Synthese
https://doi.org/10.1007/s11229-018-01960-6

S.I.: PHILOSOPHY OF EPIDEMIOLOGY

Critical epidemiological literacy: understanding ideas better when placed in relation to alternatives

Peter J. Taylor[1]

Received: 17 November 2017 / Accepted: 21 September 2018

Abstract

This article describes contrasting ideas for a set of topics in epidemiological thinking. The premise underlying this contribution to the special edition is that researchers develop their epidemiological thinking 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. I encourage discussants from philosophy and epidemiology to draw purposefully from across a 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*. After an initial topic concerning practices for developing epidemiological literacy, 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 conceptual 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. The organization of topics derives from a graduate course that I teach that aims for epidemiological literacy, not technical ability in statistical formulas and data analysis, and shares the underlying premise and critical thinking goals of this article. During the topic-by-topic 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 concludes with a "Limitations of this Study" section that teases out different kinds of description–prescription relationship that are implied in undertaking philosophy of epidemiology and identifies some other considerations that are implied but not emphasized by this article.

✉ Peter J. Taylor
peter.taylor@umb.edu

[1] Science in a Changing World Graduate Track, University of Massachusetts, Boston, MA 02125, USA

Published online: 01 October 2018

Keywords Causality · Critical thinking · Description–prescription · Heterogeneity · Inequality · Intervention

1 Introduction

How do epidemiologists analyze data from populations with a view to identifying the biological and social influences on the development of diseases and behaviors? The premise motivating this contribution is that researchers develop their *epidemiological thinking* over time through interactions with other researchers who have a variety of in-practice commitments, such as to kinds of cases and to methods of analysis, and not simply to a philosophical framework for explanation. In such interactions it should help to be able to draw purposefully from across a range of topics and contrasting positions. The aim of this article, therefore, is to stimulate discussants from philosophy and epidemiology to consider the various positions they are taking in relation to alternatives.

It is beyond the scope of this article to assemble observations that support the premise or to relate my relevant experience as I learned epidemiology mid-career. The range of considerations introduced in the topic-by-topic presentation of Sect. 2 should, however, render the premise plausible. The detail of Sect. 2 is intentional—the quantity and variety of the contrasts is meant to keep readers' attention on the multiplicity of positions that epidemiologists take, explicitly or implicitly, and to resist any expectations that philosophical articles should tie points into a focused argument, say, for a certain explanatory strategy and against competing arguments.

The article, as its title indicates, concerns *critical thinking* construed as *understanding ideas and practices better when we examine them in relation to alternatives* (Taylor 2002). If fostering critical thinking by describing contrasting positions seems a modest goal, let me acknowledge two secondary expository goals: (a) to illustrate a *description–prescription relationship* (Stegenga 2009) that runs through epidemiology, namely, 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? and (b) to set the scene for Sect. 3, "Limitations of this study," which identifies some considerations about undertaking philosophy of epidemiology that are implied but not emphasized by this article. Beyond the article's scope, somewhat ironically, is examining the overall approach taken here in relation to alternatives. Readers' critical thinking about this contribution could hardly do better than consider the alternative approach taken by Krieger and Davey Smith (2016) (see Sect. 3.1).

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.

Table 1 Sequence of topics

Setting the conceptual scene
Practices for developing epidemiological literacy
Phenomena: Exploring the natural history of disease
The scope and challenges of epidemiology
Categories
Associations, predictions, causes, and interventions
Confounders and conditioning of analyses
Complexities of inequalities within and between populations, context, and changes over the life course
Variations in health care
Heterogeneity within populations; subgroups
Placing individuals in a multileveled context
Life course epidemiology
Multivariable "structural" models of development
Heritability, heterogeneity, and group differences
Genetic diagnosis, treatment, monitoring, and surveillance
Popular epidemiology and health-based social movements

2 Topics in epidemiological thinking and population health

The organization of topics derives from a graduate course, *Epidemiological Thinking and Population Health*, that I teach to students from public policy, nursing, gerontology, and science studies. The course aims not for technical ability in data analysis and use of statistical formulas, but for *epidemiological literacy*. That is, students should understand the contrasting positions for each topic well enough to communicate and collaborate thoughtfully with specialists, in particular, to identify the positions the specialists are taking and probe their thinking about what would be entailed to adopt or consider alternatives. This pedagogical goal parallels the aim of this article; indeed, the course and the article are motivated by the same premise.

The sequence of topics is previewed in Table 1. The initial topic is included to match the set-up session in the course. After that, a number of conceptual steps follow, starting with characterization of the very phenomena we might be concerned with, leading up to the making and confounding of explanations and causal claims. Building on that conceptual basis, the remaining topics address the complexities of inequalities within and between populations, context, and changes over the life course. (This set of topics also matches the social policy orientation of the course mentioned above.) The topics as a whole are ordered so that the issues and angles of analysis in earlier ones lay the basis for discussion of topics that come later.

For each topic I present an idea (or ideas), then (with the exception of the first topic) elaborate on the contrasts stated in, or that follow from, those ideas. Diverse kinds of points are made, with a view, as noted in the Introduction, to keeping readers' attention on the many positions they take and countering expectations that points

should get tied into a case for some explanatory strategy, as well as to illustrating the description–prescription relationship running through epidemiology and setting the scene for Sect. 3. (The range of points also matches those that I see the need to cover in the epidemiological literacy course.)

(Pedagogical note: The origins of this article in a course makes students and educators very welcome as readers, but the article cannot provide the detail needed for students to get the effect of taking the course or educators to be able to teach it themselves. 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://www.faculty.umb.edu/pjt/epi.)

2.1 Practices for developing epidemiological literacy

Idea 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.

2.2 Phenomena: exploring the natural history of disease

Idea Detailed observation (as 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.

In standard epidemiology texts analysis of data enters quickly, whether the books 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, or 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 then hypothesizing about the causes (Brody et al. 2000). In short, defining phenomena is not a simple matter of induction; this raises the perennial question for philosophy of science of where hypotheses that get assessed by research come from in the first place. That 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 et al. (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.

2.3 The scope and challenges of epidemiology

Idea 3a The uses of epidemiology are many, but shift over time, and are subject to recurrent challenges from inside and outside the field.

Idea 3b 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 epidemiology in the sense discussed in this article, namely, the "systematic approach to the population aspects of non-communicable disease" (Davey Smith 2001). In identifying seven uses of epidemiology, Morris also invites us to consider whether epidemiology is a single thing to examine and whether the best focus for philosophical attention is the currently dominant approach, namely, Morris's 7th, "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."

Brandt and Gardner's (2000) historical account shows that physicians have often opposed an increasing role for public health and, by extension, for epidemiology. Epidemiology might be valued for quantitative assessment of new interventions and evaluating patient safety and healthcare quality (fitting under Morris's 3rd use: studying the workings of health services). Its role beyond evaluation and assessment, however, 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. It may often be possible for a person to drive home even after drinking too much, but we also know that a substantial fraction of people in road accidents have high alcohol levels. Even though some people seem more susceptible than others to having their judgement and reaction times impaired by alcohol, drink-don't-drive campaigns are directed at everyone; they are population-wide measures. In contrast to such a "sick-population" approach, the "sick-individual" approach begins by assessing an individual's risk, in this case of alcohol-related accidents. A risk formula could factor in not only the proximate alcohol consumption, but also, say, visual acuity, gender, age, presence of 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. Of course, as Rose would point out, 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. Rose's "sick individuals-sick populations" contrast needs, of course, to be supplemented by practical considerations. Would the benefits minus costs of screening for high-risk drinker-driver individuals be significant relative to that from population-wide drink-don't-drive efforts? Conversely, as a political or sociological matter, would campaigns directed at the population as a whole ever get carried through to the point of eliminating all driving after drinking?

Returning to challenges to the uses of epidemiology (Idea 3a), 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. This contrast in description parallels a contrast in prescription: political engagement to change the macro-factors versus physician or patient responsibility in relation to an individual's modifiable risk factors (Putnam and Galea 2008; Krieger 2011).

2.4 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 idea and questions above extend the theme that epidemiology does not begin with data sets to analyze (Sect. 2.2). The definition of categories shapes 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 varied kinds, he decided not to record "those that were imposed by the circumstances of [people's] lives" and focus on the "effects of tendencies received at birth" (Galton 1875, p. 566). The patterns of similarity he detected among relatives may have been sound, but they allowed only for hypotheses 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 investigations to expose the genes that influence so-called schizophrenia (see Sects. 2.12, 2.13).

Demarcating categories is one link in the chain of steps in scientific inquiry—from all possible phenomena that could be inquired into, through observations made using the chosen categories, to actions supported by predictions or to causal claims. Teasing out the assumptions at each step is obviously a matter for philosophy of science. Because the assumptions are not always dictated by the phenomena or justified by the results, there is room for attention to the negotiations and wider influences that shape how the steps end up being made (Taylor 2005, pp. 33–46, 2008).

Let me note three specific category choices in epidemiology that have prescriptive implications. First, use of the category 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, focus on the absolute incidence of an illness versus 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 Sects. 2.6, 2.7 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 5 mm Hg) animates various disputes in epidemiology about prescriptively relevant associations (Huxley et al. 2002; see Sect. 2.10).

2.5 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 then 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 affecting which elderly patients in the original study had been prescribed statins, that is, patients with yet-to-diagnosed dementia were less likely to receive treatment. RCTs provide an even stronger check on results from observational studies (Lawlor et al. 2004), as illustrated when the Women's Health Initiative clinical trial reported that hormone therapy increased rather than decreased, as had been previously claimed, the risk of coronary heart disease in women.

The use of RCTs incorporates what we might call an *interventionist* model of causality. That is, of the many factors possibly associated with the outcome of interest, it is possible to intervene to modify one factor in members of 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 intervene and modify the factor going forward. 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. For example, a small subset of people has a genetic variant that leads to life-long elevated c-Reactive Protein (CRP) levels, but otherwise vary randomly on other risk factors for CHD (such as smoking, body-

mass index, and blood pressure). Such studies then examine whether the association between levels of CRP in the blood and coronary heart disease (CHD) holds for this subset. (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 *risk factor* for variables associated with an outcome of interest. 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 modifying those factors, such as age or gender, is not possible, 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.

In saying "not surprisingly" I am going along with the expectation that 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 contrasting 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 are 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 of 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. Suppose, say, the genetic variant inducing lifelong elevated CRP levels had been associated with CHD. Reducing CRP for future patients then would have to be by means other than giving them the rare genetic variant at conception.)

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 individuals to take tablets, 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 so-called compliance whenever 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? If we reject group-specific treatment, we might go on to ponder 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; also Sects. 2.8, 2.12, and 3.2.]

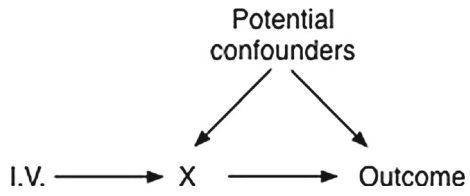
2.6 Confounders and conditioning of analyses

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?

To take conditionality into account in analysis of associations is, at one level, a descriptive matter. Envisage the observations as divided into slices, each slice containing only observations that share the same value (or limited range of values) 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). One could conduct a separate analysis on each slice, that is, *conditional* on the value of the variable defining the slice. Typically, however, the statistical analysis averages those separate comparisons into a single conditioned or *adjusted* comparison. In effect, the analysis is run on an adjusted data set in which the number of observations in each slice has been balanced out.

Such conditioning of analyses becomes a prescriptive matter once there are disputes around adjustments not made or inappropriate adjustments. 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. For example, the original association between hormone replacement therapy and lowered CHD incidence 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 and Harding (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 job control) diverts the focus of health promotion away from the conventional risk factors (i.e., 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* incidence) 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 Sect. 2.10); but see Davies et al. 2006.]

Under the interventionist model of causality, it is clear how to resolve disputes around adjustments not made or inappropriate adjustments. RCTs and Mendelian randomization demonstrate that a variable is a confounder if its association with the outcome of interest disappears when it is modified while all other relevant factors vary randomly across all subjects (where modified means takes one value in a subset of subjects and another in the rest). In the schema below, this test amounts to there being an *instrumental variable* (I.V.)—the presence of the drug being tested versus 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 arrows are used to represent both the effect of a modifiable factor (I.V. ->X) and statistical associations. When for the broader class of statistical, differences-that-make-a-difference analyses such diagrams are examined to decide whether adjustment is appropriate, doing so entails bringing in qualitative, a priori, subject-matter knowledge of causal connections (Hernan 2002). Whether, in practice, this extension beyond RCTs and Mendelian randomization (i.e., to situations where there is no obvious instrumental variable) discounts the kinds of issues about causality mentioned in Sect. 2.5, 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. For example, the use of the Reynolds Risk Score (Ridker et al. 2007; Sect. 2.5) in effect separates the slices that have high levels of CRP from those that have low levels; it has the potential to improve the assignment of people to high, medium, or low risk for CHD and thus to make preventative measures more effective.

Building on the preceding topics, the remaining topics 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. As mentioned at the start of Sect. 2, the topics reflect the social policy orientation of the course from which this article derives. They are ordered so that the issues and angles of analysis in earlier ones lay the basis for discussion of topics that come later.

2.7 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 (Krieger 2010a). 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) does, however, leave as a separate matter the *how* and *by whom* of the health-income improvement.

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 with income reported experience of racial discrimination, after adjusting not only for income but also for alcohol and tobacco use, depression, and education.

As noted earlier (Sects. 2.4, 2.6), 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. A logical extension of their argument would be to question the value of research to pin down risk factors for any gradients in health that remain after adjusting for SES. Addressing the major risk factor—lower SES—should be the priority (see, e.g., Krieger et 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 given the political economic changes over the last 40 years that continue to *enlarge* such inequalities. For both sides of this argument then, prescriptive assumptions shape the descriptive exercise of focusing on finding statistical associations. Similarly, even if research pinned down risk factors for gradients in health that remain 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 (Sect. 2.4). For example, (a) the analysis of Krieger et al. (2005) responds 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; (b) Krieger (2010b) acknowledges the social and historically changing definitions of race and ethnicity as well as the necessity of employing them in order to expose the ways that "racism harms health"; (c) Krieger et al. (2014) notes that, even with the increase in studies that include experience of discrimination as a risk factor, the emphasis remains on person–person discrimination, not structural.

2.8 Heterogeneity within populations; 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, they could have been 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 then 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 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 something even more heterogeneous. Heterogeneity can have health care implications. A few examples: 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." Fazel's (2013) review of instruments used for making decisions about sentencing, release or preventative detention in the criminal justice system (widening here what comes under the umbrella of epidemiology) 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 (as well as against putting much hope in personalized medicine). Considerable randomness at the individual level means, in his view, 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 associations with risk factors and other patterns in data, 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 Sects. 2.10, 3.2).

2.9 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 Sect. 2.10).

Scrutiny of suggestions is also needed when the aggregate-level variables have no equivalent for individuals. 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 opposite turns out to be the case.

Finally, when individual-level associations are not as clear as associations for aggregate-level variables, it may be worth scrutinizing whether the latter subsume a heterogeneity of conditions (Sect. 2.8) 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 paragraphs, alternatives to the obvious inferences from individuals to aggregate units could be considered.

Hierarchical linear modeling addresses 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, even, "complex causal chains with feedback loops and reciprocal effects" (Diez Roux 2002, p. 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. They are 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 conclude that more research is needed 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.

2.10 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 for an association between a disease in later life and conditions during gestation or early life to be isolated from similar conditions persisting 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). For them, the fetal origins hypothesis had the potential to distract attention from life-extending changes in adult life, such as smoking cessation and cholesterol-lowering use of statins. Yet, suppose that transitions across generations (e.g., rural to urban migration, public health measures, nutritional improvements) influence the nutrition that mothers are able to provide their fetuses as well as the subsequent conditions for the offspring. If such transitions could be shown to be associated with the rise and subsequent decline in CHD incidence in a country (Barker and Osmond 1987; Barker 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 is 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, can it be established 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. As discussed in Taylor (2014a, b), the method combines 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 (Sect. 2.4)—differences that Brown and Harris's method accommodates. 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 (Sects. 2.4 and 2.8). For example, in the earliest work of Brown and Harris concerning a district of London in the early 1970s, four factors were identified 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 and Harris 1978, 1989b) (recalling Sect. 2.9). 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 and Harris 1989a, p. x) could be taken up more widely in life course epidemiology (Brown and Harris 1989b).

2.11 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. In Kendler et al. (2002), for example, data on over 1900 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 Sect. 2.12). 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 indirect, i.e., mediated through other factors (Lynch and Walsh 1998, p. 823). Kendler et al.'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. However, having paths pointed in one direction and calling the networks of linked factors "structural"—or my describing the picture of development in Kendler et al.'s model as "plausible"—suggests stronger causal claims. However, where Pearl (2000, pp. 135 and 344–345) 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 Sect. 2.5). Freedman's skepticism may be seen to temper the call of Diez Roux (2002, p. 516) (noted in Sect. 2.9) for more attention by epidemiologists to "complex causal chains with feedback loops and reciprocal effects."

Kendler et al. (2002, p. 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 (Sects. 2.2, 2.3), but sensitivity of the analysis to inclusion or exclusion of such factors warrants attention given the potential prescriptive implications (Sect. 2.12).

2.12 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 four components: (a) between the means for each variety when averaged across locations; (b) between the means for each location when averaged across varieties; (c) between the means for each variety–location combination when averaged across plots (and after taking out a and b); (d) what is left over or *residual*. 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 Sect. 2.11.)

The contingent, descriptive quality of partitioning of variation becomes harder to keep in mind, 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. However, given that the partitioning is of variation in *traits*, it must be possible to partition variation without using models of unobservable genes and their hypothetical effects (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 hypothetical genes and their effects 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 epidemiologists typically do, 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 (Sect. 2.8), in the following way. Consider how 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), for example, can 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 by parents, teachers, and so on more similarly than are dizygotic twins). Yet, 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 only 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 population attributable fraction (Sect. 2.7) 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 are 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 Sect. 2.13).

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 et al. 1977; but see Taylor 2012).

The distinction between the components of partitioned variation in a trait and analyses involving measurable factors has relevance to the perennially reemerging two-part hypothesis (Taylor 2014b): 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 better read as "lends plausibility to the belief that such genetic factors exist.") Yet, given that 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, which takes us back to the earlier remark (Sect. 2.5) 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). Granted, 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.

2.13 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 Genome-Wide Association (GWA) study. In the search for genetic risk factors for seven common diseases, 500,000 Single-Nucleotide Polymorphisms (SNPs) were examined from the genomes of 17,000 individuals. The number and scope of GWA studies continue to increase. Frank (2005) reminds 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 so 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 (Sects. 2.2, 2.4), 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, infrastructure to help make research reliable is needed in the form of standards for "presenting and interpreting cumulative evidence on gene-disease associations," as Khoury et al. (2007) point out, in order to reduce the frequency of un-replicable 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 Sects. 2.4–2.6). Other kinds of infrastructure would be needed if it happened that an 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 need for infrastructure 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 to secure health insurance coverage for the special diet, at least in the USA, 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. Now, PKU is a simple case—a mutation in a single gene—which

implies that translating post-natal genetic screening into health improvements at a population level can only require more elaborate infrastructure (Taylor 2009).

2.14 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 (Sects. 2.2 and 2.4) in which to make observations and to look for associations; (b) people 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) (Sects. 2.2 and 2.5); and (c) 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 (Sects. 2.2, 2.4, 2.7, 2.9, 2.10).

By teasing out epidemiological thinking through the topic-by-topic presentation, my primary aim has been to encourage discussants from epidemiology and philosophy of epidemiology 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.

3 Limitations of this study

When biomedical journals require a "Limitations of this Study" section, they are allowing authors to acknowledge additional work needed to strengthen their findings or make them more general. In this spirit, this section identifies some considerations about undertaking philosophy of epidemiology that are implied but not emphasized by this article.

3.1 Integration of different description–prescription relationships

The previous section's review of contrasting positions for a sequence of topics can be seen as one way to address a set of underlying description–prescription relationships that invite more attention than could be given here: (a) 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? (b) The latter effort implies that the views or practices of scientists can be improved. By what means then do philosophers envisage that their accounts can influence researchers? (c) Whether the accounts made by philosophers are 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? These different kinds of description–prescription relationship are all implicated in doing philosophy of epidemiology, as is the description–prescription relationship identified in the introduction, namely, the two-way relationship between patterns that epidemiologists detect in relation to illness measures (description) and action to change those measures (prescription). Further work is entailed to integrate the different description–prescription relationships into a unified framework. That would be also the case for approaches that differ from mine, such as that of epidemiologists Krieger and Davey Smith (cited often under the topics in Sect. 2). (Striking a pragmatic balance between referring to what epidemiologists do and what they need to clarify or do differently, those two researchers end up with a position aligned with philosopher of science Lipton's *inference to the best explanation*; Krieger and Davey Smith 2016.)

3.2 Critical thinking: two implied positions

I can imagine epidemiological or philosophical readers who disagree with some of the positions or their description in Sect. 2. The existence of such disagreement does not, however, count as a limitation of the study given the critical thinking premise of the article—we come to understand ideas and practices better when we examine them in relation to alternatives. Such critical thinking can be stimulated even by positions that are currently espoused by few epidemiologists. Nevertheless, the goal of this final section—inviting more attention than is possible within the scope of this article—will be served if I end by noting two positions implied by my review of critical thinking themes.

1. 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, holding the expectation that that factor, when modified, will generate that difference going forward. This distinction might seem more obviously applicable to the statistical, differences-that-make-a-difference model of causality, but it also applies to the interventionist model (Sects. 2.5, 2.6, and the end of 2.12).
2. More attention should also be given to the *possibility of underlying heterogeneity* (which informs my review of topics 2.4, 2.8, 2.9 and 2.12). 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 Sect. 2.5 and Taylor (2014a, b) [especially Part III]). Examining when researchers are and are not *troubled* by heterogeneity motivates my ongoing inquiries in epidemiological thinking (Taylor 2011). I invite other epidemiologists—and philosophers who descriptively and prescriptively discuss epidemiology—to join in examining this area further.

Acknowledgements I am grateful for comments on drafts by Sam Friedman, Brian Lax, Barbara Mawn, Karin Patzke, three anonymous reviewers, and the editors. Research underlying this article was, in part, supported by the National Science Foundation under Grant No. SES-0327696.

References

Allchin, D. (Ed.). (2013). Christiaan Eijkman and the case of beriberi. In *Teaching the nature of science: Perspectives & resources* (pp. 165–185). St. Paul, MN: SHiPS Press.

Alter, D. A., Naylor, C. D., Austin, P., & Tu, J. V. (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., & Osmond, C. (1986). Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet, 1*(8489), 1077–1081.

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

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

Ben-Shlomo, Y., & Kuh, D. (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., & Gardner, M. (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., Rip, M. R., Vinten-Johansen, P., Paneth, N., & Rachman, S. (2000). Map-making and myth-making in Broad Street: the London cholera epidemic, 1854. *The Lancet, 356,* 64–68.

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

Brown, G. W., & Harris, T. (1978). *Social origins of depression*. New York: The Free Press.

Brown, G. W., & Harris, T. O. (1989a). Depression. In G. W. Brown & T. O. Harris (Eds.), *Life events and illness* (pp. 49–93). New York: Guilford Press.

Brown, G. W., & Harris, T. O. (Eds.). (1989b). *Life events and illness*. New York: Guilford 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., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., et al. (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., & Ebrahim, S. (2007). Mendelian randomization: Genetic variants as instruments for strengthening causal influences in observational studies. In M. Weinstein, J. W. Vaupel, & K. W. Wachter (Eds.), *Biosocial surveys* (pp. 336–366). Washington, DC: National Academies Press.
Davey Smith, G., & Harding, S. (1997). Is control at work the key to socioeconomic gradients in mortality? *Lancet, 350,* 1369–1370.
Davies, A., Davey Smith, G., May, M. T., & Ben-Shlomo, 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., & Flynn, J. R. (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., & Davey-Smith, G. (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, 2946*. http://bit.ly/NS2946. Accessed 30 September 2018.
Frank, J. (2005). A tale of (more than?) two cohorts: From Canada. In *3rd International conference on developmental origins of health and disease*.
Freedman, D. A. (2001). Ecological inference and the ecological fallacy. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia for the social and behavioral sciences* (pp. 4027–4030). Oxford: Pergamon.
Freedman, D. A. (2005). Linear statistical models for causation: A critical review. In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in the behavioral sciences* (pp. 1061–1073). Hoboken, NJ: Wiley.
Friedman, S. R., West, B. S., Tempalski, B., Morton, C. M., Cleland, C. M., Jarlais, D. C. D., et al. (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., & Levitt, S. (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.
Huxley, R., Neil, A., & Collins, R. (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., Zomberg, G. L., Jick, S. S., et al. (2000). Statins and the risk of dementia. *Lancet, 356,* 1627–1631.
Kahn, J. (2007). Race in a bottle. *Scientific American, 297*(2), 40–45.
Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2002). Towards a comprehensive developmental model for major depression in women. *American Journal of Psychiatry, 159,* 1133–1145.
Khodarahmi, M., & Azadbakht, L. (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., Little, J., Gwinn, M., & Ioannidis, J. P. (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. In J. Olsen, R. Saracci, & D. Trichopoulos (Eds.), *Teaching epidemiology* (pp. 215–240). Oxford: Oxford University Press.

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 I. Whitmarsh & D. S. Jones (Eds.), *What's the use of race? Modern Governance and the biology of difference* (pp. 225–255). Cambridge, MA: MIT Press.

Krieger, N. (2011). *Epidemiology and the people's health. Theory and context*. Oxford: Oxford University Press.

Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., & Subramanian, S. V. (2005). Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: The public health disparities geocoding project. *American Journal of Public Health, 95,* 312–323.

Krieger, N., & Davey Smith, G. (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., Kosheleva, A., Waterman, P. D., Chen, J. T., Beckfield, J., & Kiang, M. V. (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., Loy, E. Y., Pawitan, Y., & Chia, K. S. (2010). The pursuit of genome-wide association studies: Where are we now? *Journal of Human Genetics, 55*(April), 195–206.

Lagakos, S. W. (2006). The challenge of subgroup analysis—Reporting without distorting. *New England Journal of Medicine, 354,* 1667–1669.

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

Lawlor, D. A., Frankel, S., Shaw, M., Ebrahim, S., & Davey Smith, G. (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.

Lynch, J., Davey Smith, G., Harper, S., & Bainbridge, K. (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., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer.

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

Majumder, P. P., & Ghosh, S. (2005). Mapping quantitative trait loci in humans: Achievements and limitations. *Journal of Clinical Investigation, 115*(6), 1419–1424.

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

Miele, F. (2002). *Intelligence, race, and genetics: Conversations with Arthur Jensen*. Boulder, CO: Westview Press.

Moffitt, T. E., Caspi, A., & Rutter, M. (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., Krieger, N., Gunderson, E. P., Sidney, S., McCreath, H., & Kiefe, C. I. (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. Accessed 24 October 2017.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer. E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67*(2), 130–159.

Nuffield Council on Bioethics. (2002). Genetics and human behavior: The ethical context. http://www.nuffieldbioethics.org. Accessed 22 June 2007.

Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology. *Social Science and 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., Lambkin, R., Sefton, A., Daniels, R., Elliot, A., Brown, R., et al. (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., Dillon, L., Pierre, J., Pasquetto, I. V., Marquez, E., Wylie, S., Murphy, M., Brown, P., Lave, R., Sellers, C., Mansfield, B., Fredrickson, L., Shapiro, N., & EDGI. (2017). Pursuing a toxic agenda: Environmental Injustice in the Early Trump Administration. https://envirodatagov.org/publication/pursuing-toxic-agenda/. Accessed 13 September 2018.
Paul, D. B., & Brosco, J. P. (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., DeFries, J. C., & Loehlin, J. C. (1977). Genotype–environment interaction correlation in analysis of human behavior. *Psychological Bulletin, 84*(2), 309–322.
Poland, J. (2004). Bias and schizophrenia. In P. Caplan & L. Cosgrove (Eds.), *Bias in psychiatric diagnosis* (pp. 149–162). Lanham, MD: Rowman & Littlefield.
Putnam, S., & Galea, S. (2008). Epidemiology and the macrosocial determinants of health. *Journal of Public Health Policy, 29,* 275–289.
Regan, M. M., & Gelber, R. D. (2005). Predicting response to systematic treatments: Learning from the past to plan for the future. *The Breast, 14,* 582–593.
Ridker, P. M., Buring, J. E., Rifai, N., & Cook, N. R. (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., Greenland, S., & Lash, T. L. (2012). *Modern epidemiology*. Boston: Lippincott Wolter.
Sampson, R. J. (2012). *Great American City: Chicago and the enduring neighborhood effect*. Chicago: Chicago University Press.
Stegenga, J. (2009). Marcel Weber: Philosophy of experimental biology. *Erkenntnis, 71*(3), 431–436.
Steinbach, R., Green, J., Kenward, M. G., & Edwards, P. (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, 21*(1), 1–19.
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. Accessed 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. (2011). Heterogeneity, control, social infrastructure, and possibilities of participation: Their interplay in modern understandings of heredity and in interpretation of science. Working papers on science in a changing world. http://scholarworks.umb.edu/cct_sicw/7. Accessed 3 September 2018.
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. Accessed 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., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (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). http://www.amstat.org/publications/jse/secure/v8n3/weldon.cfm. Accessed 27 September 2018.

Wright, R. J., Mitchell, H., Visness, C. M., Cohen, S., Stout, J., Evans, R., et al. (2004). Community violence and asthma morbidity: The inner-city asthma study. *American Journal of Public Health, 94,* 625–632.

Zandi, P. P., Sparks, D. L., Khachaturian, A. S., Tschanz, J., Norton, M., Steinberg, M., et al. (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., Hechtera, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *PNAS, 109*(4), 1193–1198.