Studies in History and Philosophy of Biological and Biomedical Sciences 52 (2015) 56–64

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

A reconsideration of the role of self-identified races in epidemiology and biomedical research

Ludovica Lorusso [a,b,*], Fabio Bacchini [c,d]

[a] Department of Political Science, Communication, Engineering and Information Technologies, University of Sassari, Italy
[b] Department of Philosophy, University of San Francisco, USA
[c] Laboratory of Applied Epistemology, DADU, University of Sassari, Italy
[d] Department of Philosophy, Stanford University, USA

ARTICLE INFO

Article history:
Available online 16 March 2015

Keywords:
Self-identified race
Biomedicine
Epidemiology
Genetic variation
Exposome
Epigenome

ABSTRACT

A considerable number of studies in epidemiology and biomedicine investigate the etiology of complex diseases by considering (self-identified) race as a relevant variable and focusing on the differences in risk among racial groups in the United States; they extensively draw on a genetic hypothesis—viz. the hypothesis that differences in the risk of complex diseases among racial groups are largely due to genetic differences covarying with genetic ancestry—that appears highly problematic in the light of both current biological evidence and the theory of human genome evolution. Is this reason for dismissing self-identified races? No. An alternative promising use of self-identified races exists, and ironically is suggested by those studies that investigate the etiology of complex diseases without focusing on racial differences. These studies provide a large amount of empirical evidence supporting the primacy of the contribution of non-genetic as opposed to genetic factors to the risk of complex diseases. We show that differences in race—or, better, in racial self-identification—may be critically used as proxies for differences in risk-related exposomes and epigenomes in the context of the United States. Self-identified race is what we need to capture the complexity of the effects of present and past racism on people's health and investigate risk-related external and internal exposures, gene–environment interactions, and epigenetic events. In fact patterns of racial self-identifications on one side, and patterns of risk-related exposomes and epigenomes on the other side, constantly coevolve and tend to match each other. However, there is no guarantee that using self-identified races in epidemiology and biomedical research will be beneficial all things considered: special attention must be paid at balancing positive and negative consequences.

1. Introduction

In the contemporary United States, the risk of morbidity and mortality from most complex or multifactorial disease is patterned along racial lines. For example, in 2009, the non-Hispanic black population had 141.3 deaths per 100,000 (age adjusted) that were due to coronary heart disease, compared with 117.7 deaths per 100,000 in the non-Hispanic white population.[1] In the same year the non-Hispanic white population had 37.8 deaths per 100,000 that were due to stroke, compared to 55.7 deaths per 100,000 in the non-Hispanic black population; non-Hispanic black females had 31.2 deaths per 100,000 population (age adjusted) that were due to breast cancer, over two and a half times the rate among Asian or Pacific Islander females, 11.4 per 100,000 (Healthy People, 2014).

* Corresponding author.
E-mail address: lorusso@uniss.it (L. Lorusso).

[1] According to the U.S. Office of Management and Budget (OMB), races in the U.S. are Black, White, Asian, American Indian, and Pacific Islander. "Hispanic" and "non-Hispanic" are ethnicities, not races.

http://dx.doi.org/10.1016/j.shpsc.2015.02.004
1369-8486/

The main aim of this article is to critically examine whether these differences are important in the study of the etiology of complex diseases—that is, whether "race" (specifically, self-identified race) can be usefully employed as a variable in the research into the causes of the susceptibility to complex diseases in the human population.

Indeed a distinction can be drawn in epidemiology and biomedical research among two categories of studies, which we call, respectively, "race-based studies" (RBS) and "race-neutral studies" (RNS). We define RBS as those epidemiological and biomedical studies investigating the etiology of complex diseases which do employ race as a relevant variable in their study design, thus focusing (among other things) on racial differences in the disease risks in the search for determinants of disease; as a consequence, these studies assume race to be a proxy for some causal factors on the pathway leading to disease which can either be specified or remain unspecified. On the other hand, we define RNS as those epidemiological and biomedical studies investigating the etiology of complex diseases which do not employ race as a relevant variable, do not focus on racial differences in the disease risks in the search for determinants of disease, and therefore avoid considering race as a proxy for any causal factors on the pathway leading to disease. While RBS consider it important to direct attention to associations between (self-identified) race and complex disease phenotypes, one of their goals being to determine a "racial" susceptibility to complex diseases, RNS do not consider such associations as relevant and only aim to determine a susceptibility to complex diseases which is not supposed to be race-specific.

This distinction clearly emerges in the literature, where many authors have similarly opposed "race-based" to "race-neutral" research, especially in genetic epidemiology and biomedicine—arguably because the use of race as a proxy for genetic features is considered ethically problematic and scientifically controversial. For instance, Fujimura and Rajagopalan (2011) remarks that in the field of biomedical genetic research it is possible to distinguish among those scientists who employ the race variable in the investigation of the etiology of complex diseases and those who think that "race categories […] are not appropriate tools to search for disease-related genes" (p. 6). Paradies, Montoya, and Fullerton (2007) differentiate between "race-neutral approaches" and approaches acknowledging race as a proxy of either genetic or social and environmental factors. In examining the use of racial categories deployed to explain specific disease patterns, Fausto-Sterling (2008) explicitly contrasts "medical scientists continu[ing] to study racial differences" (p. 659), "addressing racial differences" in disease study (p. 661), and assuming that "race might be an important study variable" (p. 662), to "race-neutral approaches" (p. 666). Similarly, Baer et al. (2013) oppose "health researchers consider[ing] race and ethnicity useful categories for health research" (p. 212) to those avoiding "the concept of race as a useful unit of analysis" (p. 213). Evidently the distinction between RBS and RNS is not new and has emerged various times in the relevant literature, however in slightly different forms (see also e.g. Fujimura, Duster, & Rajagopalan, 2008; Shields et al., 2005).

Notice that the distinction we introduce between RBS and RNS is orthogonal to the distinction among those studies that stress the role of genetic factors, and those ones that stress the role of non-genetic factors, in the explanation of the risks of complex diseases. Although assigning importance to information on race in the investigation of the etiology of complex diseases does not necessarily require favoring the causal role of genetic factors, a noticeable number of RBS in epidemiology and biomedical research[2] assume (however sometimes implicitly) that the race variable can play a relevant role as a proxy for genetic causal factors—as opposed to non-genetic causal factors—importantly contributing to the risk of complex diseases (Frank, 2007; Lee, 2009; Megyesi, Hunt, & Brody, 2011; Paradies et al., 2007; Rebbeck, Halbert, & Sankar, 2006).

In particular, RBS extensively adopt what we call the *genetic hypothesis*: they assume that differences in the risk of complex diseases among racial groups are largely due to genetic differences covarying with genetic ancestry which self-identified races are supposed to be good proxies for (e.g. Burchard et al., 2003; Campbell & Tishkoff, 2008; Drake, Galanter, & Burchard, 2008; Eeles et al., 2014; Kistka et al., 2007; Levin et al., 2014). Such genetic differences may consist in differences either in population-specific genetic variants or in genetic variants differentially distributed among populations (e.g. Aldrich et al., 2012; Kumar et al., 2010). Thus self-identified races can be used as proxies for a population-specific genetic component to the risk of complex diseases.

In this paper we argue that the genetic hypothesis is not immune from decisive criticism, and show that RBS seem up a blind alley in their use of self-identified races as proxies for genetic risk variants. However, this does not mean—as one may expect—that as a consequence self-identified races have no useful role to play in the research into the etiology of complex diseases. An alternative and epistemologically correct use of self-identified races exists, and ironically is suggested by those very studies that do not focus on races and make no use of self-identified races—viz. RNS.

In fact RNS provide a large amount of empirical evidence supporting the primacy of the contribution of a non-genetic as opposed to a genetic kind of variation to the risk of complex diseases (e.g. Miller & Jones, 2014; Rappaport & Smith, 2010; Vineis, Khan, Vlaanderen, & Vermeulen, 2009). We show, then, that self-identified races may be critically used as proxies for a risk-related environmental and epigenetic variation in the context of the United States. Self-identified race is what we need to capture the complexity of the effects of present and past racism on people's health and investigate risk-related external and internal exposures, gene–environment interactions, and epigenetic events. Our point is that a promising category of studies into the etiology of complex diseases is that of RBS focusing on non-genetic causal factors: self-identified race can be correctly employed as a useful variable in epidemiology and biomedical research, provided that the genetic hypothesis is dismissed.

2. The genetic hypothesis in RBS

In the genetic hypothesis self-identified races are considered as proxies for a specific genetic ancestry associated with specific genetic variants contributing to the risk of complex diseases (e.g. Bustamante, Burchard, & De La Vega, 2011; Fejerman et al., 2013; Kumar et al., 2010). Generally genetic ancestry itself is considered unlikely to be the cause of the population-specific genetic susceptibility, but is taken as a proxy for genetic variants contributing to the risk, which are supposed to be either population-specific or differentially distributed among populations (Aldrich et al., 2012). So the genetic hypothesis is based on three main assumptions:

1. Self-identified race is a good proxy for a specific genetic ancestry.
2. Different specific genetic ancestries can be correctly and unambiguously identified.
3. A specific genetic ancestry can be used as a proxy for unknown genetic variants contributing to the risk of a complex disease, which are supposed to be either population-specific or differentially distributed among populations.

These three assumptions are problematic for different reasons. The first assumption is problematic because racial self-

[2] Note, however, that in epidemiology the use of (self-identified) race as a proxy for risk-related genetic factors seems less common than it is in biomedical research.