of ecainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction." *New England Journal of Medicine* 321 (6):406-412.

Caughey, Gillian E., Elizabeth E. Roughead, Sepehr Shakib, Agnes I. Vitry, and Andrew L. Gilbert. 2011. "Co-morbidity and potential treatment conflicts in elderly heart failure patients: A retrospective, cross-sectional study of administrative claims data." *Drugs & Aging* 28 (7):575-581.

Charlton, Bruce G., and Andrew Miles. 1998. "The Rise and Fall of EBM." *Quarterly Journal of Medicine* 91:371 - 374.

Chobanian, A. V., G. L. Bakris, H. R. Black, W. C. Cushman, L. A. Green, J. L. Izzo, D. W. Jones, B. J. Materson, S. Oparil, J. T. Wright, E. J. Roccella, and Prog Natl High Blood Pressure Educ. 2003. "The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure - The JNC 7 Report." *Jama-Journal of the American Medical Association* 289 (19):2560-2572. doi: 10.1001/jama.289.19.2560.

CIHR. 2015. "Knowledge Translation at CIHR." Accessed 29 January 2016. http://www.cihr-irsc.gc.ca/e/29418.html.

Clarke, Brendan. 2011. "Causation and melanoma classification." *Theoretical Medicine and Bioethics* 32 (1):19-32.

Clarke, Brendan, Donald Gillies, Phyllis McKay Illari, Federica Russo, and Jon Williamson. 2014. "Mechanisms and the Evidence Hierarchy." *Topoi* 33 (2):339-360.

Contopoulos-Ioannidis, D.G., E. Ntzani, and J.P. Ioannidis. 2003. "Translation of highly promising basic science research into clinical applications." *The American Journal of Medicine* 114 (6):477-484.

Cooper, Rachel. 2002. "Disease." *Studies in History and Philosophy of Biological and Biomedical Sciences* 33:263-282.

Cowan, C. D., and J. Wittes. 1994. "Intercept studies, clinical trials, and cluster experiments: to whom can we extrapolate." *Controlled Clinical Trials* 15 (1):24-29. doi: 10.1016/0197-2456(94)90025-6.

Cox, E., Borio, L., & Temple, R. (2014). Evaluating Ebola therapies--the case for RCTs. *N Engl J Med, 371*, 2350-2351.

Cox, J., A. Campbell, and W. Fulford. 2007. *Medicine of the Person: Faith, Science and Values in Healthcare Provision*. London: Jessica Kingsley.

Craver, Carl. 2007. *Explaining The Brain*. Oxford: Clarendon Press.

Craver, Carl F., and Lindley Darden. 2013. *In Search of Mechanisms: Discoveries Across the Life Sciences*. Chicago: University of Chicago Press.

D'Agostino, Ralph B., Ramachandran S. Vasan, Michael J. Pencina, Philip A. Wolf, Mark Cobain, Joseph M. Massaro, and William B. Kannel. 2008. "General cardiovascular risk profile for use in primary care - The Framingham Heart Study." *Circulation* 117 (6):743-753.

Dans, A. L., L. F. Dans, G. H. Guyatt, and S. Richardson. 1998. "Users' guides to the medical literature. XIV. How to decide on the applicability of clinical trial results to your patient." *Journal of the American Medical Association* 279 (7):545-549.

Darden, L. 2008. "Thinking Again about Biological Mechanisms." *Philosophy of Science* 75 (5):958-969.

Dawes, Martin. 2010. "Co-morbidity: we need a guideline for each patient not a guideline for each disease." *Family Practice* 27 (1):1-2. doi: 10.1093/fampra/cmp106.

Dawid, Philip. 2015. "On individual risk." *Synthese*. doi: 10.1007/s11229-015-0953-4.

De Vreese, Leen, Erik Weber, and Jeroen Bouwel. 2010. "Explanatory pluralism in the medical sciences: Theory and practice." *Theoretical Medicine and Bioethics* 31 (5):371-390.

Deeks, J. J. 2002. "Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes." *Statistics in Medicine* 21 (11):1575-1600.

Dekkers, O. M., E. von Elm, A. Algra, J. A. Romijn, and J. P. Vandenbroucke. 2010. "How to assess the external validity of therapeutic trials: a conceptual approach." *International Journal of Epidemiology* 39 (1):89-94.

Diederichs, C., K. Berger, and D. B. Bartels. 2011. "The measurement of multiple chronic diseases--a systematic review on existing multimorbidity indices." *J Gerontol A Biol Sci Med Sci* 66 (3):301-11.

Djulbegovic, Benjamin, Gordon H. Guyatt, and Richard E. Ashcroft. 2009. "Epistemologic Inquiries in Evidence-Based Medicine." *Cancer Control* 16 (2):158-168.

Djulbegovic, Benjamin, Iztok Hozo, and Sander Greenland, eds. 2011. *Uncertainty in Clinical Medicine*. Edited by D. M. Gabbay, P. Thagard and J. Woods, *Handbook of the Philosophy of Science: Volume 16: Philosophy of Medicine*. Amsterdam: Elsevier.

Doll, Richard. Richard Peto. "Randomised controlled trials and retrospective controls." *British Medical Journal* 280:44.

Douglas, H. 2000. "Inductive risk and values in science." *Philosophy of Science* 67 (4):559-579.

Duffin, Jacalyn. 2010. *History of Medicine: A Scandalously Short Introduction*. Toronto: University of Toronto Press.

Dupre, J. 1984. "Probabilistic Causality Emancipated." *Midwest Studies in Philosophy* 9:169-175.

EBM_Working_Group. 1992. "Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine." *Journal of the American Medical Association* 268 (17):2420-2425.

Eells, E. 1991. *Probabilistic Causality*. Cambridge: Cambridge University Press.

Engel, George L. 1977. "The Need for a New Medical Model: A Challenge for Biomedicine." *Science* 196:129 - 136.

Engel, George. L. 1981. "The clinical application of the biopsychosocial model." *Journal of Medicine and Philosophy* 6 (2):101-123.

Engel, George. L. 2012. "The need for a new medical model: a challenge for biomedicine." *Psychodyn Psychiatry* 40 (3):377-96.

Engelhardt, Tristam. 1986. *The Foundations of Bioethics*. New York: Oxford University Press.

Ereshefsky, Marc. 2009. "Defining 'health' and 'disease'." *Studies in history and philosophy of biological and biomedical sciences* 40 (3):221-7. doi: 10.1016/j.shpsc.2009.06.005.

Estol, C. J. 2011. "Atherosclerosis: the XXIst century epidemic: a meeting at the Vatican." *Stroke* 42 (12):3338-9. doi: 10.1161/strokeaha.111.640300.

Evans, Alfred S. 1993. *Causation and Disease: A Chronological Journey*. New York: Plenum.

Fauci, Anthony S., and H. Clifford Lane. 2012. "Chapter 189. Human Immunodeficiency Virus Disease: AIDS and Related Disorders." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Federoff, Howard J., and Lawrence O. Gostin. 2009. "Evolving from Reductionism to Holism." *Journal of the American Medical Association* 302:994 - 996.

Feinstein, A. 1970. "Pre-therapeutic classification of co-morbidity in chronic disease." *Journal of Chronic Diseases* 23 (7):455-468. doi: 10.1016/0021-9681(70)90054-8.

Feinstein, A. R., and R. I. Horwitz. 1997. "Problems in the "evidence" of "evidence-based medicine"." *American Journal of Medicine* 103 (6):529-535. doi: 10.1016/s0002-9343(97)00244-1.

Felson, David T. 2012. "Chapter 332. Osteoarthritis." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Fine, P. E. M. 1995. "Variation in protection by BCG: implications of and for heterologous immunity." *Lancet* 346 (8986):1339-1345.

Fleming, T. R., and S. S. Ellenberg. 2016. "Evaluating interventions for Ebola: The need for randomized trials." *Clin Trials*. doi: 10.1177/1740774515616944.

Flores, L. 2015. "Therapeutic inferences for individual patients." *Journal of Evaluation in Clinical Practice* 21 (3):440-447. doi: 10.1111/jep.12293.

Fortin, M., G. Bravo, C. Hudon, A. Vanasse, and L. Lapointe. 2005. "Prevalence of multimorbidity among adults seen in family practice." *Annals of Family Medicine* 3 (3):223-228.

Fortin, Martin, Lise Lapointe, Catherine Hudon, Alain Vanasse, Antoine L. Ntetu, and Danielle Maltais. 2004. "Multimorbidity and quality of life in primary care: a systematic review." *Health and Quality of Life Outcomes* 2:51-63. doi: 10.1186/1477-7525-2-51.

Foss, L. 1989. "The challenge to biomedicine - A foundations perspective." *Journal of Medicine and Philosophy* 14 (2):165-191.

Fournier, Jay C., Robert J. DeRubeis, Steven D. Hollon, Sona Dimidjian, Jay D. Amsterdam, Richard C. Shelton, and Jan Fawcett. 2010. "Antidepressant Drug Effects and Depression Severity A Patient-Level Meta-analysis." *Jama-Journal of the American Medical Association* 303 (1):47-53. doi: 10.1001/jama.2009.1943.

Fuller, Jonathan. 2013a. "Rhetoric and argumentation: how clinical practice guidelines think." *Journal of Evaluation in Clinical Practice* 19 (3):433-441.

Fuller, Jonathan. 2013b. "Rationality and the generalization of randomized controlled trial evidence." *Journal of Evaluation in Clinical Practice* 19 (4):644-647.

Fuller, Jonathan, Alex Broadbent, and Luis J. Flores. 2015. "Prediction in epidemiology and medicine." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 54:45-48. doi: http://dx.doi.org/10.1016/j.shpsc.2015.07.001.

Fuller, Jonathan, L. J. Flores, R. E. Upshur, and M. J. Goldenberg. 2014. "Renaissance or reformation for evidence based medicine?" *Bmj* 349:g4902. doi: 10.1136/bmj.g4902.

Fuller, Jonathan, Ashwin Sankar, and Ross E.G. Upshur. 2013. "Concepts in Evidence-Informed Medical Practice." In *The Essentials of Clinical Examination Handbook 7th Edition*, edited by Justin Hall, Katrina Piggott, Miliana Vojvodic and Kirill Zaslavsky, 571-583. Toronto: Thieme.

Fuller, Jonathan, and Ross E. G. Upshur. 2011. "Medication regimen complexity and the care of the chronically ill patient." *International Journal of Person Centered Medicine* 1 (4):719-725.

Fuller, Jonathan, and Ross E. G. Upshur. 2014. "Logos, ethos and pathos in balance –the care of the patient and the soul of the clinic: person-centered medicine as an emergent model of modern clinical practice. ." *European Journal for Person Centered Healthcare* 2 (1):22-29.

Furukawa, T. A., G. H. Guyatt, and L. E. Griffith. 2002. "Can we individualize the "number needed to treat"? An empirical study of summary effect measures in meta-analyses." *International Journal of Epidemiology* 31 (1):72-76.

Gadsboll, N., P. F. Hoilundcarlsen, G. G. Nielsen, J. Berning, N. E. Brunn, P. Stage, E. Hein, J. Marving, H. Longborgjensen, and B. H. Jensen. 1989. "SYMPTOMS AND SIGNS OF HEART-FAILURE IN PATIENTS WITH MYOCARDIAL-INFARCTION - REPRODUCIBILITY AND RELATIONSHIP TO CHEST-X-RAY, RADIONUCLIDE VENTRICULOGRAPHY AND RIGHT HEART CATHETERIZATION." *European Heart Journal* 10 (11):1017-1028.

Gale, E. A. 2002. "The rise of childhood type 1 diabetes in the 20th century." *Diabetes* 51 (12):3353-61.

Gallacher, K., C. R. May, V. M. Montori, and F. S. Mair. 2011. "Understanding patients' experiences of treatment burden in chronic heart failure using normalization process theory." *The Annals of Family Medicine* 9 (3):235-243. doi: 10.1370/afm.1249.

Genest, Jacques, Ruth McPherson, Jiri Frohlich, Todd Anderson, Norm Campbell, Andre Carpentier, Patrick Couture, Robert Dufour, George Fodor, Gordon A. Francis, Steven Grover, Milan Gupta, Robert A. Hegele, David C. Lau, Lawrence Leiter, Gary F. Lewis, Eva Lonn, G. B. John Mancini, Dominic Ng, Glen J. Pearson, Allan Sniderman, James A. Stone, and Ehud Ur. 2009. "2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of cardiovascular disease in the adult-2009 recommendations." *Canadian Journal of Cardiology* 25 (10):567-579.

Giere, Ronald. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Giere, Ronald. 1999. *Science Without Laws*. Chicago: University of Chicago Press.

Gifford, Fred. 2011. "Philosophy of Medicine: Introduction." In *Handbook of the Philosophy of Science. Volume 16: Philosophy of Medicine*, edited by D. M. Gabbay, P. Thagard and J. Woods, 1-12. Amsterdam: Elsevier BV.

Gijsen, R., N. Hoeymans, F. G. Schellevis, D. Ruwaard, W. A. Satariano, and G. A. M. van den Bos. 2001. "Causes and consequences of comorbidity: A review." *Journal of Clinical Epidemiology* 54 (7):661-674. doi: 10.1016/s0895-4356(00)00363-2.

Gillies, D. 2000. *Philosophical Theories of Probability*. London, New York: Routledge.

Gillies, Donald. 2005. "Hempelian and Kuhnian approaches in the philosophy of medicine: the Semmelweis case." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (1):159-81. doi: 10.1016/j.shpsc.2004.12.003.

Glasziou, P. P., and L. M. Irwig. 1995. "An evidence based approach to individualising treatment." *British Medical Journal* 311 (7016):1356-1359.

Glasziou, Paul, and David Mant. 2007. "Applying results to treatment decisions in primary care." In *Treating Individuals: From Randomized Trials to Personalised Medicine*, edited by Peter M. Rothwell, 83-95. The Lancet.

Glennan, Stuart S. 1996. "Mechanisms and the nature of causation." *Erkenntnis* 44:49-71.

Glennan, Stuart S. 2002. "Rethinking mechanistic explanation." *Philosophy of Science* 69 (3):S342-S353.

Goldenberg, Maya J. 2009. "Iconoclast or Creed?: Objectivism, Pragmatism, and the Hierarchy of Evidence." Perspectives in Biology and Medicine 52 (2):168-187.

Good, I. J. 1967. "On the Principle of Total Evidence." *British Journal for the Philosophy of Science* 17 (4):319-321.

Goodman, R. A., S. F. Posner, E. S. Huang, A. K. Parekh, and H. K. Koh. 2013. "Defining and Measuring Chronic Conditions: Imperatives for Research, Policy, Program, and Practice." *Preventing Chronic Disease* 10. doi: 10.5888/pcd10.120239.

Goodman, S. N. 1999. "Probability at the bedside: The knowing of chances or the chances of knowing?" *Annals of Internal Medicine* 130 (7):604-606.

Goulding, M. R., M. E. Rogers, and S. M. Smith. 2003. "Public health and aging: trends in aging - United States and worldwide." *Journal of the American Medical Association* 289 (11):1371-1373.

Graham, Daniel. 1991. "Aristotle." In *Handbook of Metaphysics and Ontology*, edited by Hans Burkhardt and Barry Smith, 50-54. Munich, Philadelphia: Philosophia Verlag.

Greenfield, S., J. Billimek, F. Pellegrini, M. Franciosi, G. De Berardis, A. Nicolucci, and S. H. Kaplan. 2009. "Comorbidity affects the relationship between glycemic control and cardiovascular outcomes in diabetes: a cohort study." *Ann Intern Med* 151 (12):854-60. doi: 10.7326/0003-4819-151-12-200912150-00005.

Greenhalgh, T., J. Howick, and N. Maskrey. 2014. "Evidence based medicine: a movement in crisis?" *BMJ* 348:g3725.

Greenland, S., and J. M. Robins. 1986. "Identifiability, exchangeability, and epidemiological confounding." *International Journal of Epidemiology* 15 (3):413-9.

Gross, C. P., R. Mallory, A. Heiat, and H. M. Krumholz. 2002. "Reporting the recruitment process in clinical trials: who are these patients and how did they get there?" *Ann Intern Med* 137 (1):10-6.

Grumbach, Kevin. 2003. "Chronic illness, comorbidities, and the need for medical generalism." *Annals of Family Medicine* 1 (1):4-7. doi: 10.1370/afm.47.

Guthrie, Bruce, Katherine Payne, Phil Alderson, Marion E. T. McMurdo, and Stewart W. Mercer. 2012. "Adapting clinical guidelines to take account of multimorbidity." *British Medical Journal* 345:e6341. doi: e634110.1136/bmj.e6341.

Guyatt, G., D. Rennie, M. O. Meade, and D. J. Cook. 2008. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice*. 2nd ed. New York: McGraw-Hill Medical.

Hacking, Ian. 2001. *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.

Haggerty, Jeannie L. 2012. "Ordering the chaos for patients with multimorbidity." *British Medical Journal* 345. doi: e591510.1136/bmj.e5915.

Hankey, Graeme J. 2007. "Are n-of-1 trials of any practical value to clinicians and researchers?" In *Treating Individuals: From Randomized Trials to Personalised Medicine*, edited by Peter M. Rothwell, 231-246. The Lancet.

Hansson, G. K. 2005. "Mechanisms of disease - Inflammation, atherosclerosis, and coronary artery disease." *New England Journal of Medicine* 352 (16):1685-1695. doi: 10.1056/NEJMra043430.

Hartzband, P., and J. Groopman. 2016. "Medical Taylorism." *N Engl J Med* 374 (2):106-8. doi: 10.1056/NEJMp1512402.

Hauben, Manfred, and Jeffrey K. Aronson. 2006. "Paradoxical reactions: under-recognized adverse effects of drugs." *Drug Safety* 29:970.

Hay, M., D. W. Thomas, J. L. Craighead, C. Economides, and J. Rosenthal. 2014. "Clinical development success rates for investigational drugs." *Nat Biotechnol* 32 (1):40-51.

Hempel, Carl G. 1962. "Deductive-Nomological vs. Statistical Explanation." In *Minnesota Studies in the Philosophy of Science*, edited by H. Feigl, 98-169. Minneapolis.

Hempel, Carl G. 1965. "Science and human values." In *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, 81-96. New York: The Free Press.

Hempel, Carl G., and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15 (2):135-175.

Hennekens, Charles H., and Julie E. Buring. 1987. *Epidemiology in Medicine*. Boston: Little, Brown.

Hesslow, Germund. 1976. "Two notes on the probabilistic approach to causality." *Philosophy of Science* 43:290-292.

Hewa, S., and R. W. Hetherington. 1995. "Specialists without spirit - limitations of the mechanistic biomedical model." *Theoretical Medicine* 16 (2):129-139. doi: 10.1007/bf00998540.

Hill, A. B. 1965. "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine* 58:295 - 300.

Hill, A. B., and I. D. Hill. 1991. *Bradford Hill's Principles of Medical Statistics*. 12 ed. London: Edward Arnold.

Hlatky, Mark A., Derek B. Boothroyd, Dena M. Bravata, Eric Boersma, Jean Booth, Maria M. Brooks, Didier Carrie, Tim C. Clayton, Nicolas Danchin, Marcus Flather, Christian W. Hamm, Whady A. Hueb, Jan Kaehler, Sheryl F. Kelsey, Spencer B. King, Andrzej S. Kosinski, Neuza Lopes, Kathryn M. McDonald, Alfredo Rodriguez, Patrick Serruys, Ulrich Sigwart, Rodney H. Stables, Douglas K. Owens, and Stuart J. Pocock. 2009. "Coronary artery bypass surgery compared with percutaneous coronary interventions for multivessel disease: a collaborative analysis of individual patient data from ten randomised trials." *Lancet* 373 (9670):1190-1197. doi: 10.1016/s0140-6736(09)60552-3.

Hoffman, C., D. Rice, and H. Y. Sung. 1996. "Persons with chronic conditions - Their prevalence and costs." *Journal of the American Medical Association* 276 (18):1473-1479. doi: 10.1001/jama.276.18.1473.

Hofmann, B. 2001. "Complexity of the concept of disease as shown through rival theoretical frameworks." *Theoretical Medicine and Bioethics* 22 (3):211-236. doi: 10.1023/a:1011416302494.

Horsburgh, C. Robert. 2015. Epidemiology of tuberculosis. *UpToDate*.

Horton, R. 2000. "Common sense and figures: the rhetoric of validity in medicine. Bradford Hill Memorial Lecture 1999." Statistics in Medicine 19 (23):3149-3164. doi: 10.1002/1097-0258(20001215)19:23<3149::aid-sim617>3.0.co;2-e.

Horwitz, R. I., B. H. Singer, R. W. Makuch, and C. M. Viscoli. 1996. "Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation." *J Clin Epidemiol* 49 (4):395-400.

Howick, Jeremy. 2011a. "Exposing the Vanities-and a Qualified Defense-of Mechanistic Reasoning in Health Care Decision Making." *Philosophy of Science* 78 (5):926-940.

Howick, Jeremy. 2011b. *The Philosophy of Evidence-Based Medicine*. Chichester, UK: Wiley-Blackwell.

Howick, Jeremy, Paul Glasziou, and Jeffrey K. Aronson. 2013. "Problems with Using Mechanisms to Solve the Problem of Extrapolation." *Theoretical Medicine and Bioethics* 34:275 - 291.

Howick, Jeremy, Ashley Graham Kennedy, and Alexander Mebius. 2015. Philosophy of Evidence-Based Medicine. *Oxford Bibliographies*. doi:10.1093/OBO/9780195396577-0253.

Howson, Colin, and Peter Urbach. 2006. *Scientific Reasoning: the Bayesian Approach*. 3 ed. Chicago, La Salle: Open Court Publishing Company.

Hucklenbroich, P. 2014. ""Disease entity" as the key theoretical concept of medicine." *J Med Philos* 39 (6):609-33.

Hume, David. 1777. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Clarendon Press.

Illari, Phyllis M. 2013. "Mechanistic Explanation: Integrating the Ontic and Epistemic." *Erkenntnis* 78:237-255.

Illari, Phyllis McKay, and Jon Williamson. 2012. "What is a mechanism? Thinking about mechanisms across the sciences." *European Journal for Philosophy of Science* 2:119-135.

IOM. 2012. *Living Well with Chronic Illness*: National Academies Press.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. New York: Oxford University Press.

Jadad, A.R., A Cabrera, F Martos, R Smith, and R.F. Lyons. 2010. *When People Live with Multiple Chronic Diseases: A Collaborative Approach to an Emerging Global Challenge*. Granada: Andalusian School of Public Health.

Jansen, Ludger. 2007. "Tendencies and other realizables in medical information sciences." *Monist* 90 (4):534-554.

Johnston, B. C., P. Alonso-Coello, J. O. Friedrich, R. A. Mustafa, K. A. Tikkinen, I. Neumann, P. O. Vandvik, E. A. Akl, B. R. da Costa, N. K. Adhikari, G. M. Dalmau, E. Kosunen, J. Mustonen, M. W. Crawford, L. Thabane, and G. H. Guyatt. 2016. "Do clinicians understand the size of treatment effects? A randomized survey across 8 countries." *Cmaj* 188 (1):25-32. doi: 10.1503/cmaj.150430.

Jutel, Annemarie. 2011. "Classification, disease, and diagnosis." *Perspectives in Biology and Medicine* 54 (2):189-205.

Juurlink, D. N., M. M. Mamdani, D. S. Lee, A. Kopp, P. C. Austin, A. Laupacis, and D. A. Redelmeier. 2004. "Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study." *N Engl J Med* 351 (6):543-51.

Keynes, J. M. 1921. *A Treatise on Probability*: Macmillan.

King, Lester S. 1954. "What is disease?" *Philosophy of Science* 21 (3):193-203.

King, Lester S. 1982. *Medical Thinking: A Historical Preface*. Princeton: Princeton University Press.

Kingma, Elselijn. 2007. "What Is It to Be Healthy?" *Analysis* 67:128 - 133.

Koné Pefoyo, A. J., S. E. Bronskill, A. Gruneir, A. Calzavara, K. Thavorn, Y. Petrosyan, C. J. Maxwell, Y. Bai, and W. P. Wodchis. 2015. "The increasing burden and complexity of multimorbidity." *BMC Public Health* 15:415. doi: 10.1186/s12889-015-1733-2.

Kontos, N. 2011. "Perspective: biomedicine--menace or straw man? Reexamining the biopsychosocial argument." *Acad Med* 86 (4):509-15. doi: 10.1097/ACM.0b013e31820e0d16.

Kosinski, Neuza Lopes, Kathryn M. McDonald, Alfredo Rodriguez, Patrick Serruys, Ulrich Sigwart, Rodney H. Stables, Douglas K. Owens, and Stuart J. Pocock. 2009. "Coronary artery bypass surgery compared with percutaneous coronary interventions for multivessel disease: a collaborative analysis of individual patient data from ten randomised trials." *Lancet* 373 (9670):1190-1197.

Kotchen, Theodore A. 2012. "Chapter 247. Hypertensive Vascular Disease." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Kravitz, R. L., N. Duan, and J. Braslow. 2004. "Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages." *Milbank Quarterly* 82 (4):661-87.

Krieger, Nancy. 1994. "Epidemiology and the Web of Causation: Has Anyone Seen the Spider?" *Social Science & Medicine* 39:887 - 903.

La Caze A. 2009. "Evidence-based medicine must be. . ." *Journal of Medicine and Philosophy* 34:509 - 527.

La Caze, Adam. 2013. "Why Randomized Interventional Studies." *Journal of Medicine and Philosophy* 38 (4):352-368. doi: 10.1093/jmp/jht028.

LaCaze, Adam, Benjamin Djulbegovic, and Stephen Senn. 2012. "What does randomisation achieve?" *Evidence-Based Medicine* 17 (1):1-2.

Lange, Marc. 2007. "The end of diseases." *Philosophical Topics* 35:265-292.

Leder, D. 1984. "Medicine and paradigms of embodiment." *Journal of Medicine and Philosophy* 9 (1):29-43.

Lemoine, Mael. 2013. "Defining disease beyond conceptual analysis: an analysis of conceptual analysis in philosophy of medicine." *Theoretical Medicine and Bioethics* 34 (4):309-325. doi: 10.1007/s11017-013-9261-5.

Lemoine, Mael. 2015. "The naturalization of the concept of disease." In *Classification, Disease and Evidence: New Essays in the Philosophy of Medicine*, edited by Philippe Huneman, Gerard Lambert and Marc Silberstein. Dordrecht: Springer Netherlands.

Levenstein, J. H., E. C. McCracken, I. R. McWhinney, M. A. Stewart, and J. B. Brown. 1986. "The patient-centred clinical method. 1. A model for the doctor-patient interaction in family medicine." *Family Practice* 3 (1):24-30.

Libby, Peter. 2012. "Chapter 241. The Pathogenesis, Prevention, and Treatment of Atherosclerosis." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Lindsay, Robert, and Felicia Cosman. 2012. "Chapter 354. Osteoporosis." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Longo, Dan L., ed. 2012. *Harrison's Principles of Internal Medicine*. 18 ed. New York: McGraw-Hill.

Lubsen, J., and J. G. Tijssen. 1989. "Large trials with simple protocols: indications and contraindications." *Control Clin Trials* 10 (4 Suppl):151s-160s.

Machamer, P., L. Darden, and C. F. Craver. 2000. "Thinking about mechanisms." *Philosophy of Science* 67 (1):1-25.

Mackie, J. L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2 (4):245-264.

Mackie, J. L. 1977. "Dispositions, grounds and causes." *Synthese* 34:361-370.

Mackie, J. L. 1980. *The Cement of the Universe: A Study of Causation*. Oxford Scholarship Online.

Macmahon, B., and T.F. Pugh. 1970. *Epidemiology, Principles and Methods*. Boston: Little Brown.

Malterud, K. 1995. "The legitimacy of clinical knowledge: towards a medical epistemology embracing the art of medicine." *Theor Med* 16 (2):183-98.

Mangin, Dee, Iona Heath, and Marc Jamoulle. 2012. "Beyond diagnosis: Rising to the multimorbidity challenge." *British Medical Journal* 344. doi: e352610.1136/bmj.e3526.

Mann, Douglas L., and Murali Chakinala. 2012. "Chapter 234. Heart Failure and Cor Pulmonale." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Marcellesi, A. 2015. "External Validity: Is There Still a Problem?" *Philosophy of Science* 82 (5):1308-1317. doi: 10.1086/684084.

Marcum, James A. 2008a. "Reflections on humanizing biomedicine." *Perspectives in Biology and Medicine* 51 (3):392-405.

Marcum, James A. 2008b. *An Introductory Philosophy of Medicine: Humanizing Modern Medicine*. Vol. 99. New York: Springer.

Marincola, F. M. 2003. "Translational medicine: a two-way road." *Journal of Translational Medicine* 1 (1):1.

Mason, C., and P. Dunnill. 2008. "A brief definition of regenerative medicine." *Regen Med* 3 (1):1-5. doi: 10.2217/17460751.3.1.1.

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *Plos Medicine, 3*, e442.

Mavromoustakos, T., S. Durdagi, C. Koukoulitsa, M. Simcic, M. G. Papadopoulos, M. Hodoscek, and S. Golic Grdadolnik. 2011. "Strategies in the Rational Drug Design." *Current Medicinal Chemistry* 18 (17):2517-2530.

May, Carl, Victor M. Montori, and Frances S. Mair. 2009. "We need minimally disruptive medicine." *British Medical Journal* 339. doi: b280310.1136/bmj.b2803.

McAlister, F. A. 2002. "Commentary: Relative treatment effects are consistent across the spectrum of underlying risks ... usually." *International Journal of Epidemiology* 31 (1):76-77.

McCormick, J. 1988. "The multifactorial etiology of coronary heart disease: a dangerous delusion." *Perspectives in Biology and Medicine* 32 (1):103-108.

McKenna, M, and J Collins. 2010. "Current issues and challenges in chronic disease control." In *Chronic Disease Epidemiology and Control.*, edited by PL Remington, RC Brownson and MV Wegner, 1-24. Washinton (DC): American Public Health Association.

McKeown, R. E. 2009. "The Epidemiologic Transition: Changing Patterns of Mortality and Population Dynamics." *Am J Lifestyle Med* 3 (1 Suppl):19s-26s. doi: 10.1177/1559827609335350.

McMurray, J. J., S. Adamopoulos, S. D. Anker, A. Auricchio, M. Bohm, K. Dickstein, V. Falk, G. Filippatos, C. Fonseca, M. A. Gomez-Sanchez, T. Jaarsma, L. Kober, G. Y. Lip, A. P. Maggioni, A. Parkhomenko, B. M. Pieske, B. A. Popescu, P. K. Ronnevik, F. H. Rutten, J. Schwitter, P. Seferovic, J. Stepinska, P. T. Trindade, A. A. Voors, F. Zannad, and A. Zeiher.

2012. "ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC." Eur Heart J 33 (14):1787-847. doi: 10.1093/eurheartj/ehs104.

Mercer, Stewart W., Bruce Guthrie, John Furler, Graham C. M. Watt, and Julian Tudor Hart. 2012. "Multimorbidity and the inverse care law in primary care." *British Medical Journal* 344.

Mercuri, M., and A. Gafni. 2011. "Medical practice variations: what the literature tells us (or does not) about what are warranted and unwarranted variations." *J Eval Clin Pract* 17 (4):671-7. doi: 10.1111/j.1365-2753.2011.01689.x.

Miles, Andrew. 2009. "On a Medicine of the Whole Person: Away from scientistic reductionism and towards the embrace of the complex in clinical practice." *Journal of Evaluation in Clinical Practice* 15 (6):941-949. doi: 10.1111/j.1365-2753.2009.01354.x.

Miles, A., and J. E. Mezzich. 2011. "The care of the patient and the soul of the clinic: person-centered medicine as an emergent model of modern clinical practice." *International Journal of Person Centered Medicine* 1 (2):207-222.

Mill, John Stuart. 1882. *A System of Logic, Ratiocinative and Inductive*. New York: Harper and Brothers.

Moayedi, Y., and J. Kobulnik. 2015. "Chronic heart failure with reduced ejection fraction." *Cmaj* 187 (7):518. doi: 10.1503/cmaj.140430.

Moghaddam-Taaheri, S. 2011. "Understanding pathology in the context of physiological mechanisms: the practicality of a broken-normal view." *Biology & Philosophy* 26 (4):603-611.

Moher, D., S. Hopewell, K. F. Schulz, V. Montori, P. C. Gotzsche, P. J. Devereaux, D. Elbourne, M. Egger, and D. G. Altman. 2010. "CONSORT 2010 explanation and elaboration:

updated guidelines for reporting parallel group randomised trials." *BMJ* 340:c869. doi: 10.1136/bmj.c869.

Mol, Annemarie. 2002. *The Body Multiple*. Durham: Duke University Press.

Morabia, Alfredo. 2011. "History of the modern epidemiological concept of confounding." *Journal of Epidemiology & Community Health* 65:297-300.

Morabia, Alfredo. 2013. "Hume, Mill, Hill, and the sui generis epidemiologic approach to causal inference." *American Journal of Epidemiology* 178 (10):1526–1532.

Moynihan, Ray, and Alan Cassels. 2005. *Selling Sickness: How the World's Biggest Pharmaceutical Patients are Turning us all into Patients*. Vancouver: Greystone Books.

Mumford, Stephen. 1998. *Dispositions*. Oxford: Oxford University Press.

Nagel, Jennifer. 2014. *Knowledge: A Very Short Introduction*. Oxford: Oxford University Press.

Nasmith L., Ballem P., Baxter R., Bergman H., Colin-Thomé D., Herbert C., Keating N., Lessard R., Lyons R., McMurchy D., Ratner P., Rosenbaum P., Tamblyn R., Wagner E., Zimmerman B. 2010. Transforming Care for Canadians with Chronic Health Conditions: Put People First, Expect the Best, Manage for Results. Ottawa: Canadian Academy of Health Sciences.

Naylor, C. David. 1995. "Grey Zones of Clinical Practice: Some Limits to Evidence-Based Medicine." *The Lancet* 345:840 - 842.

Nissen, S. E., and K. Wolski. 2007. "Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes." *N Engl J Med* 356 (24):2457-71.

Nordenfelt, Lennart. 1995. *On the Nature of Health: An Action-Theory Approach*. 2 ed. Vol. 26. Dordrecht, The Netherlands: Kluwer.

Padovan, Richard. 1999. *Proportion*. London: Taylor & Francis.

Pang, and I. Harvey. 2010. "Dissemination and publication of research findings: an updated review of related biases." *Health Technology Assessment* 14 (8).

Papaioannou, A., S. Morin, A. M. Cheung, S. Atkinson, J. P. Brown, S. Feldman, D. A. Hanley, A. Hodsman, S. A. Jamal, S. M. Kaiser, B. Kvern, K. Siminoski, and W. D. Leslie. 2010. "2010 clinical practice guidelines for the diagnosis and management of osteoporosis in Canada: summary." *Canadian Medical Association Journal* 182 (17):1864-1873. doi: 10.1503/cmaj.100771.

Papineau, David. 1994. "The virtues of randomization." *British Journal for the Philosophy of Science* 45 (2):437-450.

Papineau, David. 2014. "The Poverty of Conceptual Analysis." In *Philosophical Methodology: the Armchair or the Laboratory?*, edited by Matthew C. Haug. Abingdon, Oxon: Routledge.

Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference*. New York: Cambridge University Press.

Peltzman, Sam. 1973. "An evaluation of consumer protection legislation: the 1962 drug amendments." *Journal of Political Economy* 81 (5):1049-1091.

Peter, H. Schwartz. 2008. "Risk and Disease." *Perspectives in Biology and Medicine* 51 (3):320-334. doi: 10.1353/pbm.0.0027.

Petticrew, M., and I. Chalmers. 2011. "Use of research evidence in practice." *Lancet* 378 (9804):1696-1696.

Piette, J. D., and E. A. Kerr. 2006. "The impact of comorbid chronic conditions on diabetes care." *Diabetes Care* 29 (3):725-31.

Pignon, Jean-Pierre, Aurelie le Maitre, Emilie Maillard, and Jean Bourhis. 2009. "Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients." *Radiotherapy and Oncology* 92 (1):4-14.

Pizzo, P. 2002. Letter from the Dean. In *Stanford Medical Magazine*.

Post, P. N., H. de Beer, and G. H. Guyatt. 2013. "How to generalize efficacy results of randomized trials: recommendations based on a systematic review of possible approaches." *Journal of Evaluation in Clinical Practice* 19 (4):638-643.

Potter, P. 1988. *Short Handbook of Hippocratic Medicine*. Quebec: Éditions du Sphinx.

Powers, Alvin C. 2012. "Chapter 344. Diabetes Mellitus." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Prior, E., R. Pargetter, and F. Jackson. 1982. "Three theses about dispositions." *American Philosophical Quarterly* 19:251-257.

Psaty, B. N., T. Lumley, C. D. Furberg, G. Schellenbaum, M. Pahor, M. H. Alderman, and N. S. Weiss. 2003. "Health outcomes associated with various antihypertensive therapies used as first-line agents - A network meta-analysis." *Jama-Journal of the American Medical Association* 289 (19):2534-2544.

Qato, Dima M., G. Caleb Alexander, Rena M. Conti, Michael Johnson, Phil Schumm, and Stacy Tessler Lindau. 2008. "Use of prescription and over-the-counter medications and dietary supplements among older adults in the United States." *Journal of the American Medical Association* 300 (24):2867-2878.

Rabi, Doreen M., Stella S. Daskalopoulou, Raj S. Padwal, Nadia A. Khan, Steven A. Grover, Daniel G. Hackam, Martin G. Myers, Donald W. McKay, Robert R. Quinn, Brenda R. Hemmelgarn, Lyne Cloutier, Peter Bolli, Michael D. Hill, Thomas Wilson, Brian Penner, Ellen Burgess, Maxime Lamarre-Cliché, Donna McLean, Ernesto L. Schiffrin, George Honos, Karen Mann, Guy Tremblay, Alain Milot, Arun Chockalingam, Simon W. Rabkin, Martin Dawes, Rhian M. Touyz, Kevin D. Burns, Marcel Ruzicka, Norman R. C. Campbell, Michel Vallée, G. V. Ramesh Prasad, Marcel Lebel, Tavis S. Campbell, M. Patrice Lindsay, Robert J. Herman, Pierre Larochelle, Ross D. Feldman, J. Malcolm O. Arnold, Gordon W. Moe, Jonathan G. Howlett, Luc Trudeau, Simon L. Bacon, Robert J. Petrella, Richard Lewanczuk,

James A. Stone, Denis Drouin, Jean-Martin Boulanger, Mukul Sharma, Pavel Hamet, George Fodor, George K. Dresser, S. George Carruthers, George Pylypchuk, Richard E. Gilbert, Lawrence A. Leiter, Charlotte Jones, Richard I. Ogilvie, Vincent Woo, Philip A. McFarlane, Robert A. Hegele, Luc Poirier, and Sheldon W. Tobe. 2011. "The 2011 Canadian Hypertension Education Program Recommendations for the Management of Hypertension: Blood Pressure Measurement, Diagnosis, Assessment of Risk, and Therapy." *Canadian Journal of Cardiology* 27 (4):415-433.e2. doi: 10.1016/j.cjca.2011.03.015.

Ramsey, F. P. 1990. "Weight or the Value of Knowledge." *British Journal for the Philosophy of Science* 41 (1):1-4. doi: 10.1093/bjps/41.1.1.

Rather, L. J. 1959. "Towards a philosophical study of the idea of disease." In *The Historical Development of Physiological Thought*, edited by Brooks and Cranefield, 353-373. New York: Hafner.

Raviglione, Mario C., and Richard J. O'Brien. 2012. "Chapter 165: Tuberculosis." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Rawlins, Michael. 2008. "De testimonio: on the evidence for decisions about the use of therapeutic interventions." *Lancet* 372 (9656):2152-2161.

Reilly Jr., John J., Edwin K. Silverman, and Steven D. Shapiro. 2012. "Chapter 260. Chronic Obstructive Pulmonary Disease." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Reznek, Lawrie. 1987. *The Nature of Disease*. London: Routledge & Kegan Paul.

Ring, A., C. F. Dowrick, G. M. Humphris, J. Davies, and P. Salmon. 2005. "The somatising effect of clinical consultation: what patients and doctors say and do not say when patients present medically unexplained physical symptoms." *Soc Sci Med* 61 (7):1505-15

Rizzi, Dominick A., and Stig Andur Pedersen. 1992. "Causality in Medicine: Towards a Theory and Terminology." *Theoretical Medicine* 13:233 - 254.

Ross, R. 1999. "Mechanisms of disease - Atherosclerosis - An inflammatory disease." *New England Journal of Medicine* 340 (2):115-126.

Rothman, Kenneth J. 1976. "Causes." *American Journal of Epidemiology* 104 (6):587-592.

Rothman, Kenneth J., and Sander Greenland. 2005. "Causation and causal inference in epidemiology." *American Journal of Public Health* 95 (S1):S144-S150.

Rothwell, P. M. 1995. "Can overall results of clinical trials be applied to all patients?" *Lancet* 345 (8965):1616-9.

Rothwell, P. M. 2005. "External validity of randomised controlled trials: "to whom do the results of this trial apply?"." *Lancet* 365 (9453):82-93.

Rothwell, Peter M. 2007. "Preface." In *Treating Individuals: From Randomized Trials to Personalised Medicine*, edited by Peter M Rothwell, ix-xii. The Lancet.

Rubins, H. B. 1994. "From clinical trials to clinical practice: generalizing from participant to patient." *Controlled Clinical Trials* 15 (1):7-10. doi: 10.1016/0197-2456(94)90022-1.

Russo, Federica, and Jon Williamson. 2007. "Interpreting Causality in the Health Sciences." *International Studies in the Philosophy of Science* 21:157 - 170.

Sackett, David L., William M. C. Rosenberg, J. A. Muir Gray, R. Brian Haynes, and W. Scott Richardson. 1996. "Evidence Based Medicine: What It Is and What It Isn't." *British Medical Journal* 312:71 - 72.

Salisbury, Chris, Leigh Johnson, Sarah Purdy, Jose M. Valderas, and Alan A. Montgomery. 2011. "Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study." *British Journal of General Practice* 61 (582):e12-21.

Schaffner, Kenneth F. 2011. "Reduction in Biology and Medicine." In *Handbook of the Philosophy of Science. Volume 16: Philosophy of Medicine*, edited by D. M. Gabbay, P. Thagard and J. Woods, 137-157. Amsterdam: Elsevier BV.

Scheuermann, Richard H., Werner Ceusters, and Barry Smith. 2009. "Toward an ontological treatment of disease and diagnosis." *Summit on translational bioinformatics* 2009:116-20.

Schmoor, C., M. Olschewski, and M. Schumacher. 1996. "Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies." *Stat Med* 15 (3):263-71.

Schumacher, H. Ralph, and Lan X. Chen. 2012. "Chapter 333. Gout and Other Crystal-Associated Arthropathies." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Schwartz, Peter H. 2008. "Risk and Disease." *Perspectives in Biology and Medicine* 51 (3):320-334.

Schwartz, Peter H. 2009. "Disclosure and rationality: comparative risk information and decision-making about prevention." *Theor Med Bioeth* 30 (3):199-213. doi: 10.1007/s11017-009-9111-7.

Scott, Ian A., and Gordon H. Guyatt. 2010. "Cautionary tales in the interpretation of clinical studies involving older persons." *Archives of Internal Medicine* 170 (7):387-395.

Senn, Stephen. 2013. "Seven myths of randomisation in clinical trials." *Statistics in Medicine* 32:1439-1450.

Severinsen, M. 2001. "Principles behind definitions of diseases - A criticism of the principle of disease mechanism and the development of a pragmatic alternative." *Theoretical Medicine and Bioethics* 22 (4):319-336. doi: 10.1023/a:1011830602137.

Shaughnessy, A. F., D. C. Slawson, and L. Becker. 1998. "Clinical jazz: harmonizing clinical experience and evidence-based medicine." *J Fam Pract* 47 (6):425-8.

Sheldon, Trevor A., Gordon. H. Guyatt, and Andrew Haines. 1998. "Getting-research findings into practice - When to act on the evidence." *British Medical Journal* 317 (7151):139-142.

Simon, J. R. 2008. "Constructive realism and medicine - an approach to medical ontology." *Perspectives in Biology and Medicine* 51 (3):353-366.

Simon, Jeremy R. 2010. "Advertisement for the ontology of medicine." *Theoretical Medicine and Bioethics* 31 (5):333-346.

Simon, Jeremy R. 2011. "Medical Ontology." In *Handbook of the Philosophy of Science. Volume 16: Philosophy of Medicine*, edited by D. M. Gabbay, P. Thagard and J. Woods, 65-114. Amsterdam: Elsevier BV.

Skipper, Robert A., Jr., and Roberta L. Millstein. 2005. "Thinking about evolutionary mechanisms: natural selection." *Studies in history and philosophy of biological and biomedical sciences* 36 (2):327-47.

Smart, B. 2014. On the classification of diseases. *Theoretical Medicine and Bioethics* 35 (4):251-269.

Smith, Susan M., Atakelet Ferede, and Tom O'Dowd. 2008. "Multimorbidity in younger deprived patients: An exploratory study of research and service implications in general practice." *BMC Family Practice* 9 (1):6.

Sober, Elliott. 2009. "Absence of evidence and evidence of absence: evidential transitivity in connection with fossils, fishing, fine-tuning, and firing squads." *Philosophical Studies* 143 (1):63-90.

Solomon, Miriam. 2015. *Making Medical Knowledge*. Oxford: Oxford University Press.

Song, F., S. Parekh, L. Hooper, Y. K. Loke, J. Ryder, A. J. Sutton, C. Hing, C. S. Kwok, C. Pang, and I. Harvey. 2010. "Dissemination and publication of research findings: an updated review of related biases." *Health Technology Assessment* 14 (8):1-+. doi: 10.3310/hta14080.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*. 2 ed. New York: Springer-Verlag.

Starfield, B. 2006. "Threads and yarns: Weaving the tapestry of comorbidity." *Annals of Family Medicine* 4 (2):101-103. doi: 10.1370/afm.524.

Steel, Daniel. 2008. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford: Oxford University Press.

Steel, Daniel. 2010. "A New Approach to Argument by Analogy: Extrapolation and Chain Graphs." *Philosophy of Science* 77 (5):1058-1069. doi: 10.1086/656543.

Steel, Daniel. 2011. "Causal Inference and Medical Experiments." In *Handbook of the Philosophy of Science. Volume 16: Philosophy of Medicine*, edited by D. M. Gabbay, P. Thagard and J. Woods, 159-185. Amsterdam: Elsevier BV.

Stegeman, B. H., M. de Bastos, F. R. Rosendaal, A. van Hylckama Vlieg, F. M. Helmerhorst, T. Stijnen, and O. M. Dekkers. 2013. "Different combined oral contraceptives and the risk of venous thrombosis: systematic review and network meta-analysis." *BMJ* 347:f5298. doi: 10.1136/bmj.f5298.

Stegenga, Jacob. 2011. "Is meta-analysis the platinum standard of evidence?" *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 42 (4):497-507. doi: 10.1016/j.shpsc.2011.07.003.

Stegenga, Jacob. 2014. "Down with the Hierarchies." *Topoi-an International Review of Philosophy* 33 (2):313-322. doi: 10.1007/s11245-013-9189-4.

Stegenga, Jacob. 2015. "Measuring effectiveness." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 54:62-71.

Stehbens, W. E. 1992. "Causality in medical science with particular reference to heart disease and atherosclerosis." *Perspectives in Biology and Medicine* 36 (1):97-119.

Straus, Sharon E., Paul Glasziou, W. Scott Richardson, and Brian Haynes. 2011. *Evidence-Based Medicine: How to Practice and Teach It*. Edinburgh: Elsevier Churchill Livingstone.

Straus, Sharon E., W. Scott Richardson, Paul Glasziou, and R. Brian Haynes. 2005. *Evidence-Based Medicine: How to Practice and Teach EBM*. 4 ed. Toronto: Elsevier.

Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.

Susser, Mervyn. 1973. *Causal Thinking in the Health Sciences*. New York: Oxford University Press.

Sydenham, Thomas. 1848. "Medical observations concerning the history and cure of acute diseases." In *The Works of Thomas Sydenham, M.D.* London: Sydenham Society.

Szklo, Moyses, and F. Javier Nieto. 2007. *Epidemiology: Beyond the Basics*. 2 ed. Boston: Jones and Bartlett Publishers.

Tanenbaum, Sandra J. 1993. "What Physicians Know." *New England Journal of Medicine* 329:1268 - 1271.

Tanenbaum, Sandra J. 2014. "Particularism in health care: challenging the authority of the aggregate." *J Eval Clin Pract* 20 (6):934-41. doi: 10.1111/jep.12249.

Taylor, K. Kraupl. 1979. The Concepts of Illness, Disease and Morbus. Cambridge: Cambridge University Press.

Temkin, O. 1961. "The scientific approach to disease." In *Scientific Change*, edited by A.C. Crombie, 629-647. London: Heinemann.

Thagard, Paul. 1999. *How Scientists Explain Disease*. Princeton: Princeton University Press.

Thompson, R. P. 2010. "Causality, mathematical models and statistical association: dismantling evidence-based medicine." *J Eval Clin Pract* 16 (2):267-75. doi: 10.1111/j.1365-2753.2010.01383.x.

Thompson, R. P. 2011a. "Theories and Models in Medicine." In *Handbook of the Philosophy of Science. Volume 16: Philosophy of Medicine*, edited by D. M. Gabbay, P. Thagard and J. Woods, 115-136. Amsterdam: Elsevier BV.

Thompson, R. P. 2011b. Causality, theories and medicine. In P. M. Illari, F. Russo & J. Williamson (Eds.), *Causality in the Sciences*: Oxford Scholarship Online.

Timmermans, S, and M Berg. 2003. *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care*: Temple University Press.

Tinetti, M. E., S. T. Bogardus, and J. V. Agostini. 2004. "Potential pitfalls of disease-specific guidelines for patients with multiple conditions." *New England Journal of Medicine* 351 (27):2870-2874. doi: 10.1056/NEJMsb042458.

Tinetti, M. E., and T. Fried. 2004. "The end of the disease era." *American Journal of Medicine* 116 (3):179-185. doi: 10.1016/j.amjmed.2003.09.031.

Tonelli, Mark R. 1998. "The Philosophical Limits of Evidence-Based Medicine." *Academic Medicine* 73:1234 - 1240.

Turnbull, F., B. Neal, T. Ninomiya, C. Algerl, H. Arima, F. Barzi, C. Bulpitt, J. Chalmers, R. Fagard, A. Gleason, S. Heritier, N. Li, V. Perkovic, M. Woodward, and S. MacMahon. 2008a. "Effects of different regimens to lower blood pressure on major cardiovascular events in older and younger people: meta-analysis of randomised trials." *British Medical Journal* 336 (7653):1121-1123.

Turnbull, F., M. Woodward, B. Neal, F. Barzi, T. Ninomiya, J. Chalmers, V. Perkovic, N. Li, and S. MacMahon. 2008b. "Do men and women respond differently to blood pressure-lowering treatment? Results of prospectively designed overviews of randomized trials." *Eur Heart J* 29 (21):2669-80.

Upshur, R. E. G. 2001. "The ethics of alpha: reflections on statistics, evidence and values in medicine." *Theoretical Medicine and Bioethics* 22 (6):565-76.

Upshur, R. E. G. 2002. "If not evidence, then what? Or does medicine really need a base?" *Journal of Evaluation and Clinical Practice* 8 (2):113-119.

Upshur, R. E. G. 2005. "Looking for rules in a world of exceptions: reflections on evidence-based practice." *Perspectives in Biology and Medicine* 48 (4):477-489. doi: 10.1353/pbm.2005.0098.

Upshur, R. E. G. 2014. Do Clinical Guidelines Still Make Sense? No. *Annals of Family Medicine, 12*, 202-203.

Upshur, R. E. G., and J. Fuller. 2016. "Randomized controlled trials in the West African Ebola virus outbreak." *Clin Trials*. doi: 10.1177/1740774515617754.

Upshur, R. E. G., and C. S. Tracy. 2008. "Chronicity and complexity: Is what's good for the diseases always good for the patients?" *Canadian Family Physician* 54 (12):1655-1658.

Upshur, R. E. G., and C. S. Tracy. 2013. "Is evidence-based medicine overrated in family medicine?: Yes." *Can Fam Physician* 59 (11):1160-1.

Urbach, Peter. 1993. "The Value of Randomization and Control in Clinical Trials." *Statistics in Medicine* 12:1421 - 1431.

USSCF. 2007. The Intimidation of Dr. John Buse and the Diabetes Drug Avandia. United States Senate Committee on Finance.

Valle, John Del. 2012. "Chapter 293. Peptic Ulcer Disease and Related Disorders." In *Harrison's Principles of Internal Medicine*, edited by Dan L. Longo. New York: McGraw-Hill.

Vallvé C. 2003. "A critical review of the pragmatic clinical trial [in Spanish]." *Medicina Clínica* 27:384e8.

van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Clarendon Press.

van Hylckama Vlieg, A., F. M. Helmerhorst, J. P. Vandenbroucke, C. J. Doggen, and F. R. Rosendaal. 2009. "The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study." *BMJ* 339:b2921. doi: 10.1136/bmj.b2921.

van Loo, H. M., and J. W. Romeijn. 2015. "Psychiatric comorbidity: fact or artifact?" *Theor Med Bioeth* 36 (1):41-60. doi: 10.1007/s11017-015-9321-0.

Van Spall, Harriette G. C., Andrew Toren, Alex Kiss, and Robert A. Fowler. 2007. "Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review." *Journal of the American Medical Association* 297 (11):1233-1240. doi: 10.1001/jama.297.11.1233.

van Weel, C., and F. G. Schellevis. 2006. "Comorbidity and guidelines: conflicting interests." *Lancet* 367 (9510):550-551. doi: 10.1016/s0140-6736(06)68198-1.

Wakefield, J. C. 1992. "The concept of mental disorder. On the boundary between biological facts and social values." *Am Psychol* 47 (3):373-88.

Walker, Agnes Emilia. 2007. "Multiple chronic diseases and quality of life: patterns emerging from a large national sample, Australia." *Chronic illness* 3 (3):202-18. doi: 10.1177/1742395307081504.

Walton, Douglas. 2006. "Rules for reasoning from knowledge and lack of knowledge." *Philosophia* 34 (3):355-376.

Weiss, Noel S. 2006. *Clinical Epidemiology: The Study of the Outcome of Illness*. Oxford: Oxford University Press.

Weiss, Noel S., T. D. Koepsell, and B. M. Psaty. 2008. "Generalizability of the results of randomized trials." *Arch Intern Med* 168 (2):133-5.

Welsby, P. D. 1999. "Reductionism in medicine: some thoughts on medical education from the clinical front line." *Journal of Evaluation in Clinical Practice* 5 (2):125-131. doi: 10.1046/j.1365-2753.1999.00188.x.

Whitbeck, Caroline. 1977. "Causation in medicine: The disease entity model." *Philosophy of Science* 44 (4):619-637.

WHO. 2005. Preparing a Health Care Workforce for the 21st Century: the Challenge of Chronic Conditions. Geneva: World Health Organization.

WHO. 2011. Global Status Report on Noncommunicable Diseases 2010. Geneva: World Health Organization.

WHO. 2014. Global Status Report on Noncommunicable Diseases 2014. Geneva: World Health Organization.

WHO. 2015. World Report on Aging and Health. Luxembourg: World Health Organization.

Williams, Neil E. 2007. "The Factory Model of Disease." *Monist* 90:555 - 584.

Wolff, J. L., B. Starfield, and G. Anderson. 2002. "Prevalence, expenditures, and complications of multiple chronic conditions in the elderly." *Archives of Internal Medicine* 162 (20):2269-2276. doi: 10.1001/archinte.162.20.2269.

Wolf-Maier, K., R. S. Cooper, H. Kramer, J. R. Banegas, S. Giampaoli, M. R. Joffres, N. Poulter, P. Primatesta, B. Stegmayr, and M. Thamm. 2004. "Hypertension treatment and control in five European Countries, Canada, and the United States." *Hypertension* 43 (1):10-17. doi: 10.1161/01.hyp.0000103630.72812.10.

Woodward, J. 2003. *Making Things Happen: A Causal Theory of Explanation*. Oxford: Oxford University Press.

Woolcock, Michael. 2013. "Using case studies to explore the external validity of 'complex' development interventions." In *HKS Faculty Research Working Paper Series*. Boston: Center for International Development at Harvard University.

Woolf, S. H. 2008. "The meaning of translational research and why it matters." *Jama* 299 (2):211-3.

World Bank. 1995. "Tamil Nadu and Child Nutrition: A New Assessment." Washington, DC: World Bank.

Worrall, John. 2002. "What Evidence in Evidence-Based Medicine?" *Philosophy of Science* 69:S316 - S330.

Worrall, John. 2007. "Why there's no cause to randomize." *British Journal for the Philosophy of Science* 58 (3):451-488. doi: 10.1093/bjps/axm024.

Worrall, John. 2010. "Do we need some large, simple randomized trials in medicine?" In *EPSA Philosophical Issues in the Sciences*, edited by M. Suarez, M. Dorato and M. Redei. Netherlands: Springer.

Yusuf, S., P. Held, K. K. Teo, and E. R. Toretsky. 1990. "Selection of patients for randomized controlled trials: implications of wide or narrow eligibility criteria." *Stat Med* 9 (1-2):73-83; discussion 83-6.

Yusuf, S., R. Collins, and R. Peto. 1984. "Why do we need some large, simple randomized trials?" *Stat Med* 3 (4):409-22.

Zuiderent-Jerak, Teun, Frode Forland, and Fergus Macbeth. 2012. "Guidelines should reflect all knowledge, not just clinical trials." *British Medical Journal* 345. doi: e670210.1136/bmj.e6702.

Zwarenstein, Merrick, and Shaun Treweek. 2009. "What kind of randomized trials do we need?" *Journal of Clinical Epidemiology* 62:461-463.

Copyright Acknowledgement