

Introduction

The influence of biomedical research extends far beyond the walls of the hospital and clinic. How scientists see the world has a significant impact on how the rest of society will see the world. The use of racial and ethnic categories in science is a prominent example of how the work of scientists and physicians can have a tremendous impact on the rest of society. As such, scientists and physicians have a responsibility to use racial and ethnic categories ethically.

The use of ethnic and racial categories in science has continued well into the genomic age. After the completion of the human genome project, some believed that the biological meaning of race would finally be resolved. Instead, there is more controversy now than ever. Genetics has opened up a completely new way to categorize human populations. Whether race can be ethically used in genetic research is still hotly debated. Some believe that racial categories are not biologically meaningful, and cannot be used in science. Meanwhile, others believe that racial categories are biologically meaningful and cannot be ignored.

In this paper I adopt the position that racial categories may be used in science as a proxy for social factors, but not for genetic or biological factors. In Section 1, I review history of race in science. In Section 2, I outline the ethical dangers of using racial categories in science. In Section 3, I briefly describe the science of genetic epidemiology.

In Section 4, I discuss the dangers of racial categories specifically in international genetic research. In Section 5 I outline the major positions regarding the use of racial categories in scientific and genetic research. In section 6, I evaluate the use of racial categories in a prominent international genetic study, the International HapMap Project.

Section 1- The history of race: An effort to categorize and sort human populations.

Race is a difficult concept with a dark and complicated history. In its most naked interpretation, stripped of any social and political baggage, race is simply a way to categorize human populations. Humans have always categorized things in an effort to understand them. We have created systematic methods for categorizing everything from music, art, and philosophy to cars, plants, and animals. Human populations are not exempt from the human drive to organize and understand. While the categorization of human populations is not intrinsically problematic, race is unique because it has evolved into a concept burdened with heavy ethical, legal, and social issues.

Historically humans have demonstrated a tendency to sort themselves by criteria such as appearance, language, and religion. Early interpretations of race were efforts to sort human populations based on common physical traits. Scholars believed that race could be an objective and neutral way of sorting human populations based on physical characteristics such as skin color, anatomical variation, and behavior. For example, in the mid 18th century, Carolus Linnaeus, the father of modern taxonomy, sorted humanity into 4 groups based on geography, skin color, and behavior: *Homo Sapiens Americanus*, *Asiaticus*, *Africanus* and *Europeanus*.ⁱ Similarly, in Johann Blumenbach's famous book,

On the Natural Varieties of Humankind, Blumenbach sorted humans into 5 populations: white, black, yellow, brown and red.ⁱⁱ

As time went on, scholars who studied the variation and distribution of traits among human populations found that there were intermediate traits that did not fit neatly into one of the discrete 4 or 5 basic racial categories presented by Linnaeus and Blumenbach. To accommodate the new populations that could not be fit into a preexisting racial category, new racial categories were formed. The number of racial categories gradually increased from 5 to 50.ⁱⁱⁱ Eventually, anthropologists concluded that there were no discrete racial groups because human traits and behaviors were found to vary continuously.^{iv}

Section 2- The ethical problems of race: From Hierarchical categories to slavery and genocide

Race is an ethically challenging concept because it has been the conceptual foundation for egregious violations of human rights including slavery and genocide. The ethical and moral problems started when scientists who studied race began to rank races by perceived standards of biological and social worth.^v Scientists would compare races based on criteria such as brain size and intelligence. Brain size was associated with intelligence. Intelligence was then associated with social worth. Finally, social worth was associated with moral status. Such rankings of race embedded racial categories with social and moral values.

What followed was the evolution of biologically deterministic theories, which suggested that scientifically observable biological criteria could conclusively indicate

social worth or status. The thrust of these theories was that observable traits like brain size or skin color could be used to evaluate an individual's or population's social worth. The logical conclusion was that people within a "scientifically determined racial category" could be less valuable or less human, and therefore were not due the rights and moral consideration that were due other superior races. Eventually race would become the conceptual foundation for slavery and genocide. Any use of racial categories in science that tends to validate and perpetuate hierarchies of race and social worth is ethically problematic.

Section 4- Genetic Epidemiology.

The discovery of the molecular structure of DNA created a revolution in science, medicine, and culture. Genetics provided a completely new way to not only understand and treat disease, but also a completely new way to categorize and understand human populations. Some scientists believed that genetics would finally prove race was a biologically meaningless, while others believed that genetics would prove race was biologically meaningful.

Human populations can be defined by genetic similarities and differences. These different genetic populations have different traits and features, as well as different rates of disease. Most traits and diseases are a combination of environmental and genetic factors. Different populations will have different physical traits as well as different predispositions and susceptibilities to disease. Genetic epidemiology is the study of the distribution of genes related to disease.

The power of genetic epidemiology is not only that it allows scientists to find genes and locations in the human genome that are associated with disease, but also that it allows scientists to identify individuals and populations who either have, or are at risk for certain disease. Being able to identify the genes will facilitate finding treatments and potentially cures. Being able to identify individuals and populations at risk will allow for interventions that may prevent death and disease. The potential power of genetic epidemiology to increase health is immense.

In order for genetic research to be efficient and meaningful, the study population needs to be genetically similar. There are 3 main reasons for why populations used in genetic research need to be genetically similar. The first is the statistical power of the study increases as the genetic similarity of the study population increases.^{vi} The second is that the utility and efficiency of interventions and subsequent research increases as the genetic similarity of a population increases.^{vii} Finally, many scientists believe that most human disease is caused by a common genes that all humans possess. This is called the Common Disease Common Variant hypothesis. If the common disease common variant is true, then the discovery of genes in one population benefits all other human populations as well.^{viii}

Section 3- The ethical danger in using racial and ethnic categories in international genetic research.

Racial categories are being used in genetic research as a means to select for genetically similar study populations. Historically, racial and ethnic categories have frequently been used to define study populations in medical and scientific research. As a

general principle of science, all variables other than the experimental variable being studied should be controlled. Race has been used as a method to select for a study population that is socially and biologically similar to reduce variability and confounding in scientific research.

The danger of using racial categories in conducting international genetic research is not only that it may reaffirm and perpetuate the notion that race is scientifically observable and biologically real, but it may also lead to the belief that a racial, national, or ethnic group may be biologically inferior. This may happen in several ways. First, a study may find that a population recruited and defined by racial category may have higher rates and a higher risk for certain diseases. This may lead to the perception that the racial category is unhealthy, diseased, weak, or unclean.^{ix} Second, international genetic epidemiology research may frame a disease or a health problem as a racial or ethnic problem, and not as a human, social, or environmental problem.^x Third, international genetic research may evaluate several different racial groups within a study, and different racial groups may be compared and ranked based on the results of the study.^{xi} Fourth, the use of racial categories in science validates race as a meaningful scientific category, and perpetuates the notion that race is natural, scientifically observable phenomenon.

Section 5- Ethical usage of racial and ethnic categories in international genetic research

5.1- The Problem of Racial Categories in genetic epidemiology research

It is clear that the use of racial and ethnic categories in science is ethically problematic, given not only the history of harms but also the debatable scientific value. It is also clear that genetically similar populations are necessary to conduct genetic epidemiology research. For many years, scientists and scholars have used race as a method to select for similar populations. Can this practice be justified in international genetic research? Is it ethically permissible to use race as a proxy for genetic similarity in international genetic epidemiology?

There are three general positions: The first position maintains that race is not biologically or genetically real, and that racial categories should not be used in science at all. The second position maintains that race is biologically meaningful, and therefore race can and should be used in genetic and medical research. The third position maintains that race is a unique concept that is both socially constructed and biologically meaningful, and race may be used as a proxy for certain social factors, but never as a proxy for genetic or biological similarity.

5.2 – Position 1: Racial and ethnic categories are not real, and cannot be ethically used

Many scientists and scholars maintain that it is unethical to use racial categories in scientific research. These scientists offer 3 main arguments in support of their position. First, they maintain that race is not biologically real, and therefore that it is not a scientifically valid concept and cannot be ethically used as a proxy for any biological or genetic properties. There is a large body of work that concludes race is socially constructed and has no biological meaning.^{xii} These authors point out that definitions of

race are fluid and imprecise, continuously varying by time and place.^{xiii} They also point out that genetic research has shown that all humans are 99.9% genetically similar,^{xiv} and that genetic populations vary continuously the same way physical traits do.^{xv} Second, they maintain that racial disparities in health and medicine can be attributed to disparities in other social and environmental factors, not biological or genetic factors. These disparities are attributed to factors such as poverty, pollution, lack of health care resources and services, and other social, economic and environmental issues that disproportionately effect certain populations.^{xvi} They believe that once these social and environmental factors are adjusted, racial disparities in health will disappear. Third, they maintain that the harms of racializing science can best be avoided by rejecting the use of racial categories in science.^{xvii} The harms of labeling a racial population in scientific research are unavoidable unless the use of racial categories is altogether stopped.

5.3 – Position 2: Racial and ethnic categories are biologically meaningful, and can be ethically used

Many scientists maintain the use of racial categories in scientific research is both necessary and ethically permissible. These scientists offer three arguments in support of their position: First, there is strong scientific evidence that racial categories are biologically and genetically meaningful. Genetic research has shown that individuals within a racial category are more genetically similar to each other than individuals outside their racial category,^{xviii} and that genetic markers can reliably predict an individual's self-selected racial category.^{xix} Second, it is argued that there is an unavoidable and irreducible affect of race on individual health that cannot be ignored.^{xx}

Even if every social, economic, and environmental factor is taken into account, race will still be an important factor in studying the interactions of biological and environmental factors. Race can impact health even after every social and environmental factor is adjusted because as long as race is a social reality, it can still cause create social and biological consequences. Third, removing race from science does not stop racism, and failing to study race actually enables racism in medicine and research.^{xxi} Without racial categories in research, scholars and scientists cannot study the effects of racism and cannot study the interactions of racism and medicine. Racism exists, and addressing the consequences of racism are important and shouldn't be ignored.

5.3 Position 3- An alternative position: Racial Categories may be used as a proxy for social factors, but not genetic or biological factors.

A few scholars have presented a novel position that reconciles some of the apparent contradiction in the discussion regarding race and biology. These scholars not only acknowledge that race is socially constructed and imprecise, but they also maintain that race is real and does affect health and biological outcomes. These scholars explain that race is a socially constructed phenomena that has biological consequences. Ossorio and Duster write,

“Race is a complex but empirically demonstrable stratifying practice that creates identity and hierarchy through social interaction...The social fact of racial stratification has biological consequences, which is why race is a relevant, appropriate variable in some biomedical research.”^{xxii}

The biological consequences of racial stratification and segmentation are important and need to be understood. As such, these scholars believe that race may be

used as category in science to study how social stratification and social factors influence health outcomes. The important issues are how and why racial categories are used in research. Race should not be used as a proxy for biological or genetic similarity because race is an imprecise cultural variable and should not ever be attached to an objective scientific variable.^{xxiii} Race should also not be used as a proxy for a social or environmental factor where there is another more direct way to measure the variable, such as measuring income or education. Race may be used as a proxy for a social factors the influence health and biological outcomes to the extent that there are no other more direct measures available.^{xxiv}

I subscribe to this position because it not only acknowledges that race is socially constructed, but also that it does have biological consequences. As such, those who hold this view neither reject race as a meaningless, nor accept race as biologically real. It allows race to be studied as what it truly is, a social construct with biological consequences.

Section 6- Applicaton to the HapMap..

The International HapMap Project is an ambitious international research program that created a public database of common patterns of human genetic variation. The ultimate purpose of this research was to create a resource that would accelerate genetic epidemiology research worldwide. This project is a prominent example of the use of racial and ethnic categories in international genomic research. As such, it has great responsibility in setting an example for the proper use of racial categories in international genetic research. I would maintain the HapMap project failed to use racial and ethnic

categories in a responsible manner because the HapMap project used racial categories as a proxy for genetic similarity. Use of racial categories as a proxy for genetic similarity in genetic research is ethically problematic because it validates an imprecise, socially constructed category as a natural, scientifically meaningful category.

To fulfill the purpose of providing a meaningful and internationally useful database of the patterns of human genetic variation, HapMap researchers would need to find study populations that meet 2 criteria. First, the study populations involved would have to be genetically similar. Genetic similarity not only increases statistical power, but also allows for more specific and efficient association studies and interventions in similar populations. Second, the study populations involved should include most of the genetic variation in the whole world. The more diverse the study populations are, the more variation of the human genome is included and the more useful the database will be to international research. To fulfill these criteria, researchers decided to recruit 4 study populations: a Han Chinese population, Japanese population, an African population from the Yoruba, and a European/American population recruited in Utah.^{xxv} Their rationale was that the genetic similarity within each study population would be sufficient, and that between all 4 study populations involved, most of the genetic variability in the world would be included in the study.

The methodology the HapMap employed to select for the individual participants from the study populations depended upon the study population. For the African study population, the participants were required to have 4 Yoruba grandparents, as identified by a physician. For the Chinese study population, the participants were required to have 3 Han Chinese grandparents. For the Japanese population, study participants were generally

informed that the goal of the study was to collect samples from people whose grandparents were from Japan. The European/American sample from Utah were taken from a previous study, where it is “generally known” that the participants were of northern and western European descent.^{xxvi} These selection methods are imprecise and inconsistent. The researchers themselves concede, “Asking individuals to identify themselves, or their grandparents, as having a particular ancestry will inevitably result in some imprecise or inaccurate assignments.”^{xxvii}

HapMap researchers offer 2 general arguments in support of their use of racial categories. The first argument is that at this point in the state of the science of genetics, use of racial categories as proxies is the most efficient, useful, and practical way to conduct this type of research. Charles Rotimi articulates 4 reasons for why use of racial categories in this research was the best approach. First, studies have shown that haplotypes vary between different populations. Second, having population information allows for more efficient association studies based on the results. Third, masking the study population’s identity or label creates a false sense of security, and cannot be maintained to protect the study population. Fourth, identification of the study population allows for better interpretation of results because it allows scientists to consider context.^{xxviii}

The second argument is that the use of categories in this research is not problematic because the purpose or consequence of the research is not to define the populations. The HapMap researchers explain,

“The lack of precise population definitions and ancestral geographic self-assignments is not problematic for developing the HapMap because exact

demarcations are not necessary for the way that the HapMap will be used, and the Project does not aim to define populations. No claims are made about the genetic “purity” of the sample sets or the populations of which donors are members; such claims would be scientifically spurious, as human populations are the products of countless social, historical and demographic processes, and therefore cannot be sharply defined.”^{xxix}

All of the arguments articulated by the HapMap researchers are true and well taken; however, they do not address the harm from using racial categories in international genetic research. First, the utility of racial categories in science does not necessarily justify their continued usage. An analysis must be made weighing the harms and benefits; it is not sufficient to note that there is scientific utility. The utility must be weighed against the harms, and an evaluation should be made as to whether an effective, alternative method of measurement exists.

Second, the purposes of the usage do not mitigate the potential harms; that the intention of the researchers was not to demarcate populations or make claims of purity is irrelevant to the harm caused by the usage of the racial categories in the research. Use of the category in research attaches scientific meaning and importance to an imprecise, inconsistent and subjective social construct. Furthermore, usage of the categories may lead to the harms of ranking, comparing, or labeling of the study populations based on the results of future research.

While the results of research may be scientifically and medically valuable, I would maintain that racial categories should never be used as a proxy for genetic similarity in international genetic research because there is no way to mitigate the harms. Scientists should try to find other criteria or methods to select for populations that are genetically similar. As progress is made in genetics, more tools should be developed that

allow for scientists to select for genetic similarity that does not involve racial categories. In general, conscious effort should be made by scientists to only use race as a proxy for social factors, and not genetic or biological factors.

References

-
- ⁱ Linneaus, C. *Systema Naturae*. 1770
- ⁱⁱ Blumenbach, JF. *De generis humani varietate native*. 1775
- ⁱⁱⁱ Marks, J. (2002) What it means to be 98% Chimapanzee. Berkeley: University of California Press.
- ^{iv} Id.
- ^v Bhopal, R. Is research into ethnicity and health racist, unsound, or important science? *BMJ* 1997; 314:1751-6
- ^{vi} Cardon, LR & Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003; 361:598-604
- ^{vii} The International HapMap Consortium. Integrating ethics and science in the International HapMap Project. *Nature Reviews Genetics* 2004;5:467-475
- ^{viii} Foster MW & Sharp RR. Will investments in biobanks, prospective cohorts, and markers of common patterns of variation benefit other populations for drug response and disease susceptibility gene discovery? *The Pharmacogenomics Journal* 2005;5:75-80
- ^{ix} Bhopal, R. Is research into ethnicity and health racist, unsound, or important science? *BMJ* 1997; 314:1751-6
- ^x id
- ^{xi} id
- ^{xii} Schwartz, RS. Racial profiling in medical research. *NEJM* 2001; 344(18):1392-1393.
- ^{xiii} Keita SOY, Kittles RA, Royal CDM, Bonney GE, Furbert-Harris P, Dunston GM & Rotimi CN. Conceptualizing human variation. *Nature Genetics Supplement* 2004;36(11):517-520
- ^{xiv} Tishkoff SA & Kidd KK. Implications of biogeography on human populations for 'race' and medicine. *Nature Genetics Supplement* 2004; 36(11): s21-s27
- ^{xv} Keita SOY, Kittles RA, Royal CDM, Bonney GE, Furbert-Harris P, Dunston GM & Rotimi CN. Conceptualizing human variation. *Nature Genetics Supplement* 2004;36(11):517-520
- ^{xvi} Chaturvedi, N. Ethnicity as an epidemiological determinant- crudely racist or crucially important? *International Journal of Epidemiology* 2001; 30:925-927.
- ^{xvii} Schwartz, RS. Racial profiling in medical research. *NEJM* 2001; 344(18):1392-1393.
- ^{xviii} Bamshad M, Wooding S, Salisbury BA & Stephens Claiborne J. Deconstructing the relationship between genetics and race. *Nature Reviews Genetics* 2004;5:598-608
- ^{xix} Id.
- ^{xx} Ossorio P & Duster T. Race and Genetics, Controversies in Biomedical, Behavioral and Forensic Sciences. *American Psychologist* 2005; 60:115-128
- ^{xxi} Burchard EG, Ziv E, Coyle N, Gomez SL. The importance of race and ethnic background in biomedical research. *NEJM* 2003; 348(12) 1170-1176.
- ^{xxii} Ossorio P & Duster T. Race and Genetics, Controversies in Biomedical, Behavioral and Forensic Sciences. *American Psychologist* 2005; 60:115-128
- ^{xxiii} Sankar P. & Cho MK. Toward a new vocabulary of human genetic variation. *Science* 2002; 298:1337-1338.
- ^{xxiv} Ossorio P & Duster T. Race and Genetics, Controversies in Biomedical, Behavioral and Forensic Sciences. *American Psychologist* 2005; 60:115-128

^{xxv} The International HapMap Consortium. The International HapMap Project. *Nature* 2003; 426:789-796

^{xxvi} The International HapMap Consortium. Integrating ethics and science in the International HapMap project. *Nature Reviews Genetics* 2004; 5:467-475.

^{xxvii} The International HapMap Consortium. Integrating ethics and science in the International HapMap project. *Nature Reviews Genetics* 2004; 5:467-475.

^{xxviii} Rotimi CN. Are Medical and nonmedical uses of large-scale genomic markers conflating genetics and 'race'? *Nature Genetics Supplement* 2004; 36(11): s43-s47

^{xxix} The International HapMap Consortium. Integrating ethics and science in the International HapMap project. *Nature Reviews Genetics* 2004; 5:467-475.