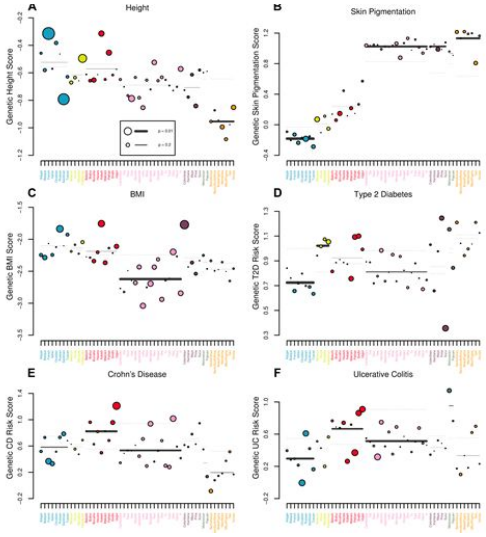
HARVARD UNIVERSITY
Edmond J. Safra Center for Ethics

Meeting Minutes

A workshop on interpreting the genetic bases of differences between populations and on the interactions among concepts used for research in social and natural sciences

Oct 25-26, 2018



Co-organizers

Danielle Allen (Harvard); Anna Di Rienzo (UChicago); Evelynn Hammonds (Harvard); Molly Przeworski (Columbia); Alondra Nelson (Columbia + SSRC)

sponsored by Harvard's Edmond J. Safra Center for Ethics

with Additional sponsorship from the Social Science Research Council and the project on race & gender in science and medicine at Harvard's Hutchins Center for African and African American research

Overview

On October 25 and 26, 2018, Harvard's EDMOND J. SAFRA CENTER FOR ETHICS hosted a small-scale academic workshop, bringing together thirty scholars from a variety of disciplines—including the social sciences, philosophy, statistics and genetics—to probe the conceptual and technical underpinnings of research in contemporary genomic science and the advent of polygenic approaches, in particular. The purpose of the workshop was to understand the stakes of the "polygenic turn" in genetic science. Given that this research draws on both the social and natural sciences, our starting premises were that (1) scholars need to work across multiple disciplines to clarify and probe the core concepts of this new paradigm; (2) the necessarily multi-disciplinary nature of the conversation requires significant mutual learning across disciplinary boundaries; and (3) substantive learning across fields requires small, workshop-style contexts in which scholars from different disciplines can begin to learn each others' vocabularies and methodologies and begin to put their diverse but complementary tools to work together.

The workshop agenda focused on two themes: (1) key concepts in social science and human genetics; and (2) the science behind polygenic risk scores and challenges facing their application in social sciences, medicine, and public policy. A number of possible action steps were discussed, and we focused on two: (1) the creation of a working group for further work on the population concept and scholarly practices for the use of population labels; (2) the creation of a working group to explore the best avenues for articulating and disseminating guidance on the valid use of polygenic risk scores in their various potential contexts of use (research, clinical contexts, etc.); the group will explore the possibility of working through professional associations and/or of developing other mechanisms and partnerships for addressing this issue.

The Discussion Questions

(1) **Key Concepts:** We asked how "populations" are defined and selected for genetic analysis in the first place? When and why are "groups" a natural unit of analysis for questions in genetics and evolutionary biology and according to what explicit and implicit criteria are individuals grouped? And with what consequences? How do we understand the concepts of heritability and heredity in genetics and social science? How is heritability estimated and how do we interpret these estimates? Can we develop precise operational definitions for these and other core concepts?

(2) **The science behind polygenic risk scores and challenges facing their applied use:** We asked: How do genome-wide association studies operate so as to develop predictive "polygenic scores"? What predictive power do polygenic risk scores have? In what contexts and in what sense are polygenic scores "causal"? What are the current challenges to the validity of prediction based on polygenic scores? Are polygenic scores portable across population contexts and to what extent? Are they portable even within a single population context? Do statistical genetics and biology have the tools to address these challenges?

The effect sizes estimated in a genome-wide association study can capture direct effects of the variants (or their proxies) but also residual confounding, indirect effects through parents or other relatives and

so forth. Also, gene by environment interactions may be common: direct and especially indirect effects are unlikely to remain the same in different environments. Social scientists have begun to use polygenic scores in their research, to try to isolate environmental effects, and their use in policy-making and clinical contexts are also under exploration. Given the instability, and uncertainty of the underlying data, how should we think about these uses? What are the current limitations for applied use of polygenic scores and how should they be communicated?

The Discussion

We sharpened our collective understanding that in the science of the polygenic turn, populations are defined in a diversity of *ad hoc* ways, and we explored many contrasting uses of the concept, some connected to evolutionary processes, some not, some connected to self-identified social categories of the contemporary world, some connected to stylized "ancestry" populations; some based on statistical clustering, and so on. We came to consensus that it would be helpful to create a working-group that would (1) create a descriptive taxonomy of current uses of the "population" concept; (2) identify the specific methodologies entailed in using one or another version of the "population" concept; and (3) develop a labelling taxonomy for reporting in research papers on the particular versions of the concept used in a given study as well as on the specific methodologies of the concepts' construction used in the study.

With regard to polygenic risk scores, our discussion focused on the facts that polygenic risk scores are developed within specific populations and that there is as of yet very little basis for considering prediction based on polygenic risk scores to be equally accurate across or even within populations. These limitations in portability pertain regardless of whether the populations are "self-identifying", based on contemporary social systems of classification, or "stylized ancestry" populations. Importantly, the stylized ancestry datasets currently used by population geneticists do not align with contemporary social categories of self-identification. The need for a consensus statement about these limitations became apparent. These limitations further raised questions of what advice to give concerning the use of polygenic risk scores in social science research and applied settings. We therefore came to consensus around the need to create a second working group to undertake two tasks: (1) to develop a consensus statement about the science and limitations of polygenic risk scores and their meaning; (2) to identify the best avenues for developing, articulating, and disseminating guidance on the valid use of polygenic risk scores in their various potential contexts of use (research, clinical contexts, etc.). With regard to the latter prong of work, the working group will explore the possibility of collaborating with professional associations and/or of developing other mechanisms and partnerships for addressing this issue.

Our Next Steps

We will continue our conversation via two working groups, as per the descriptions above, and look forward to engaging a broad network of scholars in these conversations.