Using Population Descriptors in Genetics and Genomics Research

A New Framework for an Evolving Field

Genetics and genomics research has grown exponentially over the past decade, providing exciting opportunities to transform medicine as we know it. The genetics ecosystem has expanded such that genetics and genomics research is now conducted across a wide variety of disciplines. Researchers have long used population descriptors, or concepts of categorization, to capture information related to the continuous and complex patterns of human genetic variation resulting from history, migration, and evolution. Many of these descriptors focus on *descent-associated* groupings (i.e., populations believed to share characteristics based on a common origin). Examples of population descriptors include race, geography, and nationality, which all have complex social definitions and connotations. The misuse of these descriptors—in particular, race—has persisted and influenced thinking in genetics and genomics that has benefited some while marginalizing others.

Researchers using genetic and genomic data have long struggled with a lack of clear and specific guidance regarding the use of population descriptors. Over the years, many efforts have sought to develop appropriate guidelines with little success. Chapter 5 of this report recommends new approaches for the use of population descriptors according to genomics study types. For both individual and collective behavior to change, incentives,

as well as accountability measures, should be in place. There is a great need for partnership among relevant parties to support researchers during the implementation process, as described in Chapter 6.

RECOMMENDED ACTIONS

Offer publicly available tools to facilitate implementation.
Although the genomics research ecosystem comprises many different groups, implementation of the recommended changes may disproportionately rely on individual researchers. It is important that other relevant parties, including research institutions and funders, share the responsibility and support researchers to effect lasting change.

Action: Funding agencies, research institutions, research journals, and professional societies should offer tools widely to their communities to facilitate the implementation of these recommendations. The tools should be publicly available, especially when they are supported by public funds. Such tools could include:

- Educational modules for inclusion in human research protection training

- Manuscript submission and review guidelines

- Grant submission and review criteria (example checklist below)

- Training and education of individuals at all levels

- Opportunities for continuing education for researchers

- Informatics tools, such as data structure standards for sharing labels and labeling procedures used within a study

Funding agencies and research institutions could collaborate to develop and provide training and continuing education about the proper use of population descriptors. Researchers could be required to complete training about the use of population descriptors before engaging in research with human participants, and before being granted access to data sets.

During procedural checks for human research, considering how researchers intend to use population descriptors would help achieve desirable change. A table or form would permit a more objective determination as to whether a proposal has answered important questions about using and reporting population descriptors and can be applied across all proposals.

Incentivize interdisciplinary collaborations to improve research studies.

One of the confounding issues in studying trait variation is distinguishing between genetic and environmental factors. Geneticists and researchers using genetic and genomic tools may lack the social and environmental data they need to analyze the most appropriate nongenetic variables. They may also lack the training and expertise to design and carry out a study that will collect such data. Collaborations among geneticists, epidemiologists, demographers, and other social

EXAMPLE CHECKLIST THAT FUNDERS CAN IMPLEMENT DURING GRANT REVIEW FOR GENOMICS RESEARCHERS

- What is the source of the data for your study?

- Are these individual-level data or group-level summary data?

- Have you clearly defined the purpose of your study?

- Have you engaged with the community that you would like to study?

- Has the community that is offering the use of its data to you had opportunities to identify themselves and explain why these are the descriptors they use to identify themselves? (Alternatively, has the research group sharing data provided guidance for how to develop population descriptors for the communities they have sampled?)

- Has consent been given for broad reuse of the data in research?

- Have you completed any required training on population descriptors?

- Have you determined which population descriptors are most appropriate for your study and do you understand why?

- Is interdisciplinary expertise needed to design and conduct the study and evaluate the data?

- Do you have a plan to clearly communicate the results of your study with the research community, including research participants?

- Do you plan to collect multiple descriptors, including specific measures of relevant environmental factors?

scientists can therefore improve study design and statistical analysis of the data to better differentiate between genetic and nongenetic factors and their effects.

Similarly, effective community engagement improves communication, study coordination, and trustworthiness of research. Community-engaged research requires multidisciplinary approaches that draw on diverse expertise; thus, research teams should include members with the expertise to develop ongoing partnerships with communities.

Action: Research institutions and funding agencies should embed incentives for fostering interdisciplinary collaboration among researchers with different areas of expertise, including genetics and genomics, social sciences, epidemiology, and community-based research, to facilitate the inclusion of environmental measures and the engagement of diverse communities in genomics research. Funding agencies and research institutions should develop strategies to encourage and reward such collaborations.

Create initiatives to advance best practices for the use of population descriptors.

Action: Given the persistent need to address this dynamic, high-stakes component of genomics research, funders and research institutions should create new initiatives to advance the study and methods development of best practices for population descriptor usage in genetics and genomics research, including the public availability of resources.

Computational tools offer a promising avenue to compare studies and identify differences in the metadata related to population descriptors. For example, these tools could assist a researcher in deciding whether merging data sets would be appropriate and enable them to use the merged data to address a question of interest.

Align policies and procedures and invest in strategies to support implementation.

The ability to advance trust and improve research depends on how groups implement the report's recommendations. Current systems for supporting

and rewarding genomics research may impede rather than facilitate implementation. It is important that organizations candidly evaluate how well their current practices and procedures align with the recommendations and course correct where needed. If changes cannot be made to resolve discrepancies, groups should be transparent, provide justification as to why, and determine how misalignment will be mitigated.

Action: Key partners, including funding agencies, research institutions, and scientific journals, should ensure that policies and procedures are aligned with these recommendations and invest in developing new strategies to support implementation when needed.

An additional challenge for researchers is the way Office of Management and Budget (OMB) Directive 15 has confounded the use of population descriptors in genetics and genomics research. A helpful step in implementing this report's recommendations would be for funding agencies to instruct researchers that they do not have to use OMB categories to group individuals in their analytical work. The need and rationale for reporting sampling procedures to funding agencies are distinct from researchers' reasoning and decisions around study design and data analysis. For example, race and/or ethnicity may be useful for identifying individuals to include in a study and ensuring a heterogeneous sample. However, for analysis, researchers should use the most appropriate population descriptors for the questions they are probing, instead of defaulting to the OMB categories.

CONCLUSION

Funding agencies, research institutions (including associated institutional review boards and other activities with research participants), research journals, professional societies, and lay media professionals should evaluate their processes and procedures related to the use of population descriptors in genomics research and report to their communities whether they are facilitating the recommendations in this report. A plan should be provided, along with a timeline, to change policies and procedures that are not aligned with these recommendations.

To create sustainable change, the full genetics and genomics research ecosystem—research participants, research funders, professional societies, research journals, research institutions, and individual researchers—must work together to implement these recommendations. Both transparency and communication among the relevant groups will be critical. The field is quickly evolving, and the time to meet this need is now.

Health and Medicine Division

Division of Behavioral and Social Sciences and Education

NATIONAL ACADEMIES

Sciences
Engineering
Medicine