Fall 11-18-2017

Teaching for Epidemiological Literacy: Description, Prescription, and Critical Thinking

Peter J. Taylor
University of Massachusetts Boston, peter.taylor@umb.edu

WORKING PAPER on Science in a Changing World

Paper # 4 -2017

Teaching for Epidemiological Literacy

Description, Prescription, and Critical Thinking

PETER J. TAYLOR

GRADUATE PROGRAM IN CRITICAL AND CREATIVE THINKING
University of Massachusetts, Boston, MA 02125, USA
http://scholarworks.umb.edu/cct_sicw/4

Teaching for Epidemiological Literacy: Description, Prescription, and Critical Thinking

Peter J. Taylor
Science in a Changing World graduate track
University of Massachusetts, Boston, MA 02125, USA
peter.taylor@umb.edu

Abstract

This working paper describes contrasting ideas for a sequence of topics as presented to students in a graduate course on *epidemiological literacy*. The premise of the pedagogical approach is that researchers develop their epidemiological thinking and practice over time through interactions with other researchers who have a variety of in-practice commitments, such as to kinds of cases and methods of analysis, and not simply to a philosophical framework for explanation. In descriptively teasing out what epidemiologists do in practice through a topic-by-topic presentation, I am prescriptively encouraging discussants to draw purposefully from across the range of topics and contrasting positions, and thereby pursue *critical thinking* in the sense of *understanding ideas and practices better when we examine them in relation to alternatives*. The initial topic concerns ways to learn in a community; after that, a number of conceptual steps follow—the characterization of the very phenomena we might be concerned with, the scope and challenges of the field of epidemiology, the formulation of categories—before linking associations, predictions, causes and interventions and examining the confounding of purported links. Building on that basis, the remaining topics consist of issues or angles of analysis related to the complexities of inequalities within and between populations, context, and changes over the life course. In the course of the description, some assertions about explanation and intervention emerge, notably, that epidemiological-philosophical discussion about causality often leaves unclear or unexamined whether a modifiable factor shown to have been associated with a difference in the data from past observations should be thought of as factor that, when

modified, would generate that difference going forward. The article ends with conjectures that concern heterogeneity and the agency of the subjects of epidemiology.

Keywords: causality, critical thinking, description-prescription, heterogeneity, inequality, intervention

Introduction

To undertake philosophy of epidemiology is—or should be—to get involved in relationships between *description* and *prescription* that arise at four levels (Stegenga 2009). First: when do the patterns that epidemiologists detect in observations of illness measures and other variables warrant action to change those variables—and by whom and how? Reciprocally, in what ways do ideas about actions favored by clinicians or health policymakers shape the kinds of patterns that get looked for. Then, at a second level, how much is philosophy of science about what epidemiologists do in practice versus what they leave unclear or under-examined, which philosophers try to resolve or shed light on? The latter effort implies that the views or practices of scientists can be improved, so—the third level—by what means do philosophers envisage that their accounts can influence researchers? Finally, whether the accounts made by philosophers are descriptive or prescriptive, explicit about the means of effecting change in science or not, by what means do philosophers of epidemiology envisage influencing others in their own field to change their views or practices?

Integrating the four levels of description and prescription can be straightforward if philosophy of epidemiology focuses on explanation and views soundness of explanation in terms of what I shall call an *interventionist* model of causality. That is, of the many factors possibly associated with the outcome of interest, one is modified in a randomly chosen subset of the population; the other factors—including ones that may not be modifiable—vary randomly across all subjects. When the focal factor is shown to be statistically significantly associated with the outcome, then clinical practice or health policy should modify the factor going forward. This model obviously links description to prescription at the first level. With respect to the other levels: philosophers who resolve or clarify issues about such explanations and about interventionist causality could expect to influence epidemiologists because the latter want their work to contribute ways to improvements in health; by extension, these philosophers would expect to influence colleagues who want their philosophical work to influence epidemiologists and, through them, people's health.

This article does not, however, focus on explanation or interventionist causality. Instead, it describes contrasting ideas for a sequence of topics as presented to students in a graduate course on *epidemiological literacy*. (Epidemiology here refers not to the dynamics of disease epidemics, but to the analysis of data from populations with a view to identifying the biological and social influences on the development of diseases and behaviors.) The initial topic concerns ways to learn in a community; after that, a number of conceptual steps follow, starting with characterization of the very phenomena we might be concerned with, leading up the making and confounding of explanations and causal claims. Building on that basis, the remaining topics consist of issues or angles of analysis related to the complexities of inequalities within and between populations, context, and changes over the life course (see Box below).

Sequence of Topics

1. The course as a learning community
2. Phenomena: Exploring the natural history of disease
3. The scope and challenges of epidemiology
4. Categories
5. Associations, Predictions, Causes, and Interventions
6. Confounders and conditioning of analyses
7. Variations in health care
8. Heterogeneity within populations and subgroups
9. Placing individuals in a multileveled context
10. Life course epidemiology
11. Multivariable "structural" models of development
12. Heritability, heterogeneity, and group differences
13. Genetic diagnosis, treatment, monitoring, and surveillance
14. Popular epidemiology and health-based social movements; Taking Stock of Course: Where have we come and what do we need to learn to go further?

(The body of the article provides entry points to these topics. The full set of readings and other course materials, with links to instructional aids and options for contributions from non-students, are viewable at http://ppol753.wikispaces.umb.edu/Visitors.)

The premise of this pedagogical approach is that students who are becoming researchers will continue to develop their epidemiological thinking and practice over time through interactions with other researchers who have a variety of in-practice commitments, such as to kinds of cases and methods of analysis, not simply to a philosophical framework for explanation. The literacy this course aims for, then, is one in which discussions about how to proceed in epidemiology draw purposefully from across the range of topics and contrasting positions. Literacy then is about *critical thinking* in the sense of *understanding ideas and practices better when we examine them in relation to alternatives* (Taylor 2002).

In descriptively teasing out through a topic-by-topic presentation of what epidemiologists do in practice, I cannot avoid being prescriptive. The implication is that researchers—not just students becoming researchers—would benefit from discussions that draw from across the range of topics and contrasting positions. Let me concede, however, that if I were designing a new course now, it might take the form of a semester-long unpacking of the recent article by two leading social epidemiologists, Krieger and Davey Smith (2016). It seems plausible that epidemiologists would be more likely to change their practice after hearing from epidemiologists who are philosophically informed but also pragmatic about what balance to strike between referring to what epidemiologists do versus to what they need to clarify or do differently. That possibility speaks also to the fourth level of description-prescription. I am not prepared to argue that my account of a course, which makes little use of the specialized literature or philosophical terminology, is the best way to convince philosophers of epidemiology, who probably expect some focused argument that takes on the focused arguments of others, especially about explanation and causality. My hope nonetheless is that philosophers and epidemiologists will, reading in the spirit of critical thinking, understand their expectations and practices better when they consider them in relation to the alternative exposition that this article represents.

If fostering critical thinking seems a modest goal, some stronger positions about the first level of description-prescription emerge. Most notably, it is often the case that

epidemiological-philosophical discussions about causality (including Krieger and Davey Smith 2016) leave unclear or unexamined whether a modifiable factor shown to have been associated with a difference in the data from past observations should be thought of as factor that, when modified, would generate that difference going forward. Another proposition that emerges concerns *heterogeneity* and the relationship between the patterns detected by epidemiologists and actions warranted by them. The closing section highlights the two propositions and opens up some conjectures that epidemiologists—and philosophers who descriptively and prescriptively discuss epidemiology—might examine more further.

Terminological note: Unless specifically noted otherwise, the terms *factor* and *variable* are used in this article in a non-technical sense simply to refer to something whose presence or absence can, at least in principle, be observed or whose level can be measured. Whether or not the factor or variable can be modified is a separate matter.

The Course: Epidemiological Thinking and Population Health

Week 1. The course as a learning community

Idea 1.1: Developing epidemiological literacy requires: a. collaboration with others (of differing skills and interests; b. reflection on personal and professional development; and c. establishing practices of learning from material we do not fully grasp at first reading or hearing.

Idea 1.2: Non-specialists need to become comfortable with the fundamental ideas and basic vocabulary of epidemiology in order to converse intelligently with specialists in epidemiology and biostatistics. One way to move in that direction is to practice making the ideas accessible to the layperson.

Let me elaborate on these ideas and the implied contrasts (as I will do for each of the weeks/topics to follow). The term epistemology may denote a focus on what makes beliefs in a specific knowledge claim justified, but it may also connote examining the *processes* through which knowledge gets established. In either case, epistemology

may seem to presume a confident individual knower. However, with its emphasis on developing epidemiological literacy, the course acknowledges that the processes through which knowledge or understanding gets established for an individual may be enhanced by attention to the tentativeness, cooperation, and communication.

2. Phenomena: Exploring the natural history of disease

Idea: Detailed observation (like naturalists make) or detective work—albeit informed by theoretical ideas—may be needed before we can characterize what the phenomenon is we are studying, what questions we need to ask, and what categories we need for subsequent data collection and analysis.

Analysis of data enters quickly in standard epidemiology texts, whether they are positioned at the accessible level of, say, Gordis (2013), or the advanced level of Rothman et al. (2012). But epidemiology need not begin with data sets to analyze. There may be exploratory, investigative, detective, anthropological, and naturalist inquiries before phenomena are even noticed, categories are defined, and questions are framed. Work to define phenomena is illustrated well by John Snow's famous use of maps to detect associations between cases of cholera in London in 1854 and water pumps, which supported his view that the infection spread through water not bad air (miasma) and his closing off the water supply from certain pumps. Snow, it should be noted, had clear hypotheses that guided his mapping; his action certainly did not follow from simply noticing patterns in the data and hypothesizing about the causes (Brody 2000). In short, defining phenomena is not a simple matter of induction, which raises the perennial question for philosophy of science of where hypotheses that get assessed by research come from in the first place. This question can be fruitfully explored through further examples of phenomena-defining work provided by Allchin (2013) on Eijkman's investigations of beriberi, Barker (1971) on buruli disease in Uganda, Oxford (2005) on teasing out the diverse factors that, in conjunction, led to the 1918 flu pandemic, or Cohen (2014) on chronic kidney disease of unknown etiology.

3. The scope and challenges of epidemiology

Idea 3.1: The uses of epidemiology are many, but shift over time, and are subject to recurrent challenges from inside and outside the field.

Idea 3.2: In advising on the most effective measures to be taken to improve the health of a population, epidemiologists may focus on different determinants of the disease than a doctor would when faced with sick or high-risk individuals.

Morris (1957) is a pioneering text in the kind of epidemiology discussed in this paper, namely, concerning the "systematic approach to the population aspects of non-communicable disease" Davey Smith (2001). In identifying seven uses of epidemiology (see Box below), Morris also invites us to consider whether epidemiology is a single thing to examine and whether or why the currently dominant approach, namely, #7, is the best focus for philosophical attention.

Epidemiology is the only way of asking some questions in medicine, one way of asking other [questions] (and no way at all to ask many). Seven 'uses' of epidemiology have been described:

1. In *historical study* of the health of the community and of the rise and fall of diseases in the population; useful 'projections' into the future may also be possible.
2. For *community diagnosis* of the presence, nature and distribution of health and disease among the population, and the dimensions of these in incidence, prevalence, and mortality; taking into account that society is changing and health problems are changing.
3. To study the *workings of health services*. This begins with the determination of needs and resources, proceeds to analysis of services in action and, finally, attempts to appraise. Such studies can be comparative between various populations.
4. To estimate, from the common experience, *the individual's chances* and risks of disease.

5. To *help complete the clinical picture* by including all types of cases in proportion; by relating clinical disease to the subclinical; by observing secular changes in the character of disease, and its picture in other countries.
6. In *identifying syndromes* from the distribution of clinical phenomena among sections of the population.
7. In the *search for causes* of health and disease, starting with the discovery of groups with high and low rates, studying these differences in relation to differences in ways of living; and, where possible, testing these notions in the actual practice among populations.

Brandt and Gardner's (2000) historical account shows that physicians have often opposed an increasing role for public health and, by extension epidemiology. Epidemiology might be valued for quantitative assessment of new interventions and evaluating patient safety and healthcare quality (fitting under Morris's use #3), but its role beyond evaluation and assessment, especially in regards to social, cultural, and economic factors influencing diseases, has continued to be contested. At the conceptual, more than sociological, level, the contest is between treatment of sick or high-risk individuals and taking population-wide measures to reduce the frequency of such individuals (Rose 1985 and commentaries in Ebrahim and Davey Smith 2001).

Alcohol consumption and road accidents provide a good illustration of Rose's "sick individuals-sick populations" contrast. Perhaps you have been able to get home safely even after drinking too much, but we also know that a substantial fraction of people in road accidents have high alcohol levels. Some people seem more susceptible to having their judgement and reaction times impaired by alcohol, but drink-don't-drive campaigns are directed at everyone; they are population-wide measures. It is easy, however, to imagine a formula to assess an individual's risk of accident that factors in not only the proximate alcohol consumption, but also background factors of, say, visual acuity, gender, age, driving with teenage passengers, cell phone habits, alcohol dehydrogenase gene variants, etc. More refined assessments of riskiness could, in principle, help focus risk-prevention efforts on high-risk individuals. Yet, we might ask,

would the net benefit (benefits minus costs) be significant relative to that from drink-don't-drive efforts? Indeed, in a society that had eliminated driving after drinking, discovering which genes might confer some susceptibility to alcohol among drinkers would be irrelevant to reduction in road accidents. Then again, as a political or sociological matter, would campaigns directed at everyone be allowed to go so far as to achieve the goal of no driving after drinking?

As an illustration of the idea of making epidemiological thinking accessible to the layperson (#1.2), note how the Rosean contrast and its implications arise in popular debates outside the health field. Following shooting rampages in the USA, Rosean risk reduction is put forward in a number of disparate forms: restricting availability of automatic weapons; providing less publicity to individuals who claim that they have to arm themselves against the tyrannies of the government; improving mental health funding so that help would be given to distressed individuals; and so on. For each proposed method, questions arise: would it be practical? …politically feasible? …effective? How would policy address fractions of the population (e.g., so-called "responsible gun owners") who see no benefit from the population-wide risk reduction and even harm? Discussions often shift from population responses to the notion that rampages are the work of deranged individuals. Yet, if the focus were to be on high-risk individuals, why are medical practitioners discouraged (or even prohibited) from discussing whether guns are accessible in the households of their patients?

Returning to challenges to the uses of epidemiology (#3.1), challenges *within* the field occur at regular intervals, especially around the contrast Pearce (1996) identifies as "bottom-up" versus "top-down" approaches. The latter begins at the population level in order to determine the primary socioeconomic factors that effect health. Bottom-up approaches, e.g., molecular epidemiology, begin on the individual level and aim to proceed upward toward explaining population level patterns. Description parallels prescription in the contrast between political engagement to change the macro-factors and physician or patient responsibility in relation to an individual's modifiable risk

factors. (See also Putnam and Galea 2008 and McMichael's 2011 review of Krieger's 2011 *Epidemiology and the People's Health*.)

4. Categories

Idea: Collecting and analyzing data requires categories: Have we omitted relevant categories or mixed different phenomena under one label? What basis do we have for subdividing a continuum into categories? How do we ensure correct diagnosis and assignment to categories? What meaning do we intend to give to data collected in our categories?

The theme that epidemiology does not begin with data sets to analyze (#2.1) is extended by the idea and questions above. The definition of categories shape the observations that can be made, the data collected from the observations, the associations or patterns perceived in the data, and so on. For example, early on in Galton's lifelong collection of data on human traits of many and varied kinds, he decided not to record "those that were imposed by the circumstances of their… lives" and focus on the "effects of tendencies received at birth" (Galton 1875, 566). The patterns of similarity he detected among relatives may have been sound, but only allowed for hypotheses and patterns about biological, not social inheritance, and spoke only to his prescriptive interests in the area he called eugenics (Taylor 2008). Closer to the present, Poland (2004) rejects the category of schizophrenia as defined by the Diagnostic and Statistical Manual (and elsewhere). Making use of such a category to describe patients makes it harder, he argues, for a clinician to pay attention to the contextual and life history information of patients. Even the milder position that the label schizophrenia is an umbrella term for heterogeneous conditions obviously has implications for investigation to expose the genes that influence "schizophrenia" (see #12 and 13 below).

Teasing out the assumptions along the chain of steps in scientific inquiry—from all possible phenomena that could be inquired into through categories demarcated, to observations made using those categories, to actions supported by predictions or to

causal claims—is obviously a matter for philosophical attention. Because the assumptions are not always dictated by the phenomena or justified by the results, teasing out the steps invites attention to the negotiations and wider influences that shape how the steps end up being made (Taylor 2005, 33-46; 2008). When we observe philosophers focusing on the logic but not the *sociologic* of the steps they observe epidemiologists making, we could well inquire into the prescriptive interests they might be enacting (i.e., the third level of description-prescription in the introduction).

Let me note three specific category choices that have prescriptive implications. First, incidence—new cases per unit time—versus prevalence—the caseload at any point of time. The public health burden of say, Alzheimer's dementia, is related to its prevalence; for epidemiologists to focus on its incidence is to imply that identifying risk factors for incidence can lead either to public health measures or other policies to reduce those factors in the population or to biomedical research that would trace and ultimately disrupt the pathways from the risk factor to the disease. Second, the choice to focus on the absolute incidence of an illness or on the relative incidence, in which one group is compared with another. Measures and policies to reduce the risk factors for absolute incidence may save lives even though the inequality among groups persists (Lynch et al. 2006; see #6 and 7 for further discussion). Finally, the seemingly mundane descriptive issue of how well the observations are made in the category chosen (e.g., rounding off blood pressure to the nearest 5mm Hg) animates various disputes in epidemiology about prescriptively relevant associations (Huxley et al. 2002; see #10).

5. Associations, Predictions, Causes, and Interventions

Idea: With respect to the relationships among associations, predictions, causes, and interventions that run through most cases and controversies in epidemiology, the field has two faces: One from which the thinking about associations, predictions, causes, and interventions are allowed to cross-fertilize, and the other from which the distinctions among them are vigorously maintained, as in "Correlation is not causation!" The second face views Randomized Control Trial (RCTs) as the "gold-standard" for testing

treatments in medicine. The first face recognizes that many hypotheses about treatment and other interventions emerge from observational studies and often such studies provide the only data we have to work with. What are the shortcomings of observational studies we need to pay attention to?

On this last question, examples such as the following kind are familiar: Being under treatment with statins was observed to be associated with lowered risk of dementia (Jick et al. 2000). In subsequent prospective studies, however, use of statins at the outset was not associated with lower development of Alzheimer's in the future (Zandi et al. 2005). The discrepancy seems to be consistent with an unrecognized bias in which elderly patients in the original study had been prescribed statins—patients with undiagnosed dementia were less likely to receive treatment. An even stronger check on results from observational studies are RCTs (Lawlor et al. 2004), as illustrated when the Women's Health Initiative clinical trial reported that hormone therapy increased rather than decreased the risk of coronary heart disease in women.

The use of RCTs builds in the *interventionist* model of causality defined in the introduction. To reiterate: Of the many factors possibly associated with the outcome of interest, one is modified in a randomly chosen subset; the other factors—including ones that may not be modifiable—vary randomly across all subjects. The same model of causality also informs Mendelian randomization (Davey Smith and Ebrahim 2007), but here nature modifies the factor in a randomly chosen subset. Is there an association between, for example, levels of cReactive Protein (CRP) in the blood and coronary heart disease (CHD) for people who have a rare genetic variant that leads to life-long elevated CRP levels, but otherwise vary randomly on other risk factors for CHD (such as smoking, bodymass index, and blood pressure)? (Notice that the interventionist model in epidemiology differs from typical experiments in the laboratory, in which the background factors are controlled, not randomly varying, across replicates of the experimental intervention.)

Ambiguity regarding causality is obvious in the common term for variables associated with an outcome of interest, *risk factor*. The term has connotations of interventionist causality, of something that, if altered, reduces risk. However, associations with risk factors can allow for clinically useful predictions even when those factors are not modifiable, such as age or gender, and even when modifying the level of the factor does not improve the outcome. For example, Ridker et al. (2007) propose a composite of risk factors for CHD in women, the Reynolds Risk Score, that improves on the conventional Framingham score, primarily, it seems, by including CRP levels. "Improve" here means fewer women assigned to the medium or low-risk categories had subsequent coronary events; by implication, clinicians could feel more confident in focusing their attention on individuals assigned to the high-risk category. Not surprisingly, researchers such as Ridker became interested in the idea that intervening to reduce CRP could improve CHD outcomes. Mendelian randomization subsequently cast doubt on that hypothesis (C Reactive Protein Coronary Heart Disease Genetics Collaboration 2011), yet the clinical value of the Reynolds Risk Score remains.

The phrase "not surprisingly" used above betrays the common expectation when a factor is associated with an outcome, typically as significant variable in some kind of regression equation, it is a plausible candidate for inclusion in explanations or hypotheses about interventionist causality. It may be noted, however, that, at the very foundations of fitting regression equations to data lies two alternative pictures (Weldon 2000). The first is that the so-called independent variables are combined in the regression equation to provide the best *prediction* of the dependent variable (and thus become the plausible causal candidates above). The second picture follows from seeing that, for the simplest case of one variable used to predict a second, the slope of the regression line when the two variables are scaled to have equal spread (standard deviation) is the same as their correlation; this value is also a measure of how tightly the cloud of points is packed around the line of slope 1 (or slope -1 for a negative correlation). Technically, when both measurements are scaled to have a standard deviation of 1, the average of the squared perpendicular distance from the points to the line of slope 1 or -1 is equal to 1 minus the absolute value of the correlation (Weldon

2000). This means that the larger the correlation, the tighter the packing. This *tightness-of-packing* picture of correlation—and, by extension, of regression equations—affords no priority to one measurement over the other in prediction. This second picture means that a good predictor is not in itself a basis for the causal plausibility of a variable; linking prediction and causality must depend on considerations beyond the statistical analysis of data.

A looser alternative to the interventionist model of causality is to view statistical analysis as identifying differences that make a difference. In this model, the differences—typically departures of a factor from a mean value—need not be modifiable (e.g., chromosomal sex is a commonly measured but non-modifiable genetic factor). Moreover, if the factors were modifiable, *it does not follow that modifying them would generate the differences observed in the original data set*. In other words, it does not follow that the difference that "makes" a difference as exposed by statistical analysis of data (outside RCTs and Mendelian randomization) is a factor one can *modify to make the same difference again*. For example, lower income level is a significant factor associated with smoking rates, but there is no reason to expect that disbursing $10,000 to poor smokers would lead many of them to quit. After all, the dynamics through which a person develops a low income and the dynamics through which a person becomes a smoker are separately and jointly far more complex than any static statistical, differences-that-make-a-difference model can capture. (Obviously this reservation does not apply to RCTs, but it might well apply in Mendelian randomization given that, if, say, the genetic variant inducing lifelong elevated CRP levels had been associated with CHD, modifying CRP for future patients would be by means other than giving them that variant at birth.)

A curious prescriptive implication is shared by both the interventionist and the statistical, differences-that-make-a-difference models. When a significant result becomes the basis for practice or policy, *variation around the mean gets discounted*. For example, imagine a comparison of the dental health of two communities that have the same range of health problems except that the one with naturally high level of fluorides in its

water supply has better than average dental health. In each community there will be variation around the average dental health. However, if the variation is small relative to the differences in the two averages, it might seem reasonable to advocate fluoridation of water supplies lacking natural fluoride. In doing so the variation around the average is discounted (as are other deviations from type, such as teeth discoloration that occurs in some individuals). The alternative would be for tablets to be taken by each individual, which would allow the dosage to be customized according to a person's dental health habits and disposition. This individual approach is not preferred by most public health policy-makers, who point to lack of "compliance" when individuals are responsible for administering their own preventative medicines. Discounting of variation around the mean could, however, trouble epidemiologists and population health researchers. Consider, for example, the persistent differences on average in various scholastic achievement tests between so-called racial groups. When researchers set out to explain these average differences are they assuming that educators will treat individuals according to the average of the group to which they belong? This question might even lead us to ask what exactly is meant by trying to *explain a difference between the means* of two groups. (See Davey Smith 2011 and Taylor 2014a for contrasting positions on whether and when to discount heterogeneity in favor of average differences between groups; see #8 and 12.)

6. Confounders and conditioning of analyses

Idea: Statistical associations between any two variables generally vary depending on the values taken by other potentially "confounding" variables. We need to take this dependency or conditionality into account when using our analyses to make predictions or hypothesize about causes, but how do we decide which variables are relevant and real confounders?

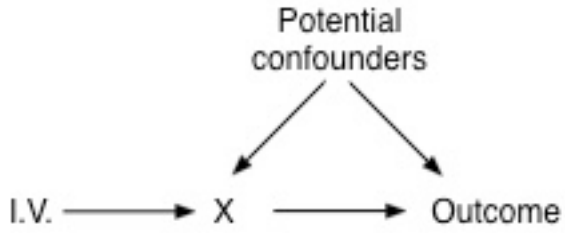
The descriptive moment in conditioning of analyses simply envisages the observations as divided into slices, each slice containing only observations that share the same value (or range) of a given variable, such as age. We do not want the comparison of two groups to be distorted by one group having a larger fraction in some slices (finding, say,

women to be at a greater risk for Alzheimer's dementia without noting that women are a larger fraction of older age classes), so the statistical analysis is run on a data set in which, in effect, the number of observations in each slice are balanced out. Perhaps the comparison would be different if we focused on each slice separately, but typically the analysis averages those separate comparisons into a single *conditioned* or *adjusted* comparison.

The prescriptive moment in conditioning of analyses emerges in disputes around adjustments not made or inappropriate adjustments, in which the implication is that actions that might be supported by the unadjusted or inappropriate adjustments are not justified. In these contexts, the term *confounder* or confounding variable is typically used. The original association between hormone replacement therapy and lowered CHD incidence, for example, was supported by studies that did not adjust for the on-average higher socioeconomic status (SES) of the women receiving the therapy. In that light, prescribing the therapy across all SES should not be expected to result in comparable lowering of CHD incidence for all (Petitti 2004); indeed that turned out to be the case. On inappropriate adjustments, Davey Smith et al. (1997) see elimination of the socioeconomic gradient in CHD incidence by statistical adjustment for self–reported job control as, in effect, an adjustment for SES given that low job control is associated with lower SES. Lynch et al. (2006) argue that the focus on the psychosocial factors (such as on job control) diverts the focus of health promotion away from the conventional risk factors (smoking, hypertension, dyslipidemia [unhealthy cholesterol levels], and diabetes), attention to which can reduce the absolute amount CHD incidence even if a socioeconomic gradient (i.e., its relative amount) were to persist. (The claim of unjustified adjustment is central to Huxley et al.'s [2002] critique of associations between early life experience and chronic adult disease [see #10 below]; but see Davies et al. [2006].)

Under the interventionist model of causality, it is clear how to show whether actions that might be supported by the unadjusted or inappropriate adjustments are justified. RCTs and Mendelian randomization demonstrate that a variable is a confounder if its

association with the outcome of interest disappears when it is modified in a subset of subjects while all other relevant factors vary randomly across all subjects. In the schema below, an instrumental variable (I.V.) would be the presence of the drug being tested or the control (in RCTs) or the rare genetic variant versus the normal variant (in Mendelian randomization). In both cases the I.V. is associated with the outcome only if it influences X (the effect of the drug or the rare variant). The I.V. is not associated with any other variable that is or might be associated with X or the outcome.



Notice that both the effect of a modifiable factor (I.V. -> X) and statistical associations are represented by arrows. The use of such diagrams to decide whether adjustment is appropriate revolves around bringing in qualitative, a priori, subject-matter knowledge of causal connections (Hernán et al. 2002) even when there is no obvious instrumental variable, that is, even for the broader class of statistical, differences-that-make-a-difference analyses. Whether, in practice, this extension discounts the kinds of issues about causality mentioned in topic #5, such as, what it means to explain a difference between averages across two (or more) groups, warrants philosophical attention. Moreover, description can underwrite prescription even without giving causal interpretation (of either type) to statistical associations. The use of the Reynolds Risk Score (Ridker et al. 2007; see #5) in effect separates the slices that have high from those that have low levels of CRP; it has the potential to improve the assignment of people to high, medium, or low risk for CHD and thus the effectiveness of preventative measures.

* * *

Building on the preceding topics, the rest of the course takes up issues or angles of analysis related to the complexities of inequalities within and between populations,

placing individuals in context, changes over the life course, and heterogeneous pathways. Before moving forward, let me acknowledge a contrasting approach, namely circling back around explanation following the article mentioned in the introduction, Krieger and Davey Smith (2016). They argue—and illustrate—that epidemiology needs to make use of a diversity of methods to determine the causal connections that are relevant to making policy and changing practice. In noting the "need to amass substantive expertise and to generate and think critically about contrastive hypotheses," these epidemiologists (who are cited often under the topics above or below) align their position with the advocacy by philosopher of science Lipton of *inference to the best explanation*.

7. Variations in health care

Idea: Inequalities in people's health and how they are treated are associated with place, race, class, gender; these inequalities may persist even after conditioning on other relevant variables.

Concern about inequalities in health among groups lies at the center of social epidemiology. (See Krieger 2010a for a detailed overview of a course on inequalities and health.) Any descriptive account of inequalities can readily be given a prescriptive interpretation. For example, after finding that "[f]or virtually all outcomes, risk increased with CT [census tract] poverty," Krieger et al. (2005) note that "[f]or half the outcomes, more than 50% of cases would not have occurred if population rates equaled those of persons in the least impoverished CTs." The prescription-by-counterfactual (technical term: population attributable fraction [PAF]) does, however, leave the *how* and *by whom* of the health-income improvement as a separate matter.

Indeed, the how of the disease-poverty association need not be obvious. Alter et al. (1999), for example, show that in Ontario where there is universal health insurance, access to specialized cardiac services is associated with SES even after statistically adjusting for factors corresponding to the reasonable assumption that specialist doctors and higher quality facilities would tend to be located in higher SES areas. What other

factors then are associated with the unequal access? Wright et al.'s (2004) study of asthma among children in low-income urban settings, after adjusting for SES and caretaker behaviors, such as smoking, found asthma to be associated with stress and exposure to violence. Krieger et al. (2005) shows the association of health inequalities with race or ethnicity is reduced after adjustment for socioeconomic deprivation (CT poverty), but not eliminated. Searching for associations that pertain specifically to race and ethnicity has led to results such as those of Mustillo et al. (2004) in which higher risk of pre-term delivery of babies to African-American women, was associated, after adjusting not only for income but also for alcohol and tobacco use, depression, and education, with income reported experience of racial discrimination.

As noted earlier (#4, 6), Lynch et al. (2006) question the value of research to pin down risk factors for the SES gradient in health when measures and policies already exist to reduce the major risk factors for absolute incidence; those measures and policies need to be the priority. A logical extension of their argument would be to question the value of research to pin down risk factors for gradients in health remaining after adjusting for SES when addressing the major risk factor, lower SES, needs to be the priority (see e.g., Krieger at al. 2005's conclusion above). The obvious counter-argument might be that while measures and policies to reduce smoking, hypertension, and so on seem feasible to Lynch and colleagues—they lie in the realms of clinical practice and health promotion—substantial reduction in SES inequalities lies beyond the ambit of epidemiology and seem difficult, even if important, especially given the political economic changes over the last 40 years that continue to *enlarge* such inequalities. In this argument, prescriptive assumptions shape the descriptive exercise of finding statistical associations. Similarly, even if research pinned down risk factors for gradients in health remaining after adjusting for SES, which in the USA might include specific features of racial discrimination, measures to change the dynamics producing, say, that discrimination may seem as difficult as they are important. Yet, a counter to this counter-argument might be that, when descriptive accounts of such associations are not available, it is harder to bring the unfairness or injustice of health inequalities to bear in the prescriptive realm of politics and policy making.

The preceding discussion of inequalities not only introduces contrasts in how to interpret variation in health, but also points to issues that readily arise about how to measure and track health variations (#4). The analysis of Krieger et al. (2005), for example, responded to the lack of socioeconomic data in US public health surveillance systems by geocoding records according to census tract, for which poverty rates were available. Krieger (2010b) acknowledges the social and historically changing definitions of race and ethnicity as well as the necessity of employing them if the ways that "racism harms health" are to be exposed. Krieger (2014) notes that, even with the increase of studies that include experience of discrimination as a risk factor, the emphasis remains on person-person discrimination, not structural.

8. Heterogeneity within populations and subgroups

Idea: How people respond to treatment may vary from one subgroup to another. When is this a matter of chance or of undetected additional variables? How do we delineate the boundaries between subgroups?

If subgroups are defined *after* exploring the data, there is an obvious risk that they are shaped with a view to finding a significant association with some outcome of interest (which was evident in the case of the purportedly race-specific medicine BiDil; Kahn 2007). More generally, as statisticians caution, the more subgroupings that are explored the more chance that a significant association will arise by chance; Lagakos (2006) recommends therefore tighter criteria for claiming that an association is statistically significant. Ioannidis's (2005) article has stimulated wider scrutiny of fishing to find and publish on associations that turn out to be false positives or, at least, hard to reproduce (so-called P-hacking).

The opposite caution is that treating everyone as if they were from the same population (even if for good statistical reasons) distracts our attention from the clues that might lead to seeing ways that the population is not one uniform whole, but is a mixture of types or even more heterogeneous than that. Heterogeneity can have health care

implications. For example, when breast cancers are subdivided according to the responsiveness of the tumor to hormones, there is a qualitative difference in effectiveness, on average, of different regimes of chemotherapy and tamoxifen (Regan and Gelber 2005). Steinbach et al. (2014) examine the not-surprising association of lower injury from pedestrian accidents for children in affluent areas, but find that the association does not hold “for those in some minority ethnic groups." If we widen further what comes under the umbrella of health, Fazel’s (2012) review of instruments used for making decisions about sentencing, release or preventative detention in the criminal justice system argues that, when low-risk and high-risk offenders are separated, the predictive value of the instruments turns out to be very poor for the high-risk offenders.

Notwithstanding the preceding health implications of heterogeneity, Davey Smith (2011) warns against paying much attention to it (and against putting much hope in personalized medicine): considerable randomness at the individual level means that epidemiology should keep its focus on modifiable causes of disease at the population level. Taylor (2014a) counters or complicates the advice of Davey Smith with examples and arguments showing that: a) it can be quite reasonable to try to differentiate among individuals so as to improve risk prediction, even if finding ways to do so may not be straightforward; and b) when researchers think about the causal dynamics underlying patterns in data (such as associations with risk factors), it may be helpful *not* to view deviations from patterns as noise but as invitations to pay attention to the multiplicity of paths to the “same” trait and to other forms that heterogeneity takes (see also #10 and the closing section of this article).

9. Placing individuals in a multileveled context

Idea: Different or even contradictory associations can be detected at different levels of aggregation (e.g., individual, region, nation), yet not all influences can be assigned to properties of the individual. Membership in a larger aggregation may be associated with outcomes even after conditioning on the attributes that individual members have.

Associations at the level of nations between incidence of a disease, say, breast cancer,

and a given risk factor, say, dietary fat intake *suggest* associations at the level of individuals: among women who consume more fat we might expect there to be a greater incidence of breast cancer. Such *ecological inference* suggests, in turn, advice to women: reduce your dietary fat intake. Alternatives to the obvious inference need, however, to be considered. Perhaps dietary fat is associated, say, with higher standard of living and some other aspects of affluence can also be shown to be risk factors. Not only such confounding variables, but also alternatives that point in the *opposite* direction to the original suggestion may need to be identified and examined. Barker and Osmond (1986), for example, studied patterns in CHD, which is associated with increasing prosperity of a country and, by inference, with some risk factor(s) for individuals that had increased with affluence. In England, however, CHD turned out to be highest in districts that had poorest conditions for health, as measured by infant mortality 55 years earlier (see #10 for discussion of associations across the life course).

Scrutiny of suggestions is also needed when the aggregate-level variables have no equivalent for individuals. In an encyclopedia entry on ecological inference, Freedman (2001) showed that in 1995 U.S. states with higher fraction of foreign born tended to be the ones with higher fractions of higher income. An individual cannot be fractionally high income or fractionally foreign born, yet the association across states might be taken to suggest that the foreign born tend to have higher incomes. The inverse turned out to be the case. Finally, when individual-level associations are not as clear as association for aggregate-level variables, it may be worth scrutinizing whether the latter subsume a heterogeneity of conditions (#8) experienced by individuals (see, e.g., Khodarahmi and Azadbakht 2014 in relation to the dietary fat-breast cancer association). This last situation points to one of the difficulties of making inferences in the *opposite* direction—from risk factors at the individual level to risk factors associated with health differences among units at some level of aggregation above the individual. Indeed, for each situation in this and the previous paragraph, alternatives to the obvious inferences from individuals to aggregate units should be considered.

Hierarchical linear modeling address the problem of inferences across levels by, in

effect, examining an association within a group, say, CHD incidence in relation to an individual's income within a neighborhood or census tract, and then comparing the slopes and intercepts of the resulting regression equations across the groups. The nesting of individuals into groups is seen to be relevant if the slopes and intercepts are significantly different (Diez Roux 2002). Interpretation of significant differences in terms of some modifiable quality of the aggregate units, such as the number of playgrounds in a neighborhood, is difficult and contested (Oakes 2004), all the more so if proposed interpretations involve aggregate-level variables with no equivalent for individuals, such as income inequality within the neighborhood, or "complex causal chains with feedback loops and reciprocal effects" (Diez Roux 2002, 516). To reprise an earlier point, the dynamics through which income inequality evolves in a neighborhood and through which individuals' health or disease develops in their neighborhoods are more complex than any static statistical, differences-that-make-a-difference model can capture—certainly more complex than addressed by social science experiments of the kind that would, say, fund new playgrounds after finding an association between childhood obesity and the number of playgrounds in a neighborhood.

The importance—and complexity—of analyzing health in a multilevel context is illustrated by the study of Friedman et al. (2014), which found that a) population density of HIV+ people who inject drugs was positively associated with the density of non-injecting drug users; b) HIV prevention programs for people who inject drugs was negatively associated with "AIDS incidence among heterosexuals and… mortality among heterosexuals living with AIDS" several years later, but c) there was no such associations for HIV+ men who have sex with men. The authors recommend more research on how the non-injecting drug users may serve as a bridge between other populations and thus how interventions in one key population affect HIV epidemics in other populations.

10. Life course epidemiology

Idea: How do we identify and disentangle the biological and social factors that build on each other over the life course from gestation through to old age?

The finding of Barker and Osmond (1986) mentioned earlier, that CHD turned out to be highest in districts of England that had poorest conditions for health as measured by infant mortality 55 years earlier, opened up inquiry into the fetal or early life origins of chronic adult diseases. Mechanisms were suggested involving adaptation of fetal growth to undernutrition at different phases of gestation, with subsequent confirmation in experiments on animals (Barker 1998). For humans, it is difficult to isolate the association between a disease in later life and conditions during gestation or early life given that such conditions tend to be associated with similar conditions during childhood and beyond (Ben Shlomo and Davey Smith 1991). Researchers who conducted large-scale clinical trials or large observational studies of factors that could be modified in adult life were especially strong in their criticisms (Huxley et al. 2002; but see Davies et al. 2006). The fetal origins hypothesis had the potential to distract attention from demonstrable life-extending changes in adult life, such as smoking cessation and cholesterol-lowering use of statins. Yet, if transitions across generations (e.g., rural to urban migration, public health measures, nutritional improvements) that influence the nutrition mothers are able to provide their fetuses as well as the subsequent conditions for the offspring could be shown to be associated with the rise and subsequent decline in CHD incidence in a country (Barker 1987, 1999), the result would be relevant to understanding epidemiological patterns even if it did not translate into clear clinical recommendations.

The challenge raised by the fetal origins hypothesis was to assemble health data across the life course and develop methods to discriminate among factors from different stages with respect to their association with diseases in later life. For example, establishing whether factors at one stage build on those of earlier stages or influence later disease separately (as would occur if there were specifically sensitive periods). This challenge was taken up by the field that emerged as life course epidemiology (Ben Shlomo and Kuh 2002; Davey Smith 2007).

An earlier line of research, initiated by the medical sociologists Brown and Harris in the late 1960s (Harris, 2000), employs a different and labor-intensive method to investigate

the role of factors from different periods of the life course. They combine wide-ranging interviews, ratings of transcripts for the significance of past events in their context (with the rating done blind, that is, without knowledge of whether the person became ill), and statistical analyses to investigate how severe events and difficulties during people's life course are associated with the onset of mental illnesses. An event, such as death of a spouse, might have very different meanings and significance for different subjects according to the context, which Brown and Harris's methods accommodate (see #4). At the same time, apparently heterogeneous events can be subsumed under one factor, such as, in explanation of depression, a severe, adverse event in the year prior to onset (see #4 and 8). For example, in the earliest work of Brown and Harris concerning a district of London in the early 1970s, they identified four factors as disproportionately the case for women with severe depression: a severe, adverse event in the year prior to the onset of depression; the lack of a supportive partner; persistently difficult living conditions; and the loss of, or prolonged separation from, the mother when the woman was a child under the age of 11 (Brown & Harris, 1978, 1989b) (see #9). In principle, even if results turned out to be specific to a given place, such an integration of "the quantitative analyses of epidemiology and [in] depth understanding of the case history approach" (Brown & Harris, 1989a, p. x) could be taken up more widely in epidemiology (Brown & Harris 1989b).

11. Multivariable "structural" models of development

Idea: Just as standard regression models allow prediction of a dependent variable on the basis of independent variables, structural models can allow a sequence of predictive steps from root ("exogeneous") through to highest-level variables. Although this kind of model seems to illuminate issues about factors that build up over the life course, there are strong criticisms of using such models to make claims about causes.

This idea is well illustrated by the work of Kendler and colleagues, who examine behavioral traits in relation to a wealth of factors or variables over the life course (Kendler and Prescott 2006). In Kendler et al. (2002), for example, data on over 1,900 twins are used to fit the incidence of major depression to a model that incorporates

many environmental factors and a so-called “genetic risk” factor. (This last factor is derived from the incidence of major depression in the co-twin and parents, with adjustments made for the degree of relatedness of the twins; monozygotic versus dizygotic; see #12). This kind of *path analysis* or *Structural Equation Modeling* (SEM) does not simply look for how the trait is associated with each of the factors, but quantifies their relative contributions (“path coefficients”) to the variation in the focal trait once a certain network of the factors has been specified. Some of these contributions are direct and others are mediated through other factors, i.e., indirect (Lynch & Walsh 1998, 823). Kendler’s model accounts for 52% of the variance in the incidence of major depression and provides a picture of development that is rich and plausible. For example, a path coefficient of .7 from neuroticism to low self-esteem and of .3 from low self-esteem to low education suggests that neuroticism makes it more likely that a person has low self-esteem and that, in turns, makes it more likely that they do not pursue education as far as others.

In one sense, interpretation of these paths is no different than for any other statistical analysis under a differences-that-make-a-difference model: no claim need be made that a given factor can be modified and if it were that the model would predict the outcome. In another sense, having paths pointed in one direction and calling the networks of linked factors “structural”—or my describing the picture of development in Kendler’s model as “plausible”—suggests stronger causal claims. But, where Pearl (2000, 135 and 344-5) sees path analysis in terms of variables that can be manipulated through their insertion or removal, Freedman (2005) argues against viewing path analysis/SEM models in interventionist terms: the equations (i.e., the coefficients and error terms) would have to be “stable under proposed interventions” and that this is difficult to verify without making the interventions. If the equations change when factors are manipulated, they have “only a limited utility for predicting the results of interventions” (matching the point made in #5). Freedman’s skepticism may be seen to temper the call (see #9) of Diez Roux (2002, 516) for attention to “complex causal chains with feedback loops and reciprocal effects.”

Kendler et al. (2002, 1133) show admirable reserve about how to interpret their model (as does Ou [2005] in SEM modeling of pathways of educational development from pre-school programs to later outcomes). Nevertheless, to the extent that this kind of model is meant to illuminate issues about factors that build up over the life course, the exclusion of certain factors and inclusion of others has prescriptive implications. The models of Kendler and colleagues, for example, do not include factors that correspond to therapeutic interventions or to social changes that have led to the rising incidence of depression. Data on these factors may not have been available or collected (#2 and 3), but sensitivity of the analysis to inclusion or exclusion of such factors warrants attention given the potential prescriptive implications (see #12).

12. Heritability, heterogeneity, and group differences

Idea: As conventionally interpreted, heritability indicates the fraction of variation in a trait associated with "genetic differences." A high value indicates a strong genetic contribution to the trait and "makes the trait a potentially worthwhile candidate for molecular research" that might identify the specific genetic factors involved. A contrasting interpretation is that there is nothing reliable that anyone can do on the basis of estimates of heritability for human traits. While some have moved their focus to cases in which measurable genetic and environmental factors are involved, others see the need to bring genetics into the explanation of differences for certain traits between the averages for groups, especially racial groups.

Partitioning of variation into fractions is the foundation of classical quantitative genetics, a field that arose in agriculture, where multiple varieties of plants can be grown in many plots in many locations. For a given trait, say, yield per plot, the variation can be partitioned (through the statistical technique of Analysis of Variance and its kin) into three components and what is left over or *residual*: between the means for each variety when averaged across locations; between the means for each location when averaged across varieties; and between the means for each variety-location combination when averaged across plots (and after taking out the preceding two components). Such partitioning is contingent on the specific set of varieties and locations. Despite its name,

quantitative genetics neither relies on nor produces knowledge about specific genetic and environmental factors that might be causing the yield in each variety-location combination. There is no obvious factor that could be modified under an interventionist model of causality. (This last point applies also to path analysis as used to partition variation; see #11.)

The contingent, descriptive quality of partitioning of variation becomes obscured, however, after the following common moves are made: varieties are referred to as genotypes; the variation among the variety or genotypic means across locations is called genotypic variance; this term is shortened to genetic variance; that quantity is interpreted as the fraction of variation in a trait associated with "genetic differences"; that quantity is called "heritability"; and it is discussed as if it had some relation with heritable in the sense of the transmission of genes from parents to offspring. The origin of these moves can be traced to the models used by quantitative genetics to partition trait variation, which, in order to take different degrees of relatedness into account (e.g., monozygotic twins being more closely related than dizygotic twins), posit theoretical, idealized genes that have simple Mendelian inheritance and direct contributions to the trait. (Given that the partitioning is of variation in *traits*, it must be possible to partition variation without using models of genes that are not observed [Taylor 2012]; such "gene-free" analyses have not been taken up in practice.)

Two developments in quantitative genetics might seem to undercut any concern that the genes in its traditional models are not observables. First, the technique of mapping quantitative trait loci (QTL) associates regions of the genome with variation in a continuously variable trait. Although most success has been had in animal and plant varieties that can be replicated and raised in controlled conditions, QTL analyses for human populations are advancing (Mackay et al. 2009; but see reservations of Majumder and Ghosh 2005). Second, in this age of genomics, it is possible to determine the presence or absence of actual genes and then, as epidemiology typically does, look for associations between variation in a trait and measured factors, in this case, levels of genes and environmental factors (Moffitt et al 2005). In short, to the

extent that molecular research now identifies specific genes or regions of the genome underlying variation certain traits, a high heritability value (in the traditional sense) would seem a plausible indicator as any that "the trait [is] a potentially worthwhile candidate for [such] molecular research" (Nuffield Council on Bioethics 2002, chapter 11).

However, the plausibility of heritability as a guide for what to investigate at molecular level may be disturbed by heterogeneity (#8), in the following way. Heritability (in the traditional sense) can be derived through partitioning of variation that employs data from relatives. The similarity of pairs of monozygotic twins (which share all their genes) can, for example, be compared with the similarity of pairs of dizygotic twins (which do not share all their genes). The more that the former quantity exceeds the latter, the higher is the trait's heritability (assuming for purposes of discussion that monozygotic twins are not treated more similarly than are dizygotic twins). Even if the similarity among twins or a set of close relatives is associated with similarity of (yet-to-be-identified) genetic factors, *the factors may not be the same from one set of relatives to the next, or from one environment to the next*. In other words, the underlying factors may be heterogeneous. It could be that pairs of alleles, say, AAbbcbDDee, subject to a sequence of environmental factors, say, FghiJ, during the development of the organism are associated, all other things being equal, with the same outcomes as alleles aabbCCDDEE subject to a sequence of environmental factors FgHiJ (Taylor 2012). Such *underlying heterogeneity* makes heritability an unreliable indicator of whether to study a trait with a view to exposing differences in actual genes associated with variation among variety or so-called genotypic means. (If we put aside traits associated with so-called high-penetrance major genes, e.g., polydactyly, there are no obvious grounds to rule out the possibility of heterogeneity in the measurable genetic and environmental factors that underlie patterns in quantitative and other complex traits, such as crop yield, height, human IQ test scores, susceptibility to heart disease, personality type, and so on.)

The possibility of underlying heterogeneity reminds us that statistical patterns such as the size of components of partitioned variation in a trait are distinct from measurable

underlying factors. This reminder has become more necessary since, in recent years, the same term heritability has been co-opted to refer to a conceptually and empirically distinct quantity, namely, the fraction of variation in a trait associated with variation in Single-Nucleotide Polymorphisms (SNPs) as examined by an extension of QTL analyses, namely, Genome-Wide Association (GWA) studies. It has turned out, however, that, for SNP loci where variants have a statistically significant association with some medically significant trait, that association corresponds to a small increase in incidence of the trait (McCarthy et al. 2008). Moreover, even when many such associations are considered jointly, most of the variation in the trait remains unaccounted for (Ku et al. 2010). The difference between high heritability in the traditional sense for, say, height, and the fraction of variation associated with SNPs (i.e., heritability in the new sense) led to discussions about so-called "missing heritability" (e.g., Zuka et al. 2012). Underlying heterogeneity provides one explanation for why GWA studies have had difficulties in identifying causally relevant genetic variants behind variation in human traits (Taylor 2014b).

When the presence or absence of actual genes can be determined and associations are found between variation in a trait and measured genetic and environmental factors, the distinction between statistical differences-that-make-a-difference and interventionist causality may get blurred. Caspi et al. (2002), for example, reports on antisocial behavior in adults in relation to the activity of monoamine oxidase type A (MAOA) and childhood maltreatment; MAOA deficiency is a strong predictor of antisocial behavior only when the child has also been maltreated. The authors conclude that their results "could inform the development of future pharmacological treatments." The obvious counter is that their results could also warrant more effort to reduce maltreatment of children. In any case, epidemiologists have noted that the PAF is very low for the Caspi et al. study, that is, few cases of anti-social behavior would be eliminated if MAOA was at the normal level or maltreatment was not present. Yet notice that, not only Caspi et al.'s conclusion, but also the critical responses rest on envisioning that the factors associated with the trait to be modifiable then assuming that modifying them would generate the differences observed in the original data set. Attempts to modify the

factors, however, may well entail new and possibly counter-productive measures, from intrusion of social services agencies into households to stereotyping and surveillance of low MAOA individuals (Taylor 2015; see also #13).

The possibility of finding associations between variation in a trait and measured genetic and environmental factors allows a further distinction to be made (or forgotten; Taylor 2015). A *genotype-* or *gene-environment interaction* in such studies means that the quantitative relation between the trait and one of the factors varies according to the measured value of the other factor. In traditional quantitative genetics, however, a variety-location interaction or *genotype-environment interaction* is high when the responses of the observed varieties across the range of the observed locations do not parallel one another. That is, one variety may be highest for the trait in one location, but another variety may be highest in another location-or, at least, the difference between any two varieties may change location to location. Because the traditional quantitative genetics analysis of trait variation requires no reference to measured factors, the order of the varieties (or genotypes) and locations (or environments) is arbitrary and adds no information to the analysis. Moreover, there is no reason for the relevant (but unknown) factors involved in the producing the trait to carry over from one variety-location to another. In short, the two senses of genotype-environment interaction are not linked at a conceptual or empirical level. There is no inconsistency, therefore, between claims of substantial human gene-environment interaction (for which there is an active research arena; National Institute of Environmental Health Sciences 2017), and negligible genotype-environment interaction, at least for IQ test scores (according to the conventional wisdom in human quantitative genetics; Plomin 1977, but see Taylor 2012).

The distinction between the components of partitioned variation in a trait and measurable underlying factors has relevance to the perennially reemerging two-part hypothesis: high heritability values for human IQ test scores (Neisser et al. 1996, but see Turkheimer et al. 2003, Nisbett et al. 2012) *coupled with* a failure of environmental hypotheses to account for the differences between the mean scores for racial groups

(but see Fryer and Levitt 2004) supports explanations of mean differences in terms of genetic factors (e.g., Jensen in Miele 2002, 111ff). (The specific factors would still have to be elucidated, so "support" may be read as "lends plausibility to the belief that such genetic factors exist.") Yet, if statistical analysis of variation among traits, which includes heritability estimation, provides little or no guidance in hypothesizing about measurable factors underlying the observations *within a population*, then it can provide little or no guidance about measurable factors associated with differences *between two groups*. (Strictly, differences between *the means for* the two groups. Recall the earlier remark [#5] that, when a significant result becomes the basis for practice or policy, variation around the mean gets discounted.) Moreover, *contra* Dickens and Flynn (2001), there is no paradox in finding high heritability for IQ test scores along with large differences in average score from one generation to the next (presumably unrelated to genetic changes). The average group and generational differences still need explanations, but heritability studies provide no warrant to center hypotheses about these differences around differences in measurable genetic factors.

13. Genetic diagnosis, treatment, monitoring, and surveillance

Idea: Genetic analysis has begun to identify genetic risk factors. We need to consider the social infrastructure needed to keep track of the genetic and environmental exposures with a view to useful epidemiological analysis and subsequent healthcare measures. Even in cases where the condition has a clear-cut link to a single changed gene and treatment is possible, there is complexity in sustaining that treatment.

Bowcock (2007) describes how a consortium of 50 British groups examined genetic variance in a GWA study. In the search for genetic risk factors for seven common diseases, 500,000 SNPs were examined from the genomes of 17,000 individuals. The number and scope of GWA studies continue to increase, but not so life course studies. Frank (2005) remind us that, surely environmental as well as genetic factors influence development of traits, but the cost to collect and store information about environmental exposures over the life course of individuals is much greater and it tends not to be collected. Indeed, these days, even the collection of environmental data at a

community level seems vulnerable (Paris et al. 2017). As noted earlier (#2 and 4), one-sidedness of data in turn shapes the associations or patterns that can be perceived (description) and thus the measures that can be supported by epidemiological data (prescription).

Even if the emphasis on GWA studies is accepted, standards for "presenting and interpreting cumulative evidence on gene-disease associations," as Khoury et al. (2007) point out, are needed to reduce the frequency of unreplicable associations (false positives) that might derive from publication and selection biases, differences in collection and analysis of samples, and the presence of undetected gene-environment interactions (recalling #4-6). While standards constitute infrastructure to help make research reliable, different kinds of infrastructure would be needed if it happened that a SNP loci identified by GWA studies led researchers to locate the genetic variant influencing the trait and then to identify a biochemical treatment to counter its effect. Paul's (2013) account of the history and sociology of the poster-child case for genetic medicine, phenylketonuria (PKU) makes that clear. Following routine screening of newborns and instituting of a special diet for individuals with PKU, the previous certainty of severe cognitive impairment has been replaced by a chronic disease with a new set of problems. There remains an ongoing struggle, at least in the USA, to secure health insurance coverage for the special diet and to enlist family and peers to support individuals with PKU staying on that diet through adolescence and into adulthood. For women who do not maintain the diet well and become pregnant, high levels of phenylalanine adversely affect the development of their non-PKU fetuses. This so-called maternal PKU is a public health concern that did not previously exist. Given that PKU is a simple case—a mutation in a single gene—health improvements through post-natal genetic screening can only be more complicated. What prescriptive idea, then, motivates the epidemiological search for associations between complex medical traits and variants at multiple sites on the genome (Taylor 2009)?

14. Popular epidemiology and health-based social movements

Idea: The traditional subjects of epidemiology become agents when: a. they draw attention of trained epidemiologists to fine scale patterns of disease in that community and otherwise contribute to initiation and completion of studies; b. their resilience and reorganization of their lives and communities in response to social changes displaces or complements researchers' traditional emphasis on exposures impinging on subjects; and c. when their responses to health risks displays rationalities not taken into account by epidemiologists, health educators, and policy makers.

The work of epidemiologists in looking for associations that have relevance for health-related practice and policy is complicated by their subjects becoming *agents*. For example, a. in *popular epidemiology* (Brown 2007) local residents use their experience and fine-grained knowledge to point to phenomena and categories (#2 and 4) in which to make observations and look for associations; b. when they change the social organization of their communities (Sampson 2012) thus altering the causal dynamics that researchers sought to illuminate on the basis of patterns in data (such as associations with risk factors) (#2 and 5); and c. when groups resist health promotion efforts, such as smoking-cessation programs, because of a *lay epidemiology* (Lawlor et al. 2003) in which individuals in lower SES groups assess the specific risk in relation to their wider life prospects (#2, 4, 7, 9, 10). Studies of these and other ways in which subjects become agents may well result in patterns and variation among people that do not extrapolate readily over time, place, and scale. Nevertheless such studies could still provide points of departure (see #2 and 4) for research and policy engagements in subsequent situations.

14b. Taking Stock of Course: Where have we come and what do we need to learn to go further?

Idea: In order to move ahead and continue developing, it is important to take stock of what went well and what needs further work.

If the initial session introduced various conditions that help in developing epidemiological literacy (#1), this final session allows us to plan ways to secure ongoing support beyond microcosm of the course.

From Critical Thinking to Conjecture

In descriptively teasing out what epidemiologists do in practice through the topic-by-topic presentation, I have also had a prescriptive goal: to encourage discussants to draw purposefully from across the range of topics and explore contrasting positions. What phenomena, the critical-thinking student or researcher might ask their colleagues, have been overlooked? What other ways are there to define the categories for making observations and detecting patterns? Should we be interested in screening and treatment of sick or high-risk individuals or taking population-wide measures to reduce the frequency of such individuals? How would our interpretations differ if we thought of regression equations in terms of tightness-of-packing, not goodness of prediction? Why are we focusing on factors associated with the relative risk when measures and policies already exist to reduce the major risk factors for absolute incidence? And so on, from one topic to the next reviewed in the course and this article.

It is possible that the reader or researcher disagrees with various positions or their description. No problem; the premise of critical thinking (as I define it; Taylor 2002) is that we come to understand ideas and practices better when we examine them in relation to alternatives. By extension, it does not matter if a position is currently espoused by few epidemiologists. Indeed, some positions I include because I believe that epidemiologists—and philosophers who descriptively and prescriptively discuss epidemiology—should examine them more deeply. Let me highlight two of these.

As prefigured in the introduction, discussion about causality should distinguish between, on one hand, showing a modifiable factor to have been associated with a difference in the data from past observations and, on the other hand, expecting that factor, when modified, to generate that difference going forward. This distinction applies to the

interventionist not only the statistical, differences-that-make-a-difference models of causality (see topics #5, 6, and the end of 12). More attention should also be given to the possibility of underlying heterogeneity (which informs my review of topics #4, 8, 9 and 12). That is, when similar responses of different individual types (i.e., values for the trait in question) are observed, it need not be the case that similar conjunctions of risk and protective factors have been involved in producing those responses. Epidemiology has traditionally been allied with population health and its focus on modifiable causes of disease at the population level (Davey Smith 2011); nevertheless, researchers might want to consider alternatives to treating individuals according to the average of the population or group to which they belong (as noted for racial group average differences in educational measures; see end of #5).

When are researchers *troubled* by heterogeneity? (Taylor 2014a,b) Consider this story, which concerns heterogeneity in the simplest sense, namely, a group made up of two distinguishable subgroups. At my annual physical when I turned 50 my doctor recommended a regimen of half an aspirin a day to help prevent a stroke or heart attack. Not long afterwards I learned that some fraction of the population is *resistant* to aspirin—it does not produce the desired anti-platelet effect. This subgroup is, however, still subject to aspirin resulting in an increased risk of serious gastrointestinal bleeding. Could I find out if I was in the resistant fraction? My doctor informed me that health insurance companies do not consider testing to be a justified expense for healthy subjects. It was, he advised, up to me to decide whether to take the daily aspirin. Some Internet follow-up on my part revealed that testing for resistance is possible, but is undertaken only when patients under treatment for a cardiovascular attack do not seem to be showing the anti-platelet effects of aspirin intake. Would I devote energy to find others with similar concerns about their aspirin-resistance status and agitate for access to testing? As it turned out, no—I went along with the health insurance company's determination and followed the doctor's advice to make a personal choice, in this case, *not* to take the daily pill.

With hindsight, my decision was a good one—recent research indicates that in all healthy subjects the decreased average risk of a cardiovascular event might not outweigh the increased average risk of gastrointestinal bleeding (Seshasai et al. 2012). Yet, these newer findings aside, consider my experience at the time. In the doctor's initial recommendation, aspirin-resistant and normal subgroups were treated as a single group of over-50s, all of us subject to the same positive trade-off between cardiovascular and gastrointestinal risks. The doctor could have been troubled by the heterogeneity within this group, especially after I raised my concerns. Instead he invoked both the rhetoric of patient choice and the constraints of the health insurance system. I did entertain the possibility of joining with others to agitate for testing to determine which subgroup we belonged to. In the end, I complied with my doctor's framing of my position, namely, I should see myself as a member of an over-50s group subject to a degree of uncertainty about the positive trade-off.

Three interrelated conjectures are illustrated by the story:

- *Research and application of resulting knowledge are untroubled by heterogeneity to the extent that populations are well controlled*—As the story conveys, my doctor wanted to treat me according to the average of the group I was a member of. At first I did not comply with his recommendation, but I did accept his subsequent advice.
- *Such control can be established and maintained, however, only with considerable effort or social infrastructure*—The authority of medical professionals was not sufficient to achieve my compliance, but eventually the rhetoric of patient choice and the reimbursement guidelines of the health insurance system were.
- *The interplay of heterogeneity, control, and social infrastructure provides an opening to give more attention to possibilities for participation instead of control of human subjects*—The Internet gave me a means to go beyond the consultation with my doctor. From this first port of call, I could have embarked on a journey of finding whom to collaborate with to agitate for change in the guidelines for aspirin-resistance testing.

Of course, my personal concerns about prophylactic aspirin do not constitute a key issue for epidemiology and population health. Indeed, only at the end of the course do

the conjectures surface, when attention is drawn to social infrastructure (#13) and to subjects of epidemiology becoming agents (#14). Moreover, there is ambiguity in the last topic about whether subjects as agents complicates or enables the work of epidemiologists in looking for associations that have relevance for health-related practice and policy. A bolder position would be that epidemiologists should relax their focus on modifiable causes of disease at the population level, revising the scope of epidemiology (#3) and the methods used (#4-13) to allow subjects, living in specific situations that continue to change, to show researchers how they connect knowledge with action. The patterns in variation among people derived from observations of communities where people are resilient and reorganize their health, lives, and communities in response to social changes (Sampson 2012) might not extrapolate readily over time, place, and scale, yet the patterns could provide a point of departure for research and policy engagements in subsequent situations that the researchers study. This bolder line of inquiry and conjecture, when considered in tension with the current expectations and practices of epidemiologists, might allow them—and philosophers who descriptively and prescriptively discuss epidemiology—to understand those expectations and practices better.

Acknowledgements

Some sentences have been drawn from Taylor (2014a, b). I am grateful for comments on drafts by Sam Friedman, Brian Lax, Barbara Mawn, and Karin Patzke.

References

Allchin, D. (2013). Christiaan Eijkman and the case of beriberi. In Teaching the Nature of Science: Perspectives & Resources. St. Paul, MN, SHiPS Press: 165-185.

Alter, D. A., C. D. Naylor, P. Austin and J. V. Tu (1999). "Effects of socioeconomic status on access to invasive cardiac procedures and on mortality after acute myocardial infarction." New England Journal of Medicine 341: 1359-1367.

Barker, D. J. (1971). "Buruli disease in a district of Uganda." Journal of Tropical Medicine and Hygiene 74: 260-264.

Barker, D. J. (1998). Mothers, Babies, and Health in Later Life. Edinburgh, Churchill Livingstone.

Barker, D. J. (1999). "Commentary: Intrauterine nutrition may be important." British Medical Journal 318: 1471-1480.

Barker, D. J. and C. Osmond (1986). "Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales." Lancet 1(8489): 1077-1081.

Barker, D. J. and C. Osmond (1987). "Death rates from stroke in England and Wales predicted from past maternal mortality." BMJ 295(6590): 83-86.

Ben-Shlomo, Y. and G. Davey Smith (1991). "Deprivation in infancy or in adult life: which is more important for mortality risk?" The Lancet 337(8740): 530-534.

Ben-Shlomo, Y. and D. Kuh (2002). "A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives." International Journal of Epidemiology 31: 285-293.

Bowcock, A. M. (2007). "Guilt by association." Nature 447: 645-646.

Brandt, A. M. and M. Gardner (2000). "Antagonism and accommodation: interpreting the relationship between public health and medicine in the United States during the 20th century." American Journal of Public Health 90(5): 707-715.

Brody, H., M. R. Rip, P. Vinten-Johansen, N. Paneth and S. Rachman (2000). "Map-making and myth-making in Broad Street: the London cholera epidemic, 1854." The Lancet 356: 64-68.

Brown, G. W. and T. Harris (1978). Social Origins of Depression. New York, The Free Press.

Brown, G. W. and T. O. Harris (1989). Depression. In Life Events and Illness. G. W. Brown and T. O. Harris (eds.). New York, Guilford Press: 49-93.

Brown, G. W. and T. O. Harris, Eds. (1989). Life Events and Illness. New York, Guilford Press.

Brown, P. (2007). Toxic exposures : contested illnesses and the environmental health movement. New York, Columbia University Press.

C Reactive Protein Coronary Heart Disease Genetics Collaboration (2011). "Association between C reactive protein and coronary heart disease: Mendelian

randomisation analysis based on individual participant data." British Medical Journal 342(15 February): d548.

Caspi, A., J. McClay, T. E. Moffitt, J. Mill, J. Martin, I. W. Craig, A. Taylor and R. Poulton (2002). "Role of Genotype in the Cycle of Violence in Maltreated Children." Science 297(5582): 851-854.

Cohen, J. (2014). "Mesoamerica's Mystery Killer." Science 344: 143-147.

Davey Smith, G. (2001). "The uses of Uses of Epidemiology." International Journal of Epidemiology 30: 1146-1155.

Davey Smith, G. (2007). "Life-course approaches to inequalities in adult chronic disease risk." Proceedings of the Nutrition Society 66: 216-236.

Davey Smith, G. (2011). "Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice " International Journal of Epidemiology 40: 537-562.

Davey Smith, G. and S. Ebrahim (2007). Mendelian randomization: Genetic variants as instruments for strengthening causal influences in observational studies. In Biosocial Surveys. M. Weinstein, J. W. Vaupel and K. W. Wachter (eds.). Washington, DC, National Academies Press: 336-366.

Davey Smith, G. and S. Harding (1997). "Is control at work the key to socioeconomic gradients in mortality?" Lancet 350: 1369-1370.

Davies, A., G. Davey Smith, M. T. May and B.-S. Y. (2006). "Association between birth weight and blood pressure is robust, amplifies with age, and may be underestimated." Hypertension 48: 431-436.

Dickens, W. T. and J. R. Flynn (2001). "Heritability estimates versus large environmental effects: The IQ paradox resolved." Psychological Review 108(2): 346-369.

Diez Roux, A. V. (2002). "Invited Commentary: Places, People, and Health." American Journal of Epidemiology 155: 516-519.

Ebrahim, S. and G. Davey-Smith (eds.) (2001). "Commentaries on Rose's 'Sick individuals and sick populations'." International Journal of Epidemiology **30**(3): 433-446.

Fazel, S. (2013). "Coin-flip judgement of psychopathic prisoners' risk." New Scientist.

Frank, J. (2005). "A Tale of (More Than ?) Two Cohorts – from Canada." 3rd

International Conference on Developmental Origins of Health and Disease.

Freedman, D. A. (2001). Ecological inference and the ecological fallacy. In International Encyclopedia for the Social and Behavioral Sciences. N. J. Smelser and P. B. Baltes (eds.).

Freedman, D. A. (2005). Linear statistical models for causation: A critical review. In Encyclopedia of Statistics in the Behavioral Sciences. B. Everitt and D. Howell (eds.). Hoboken, NJ, Wiley.

Friedman, S. R., B. S. West, B. Tempalski, C. M. Morton, C. M. Cleland, D. C. D. Jarlais, H. I. Hall and H. L. F. Cooper (2014). "Do metropolitan HIV epidemic histories and programs for people who inject drugs and men who have sex with men predict AIDS incidence and mortality among heterosexuals?" Annals of Epidemiology 24(4): 304-311.

Fryer, R. and S. Levitt (2004). "Understanding the black-white test score gap in the first two years of school." The Review of Economics and Statistics 86(2): 447-464.

Galton, F. (1875). "The history of twins, as a criterion of the relative powers of nature and nurture." Fraser's Magazine 12: 566-576.

Gordis, L. (2013). Epidemiology. Philadelphia, Elsevier Saunders.

Harris, T., Ed. (2000). Where Inner and Outer Worlds Meet. London, Routledge.

Hernan, M. A. (2002). "Causal Knowledge as a Prerequisite for Confounding Evaluation: An Application to Birth Defects Epidemiology." American Journal of Epidemiology 155(2): 176-184.

Hooper, L., C. D. Summerbell, R. Thompson, D. Sills, F. G. Roberts, H. Moore and G. Davey Smith (2011). "Reduced or modified dietary fat for preventing cardiovascular disease." The Cochrane Database of Systematic Reviews (7): CD002137.

Huxley, R., A. Neil and R. Collins (2002). "Unravelling the fetal origins hypothesis: is there really an inverse association between birthweight and subsequent blood pressure?" Lancet 360(9334): 659-665.

Ioannidis, J. P. A. (2005). "Why Most Published Research Findings Are False." PLoS Medicine 2(8): e124.

Jick, H., G. L. Zomberg, S. S. Jick and et al. (2000). "Statins and the risk of dementia."

Lancet 356: 1627-1631.

Kahn, J. (2007). "Race in a Bottle." Scientific American (July 15).

Kendler, K. S., C. O. Gardner and C. A. Prescott (2002). "Towards a comprehensive developmental model for major depression in women." American Journal of Psychiatry 159: 1133-1145.

Kendler, K. S. and C. A. Prescott (2006). Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Abuse Disorders. New York, The Guilford Press.

Khodarahmi, M. and L. Azadbakht (2014). "The Association Between Different Kinds of Fat Intake and Breast Cancer Risk in Women." International Journal of Preventive Medicine 5(1): 6-15.

Khoury, M. J., J. Little, M. Gwinn and J. P. Ioannidis (2007). "On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies." International Journal of Epidemiology 36: 439-445.

Krieger, N. (2010a). Social inequalities in health. Teaching Epidemiology. J. Olsen, R. Saracci and D. Trichopoulos. Oxford, Oxford University Press: 215-240.

Krieger, N. (2010b). The Science and Epidemiology of Racism and Health: Racial/Ethnic Categories, Biological Expressions of Racism, and the Embodiment of Inequality—an Ecosocial Perspective. In What's the use of race? Modern Governance and the biology of difference. I. Whitmarsh and D. S. Jones. Cambridge, MA, MIT Press: 225-255.

Krieger, N. (2011). Epidemiology and the People's Health. Theory and Context. Oxford, Oxford University Press.

Krieger, N., J. T. Chen, P. D. Waterman, D. H. Rehkopf and S. V. Subramanian (2005). "Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: The Public Health Disparities Geocoding Projec." American Journal of Public Health 95: 312-323.

Krieger, N. and G. Davey Smith (2016). "The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology." International Journal of Epidemiology 45(6): 1787-1808.

Krieger, N., A. Kosheleva, P. D. Waterman, J. T. Chen, J. Beckfield and M. V. Kiang

(2014). "50-year trends in US socioeconomic inequalities in health: US-born Black and White Americans, 1959–2008 " International Journal of Epidemiology 43(4): 1294-1313.

Ku, C. S., E. Y. Loy, Y. Pawitan and K. S. Chia (2010). "The pursuit of genome-wide association studies: where are we now?" Journal of Human Genetics 55(April): 195-206.

Kuh, D. and Y. Ben-Shlomo, Eds. (2004). A Life Course Approach to Chronic Disease Epidemiology. Oxford, Oxford University Press.

Lagakos, S. W. (2006). "The challenge of subgroup analysis--Reporting without distorting." New England Journal of Medicine 354: 1667-1669.

Lawlor, D. A., S. Frankel, M. Shaw, S. Ebrahim and G. Davey Smith (2003). "Smoking and Ill Health: Does Lay Epidemiology Explain the Failure of Smoking Cessation Programs Among Deprived Populations?" American Journal of Public Health 93(2): 266-270.

Lawlor, D. A., G. Davey Smith, K. R. Bruckdorfer, D. Kundu and S. Ebrahim (2004). "Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence?" The Lancet 363: 1724-1726.

Lynch, J., G. Davey Smith, S. Harper and K. Bainbridge (2006). "Explaining the social gradient in coronary heart disease: comparing relative and absolute risk approaches." Journal of Epidemiology and Community Health 60: 436-441

Lynch, M. and B. Walsh (1998). Genetics and Analysis of Quantitative Traits. Sunderland, MA, Sinauer.

Mackay, T. F. C., E. A. Stone and J. F. Ayroles (2009). "The genetics of quantitative traits: challenges and prospects." Nature Reviews Genetics 10: 565-577.

Majumder, P. P. and S. Ghosh (2005). "Mapping quantitative trait loci in humans: achievements and limitations." Journal of Clinical Investigation 115(6): 1419-1424.

McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. A. Ioannidis and J. N. Hirschhorn (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges." Nature Reviews Genetics 9(May): 356-369.

McMichael, A. J. (2011). "Review of "Epidemiology and the People's Health. Theory and Context. Nancy Krieger"." International Journal of Epidemiology 40: 1130–1132.

Miele, F. (2002). Intelligence, Race, and Genetics: Conversations with Arthur Jensen. Boulder, CO, Westview Press.

Moffitt, T. E., A. Caspi and M. Rutter (2005). "Strategy for investigating interactions between measured genes and measured environments." Archives of General Psychiatry 62(5): 473-481.

Morris, J. N. (1957). Uses of epidemiology. Edinburgh, E. & S. Livingstone.

Mustillo, S. A., N. Krieger, E. P. Gunderson, S. Sidney, H. McCreath and C. I. Kiefe (2004). "The Association of Self-Reported Experiences of Racial Discrimination with Black/White Differences in Preterm Delivery and Low Birth Weight: The CARDIA Study." American Journal Public Health 94: 2125-2131.

National Institute of Environmental Health Sciences (2017). "Gene-Environment Interaction." (viewed 24 October 2017).

Neisser, U., G. Boodoo, T. J. Bouchard, A. W. Boykin, N. Brody, S. J. Ceci, D. F. Halpern, J. C. Loehlin, R. Perloff, R. J. Sternberg and S. Urbina (1996). "Intelligence: Knowns and unknowns." American Psychologist 51: 77-101.

Nisbett, R. E. (1998). Race, genetics, and IQ. The Black-White Test Score Gap. C. Jencks and M. Phillips. Washington, DC, Brookings Institution Press**:** 86-102.

Nuffield Council on Bioethics (2002). "Genetics and Human Behavior: The Ethical Context." http://www.nuffieldbioethics.org viewed 22 June 2007.

Oakes, J. M. (2004). "The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology." Social Science & Medicine 58: 1929-1952.

Ou, S.-R. (2005). "Pathways of long-term effects of an early intervention program on educational attainment: Findings from the Chicago longitudinal study." Applied Developmental Psychology 26: 478-611.

Oxford, J. S., R. Lambkin, A. Sefton, R. Daniels, A. Elliot, R. Brown and D. Gill (2005). " A hypothesis: the conjunction of soldiers, gas, pigs, ducks, geese and horses in northern France during the Great War provided the conditions for the emergence of the "Spanish" influenza pandemic of 1918-1919." Vaccine 23(7): 940-945.

Paris, B. S., L. Dillon, J. Pierre, I. V. Pasquetto, E. Marquez, S. Wylie, M. Murphy, P. Brown, R. Lave, C. Sellers, B. Mansfield, L. Fredrickson, N. Shapiro and EDGI

(2017). Pursuing a toxic agenda: Environmental Injustice in the Early Trump Administration.

Paul, D. B. and J. P. Brosco (2013). The PKU Paradox: A Short History of a Genetic Disease. Baltimore, Johns Hopkins University.

Pearce, N. (1996). "Traditional epidemiology, modern epidemiology, and public health." American Journal of Public Health 86(5): 678-683.

Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge, Cambridge University Press.

Petitti, D. (2004). "Commentary: Hormone replacement therapy and coronary heart disease: four lessons." International Journal of Epidemiology 33(3): 461-463.

Plomin, R., J. C. DeFries and J. C. Loehlin (1977). "Genotype-environment interaction correlation in analysis of human behavior." Psychological Bulletin 84(2): 309-322.

Poland, J. (2004). Bias and Schizophrenia. Bias in Psychiatric Diagnosis. P. Caplan and L. Cosgrove. Lanham, MD, Rowman & Littlefield: 149-162.

Putnam, S. and S. Galea (2008). "Epidemiology and the Macrosocial Determinants of Health." Journal of Public Health Policy 29: 275–289.

Regan, M. M. and R. D. Gelber (2005). "Predicting response to systematic treatments: Learning from the past to plan for the future." The Breast 14: 582-593.

Ridker, P. M., J. E. Buring, N. Rifai and N. R. Cook (2007). "Development and Validation of Improved Algorithms for the Assessment of Global Cardiovascular Risk in Women: The Reynolds Risk Score." Journal of the American Medical Association 297: 611-619.

Rose, G. (1985). "Sick individuals and sick populations." International Journal of Epidemiology 14: 32-38.

Rothman, K. J., S. Greenland and T. L. Lash (2012). Modern Epidemiology. Boston, Lippincott Wolter.

Sampson, R. J. (2012). Great American City: Chicago and the Enduring Neighborhood Effect. Chicago, Chicago University Press.

Seshasai, S. R. K., S. Wijesuriya, R. Sivakumaran, S. Nethercott, S. Erqou, N. Sattar and K. K. Ray (2012). "Effect of Aspirin on Vascular and Nonvascular Outcomes: Meta-analysis of Randomized Controlled Trials." Archives of Internal Medicine

172(3): 209-216.

Stegenga, J. (2009). "Marcel Weber: Philosophy of Experimental Biology." Erkenntnis **71**(3): 431-436.

Steinbach, R., J. Green, M. G. Kenward and P. Edwards (2014). "Is ethnic density associated with risk of child pedestrian injury? A comparison of inter-census changes in ethnic populations and injury rates." Ethnicity & Health.

Taylor, P. J. (2002) "We know more than we are, at first, prepared to acknowledge: Journeying to develop critical thinking." Working Papers in Critical, Creative and Reflective Practice, http://scholarworks.umb.edu/cct_ccrp/1 (viewed 15 November 2017).

Taylor, P. J. (2005). Unruly Complexity: Ecology, Interpretation, Engagement. Chicago, University of Chicago Press.

Taylor, P. J. (2008). "Why was Galton so concerned about “regression to the mean”?—A contribution to interpreting and changing science and society." DataCrítica 2(2): 3-22.

Taylor, P. J. (2009). "Infrastructure and Scaffolding: Interpretation and Change of Research Involving Human Genetic Information." Science as Culture 18(4): 435-459.

Taylor, P. J. (2012). "A gene-free formulation of classical quantitative genetics used to examine results and interpretations under three standard assumptions." Acta Biotheoretica 60(4): 357-378.

Taylor, P. J. (2014a) "Heterogeneity, not randomness, sets challenges for quantitative genetics and epidemiology: A response to Davey Smith’s “gloomy prospect”." Working Papers on Science in a Changing World, http://scholarworks.umb.edu/cct_sicw/2 (viewed 15 November 2017).

Taylor, P. J. (2014b). Nature-Nurture? No: Moving the Sciences of Variation and Heredity Beyond the Gaps. Arlington, MA, The Pumping Station.

Taylor, P. J. (2015). "Distinctions that make a difference? (An essay review of Beyond versus: The struggle to understand the interaction of nature and nurture by J. Tabery)." Studies in History and Philosophy of Biological and Biomedical Sciences, Part C 51: 70-76.

Turkheimer, E., A. Haley, M. Waldron, B. D'Onofrio and I. I. Gottesman (2003). "Socioeconomic Status Modifies Heritability of IQ in Young Children." Psychological Science 16(6): 623-628.

Weldon, K. L. (2000). "A Simplified Introduction to Correlation and Regression." Journal of Statistics Education 8(3).

Wright, R. J., H. Mitchell, C. M. Visness, S. Cohen, J. Stout, R. Evans and D. R. Gold (2004). "Community Violence and Asthma Morbidity: The Inner-City Asthma Study " American Journal of Public Health 94: 625-632.

Zandi, P. P., D. L. Sparks, A. S. Khachaturian, J. Tschanz, M. Norton, M. Steinberg, K. A. Welsh-Bohmer, J. C. Breitner and Cache County Study investigators (2005). "Do statins reduce risk of incident dementia and Alzheimer disease? The Cache County Study." Archives of General Psychiatry 62(2): 217-224.

Zuka, O., E. Hechtera, S. R. Sunyaev and E. S. Lander (2012). "The mystery of missing heritability: Genetic interactions create phantom heritability." PNAS 109(4): 1193–1198.