Racial & Ethnic Disparities in Genetic Database Participation

Destiny Jackson & Joanna Kim

Mentors: Arnold Rosoff, Robert Field, Anthony Orlando

Intros

Destiny Jackson

Emory University ‘22

Sociology and Global Health

Joanna Kim

University of Pennsylvania ‘22

Health and Societies and American History

Intros: Mentors

Arnold Rosoff, JD

Robert Field, PhD, JD, MPH

Anthony Orlando, PhD, M.Sc.

1

Background & Objectives

Overview of Genetic Databases and Research Questions

Background

Genetic Databases

- Clinical
- Research
- Direct to Consumer (DTC)

PMBB | Penn Medicine BioBank

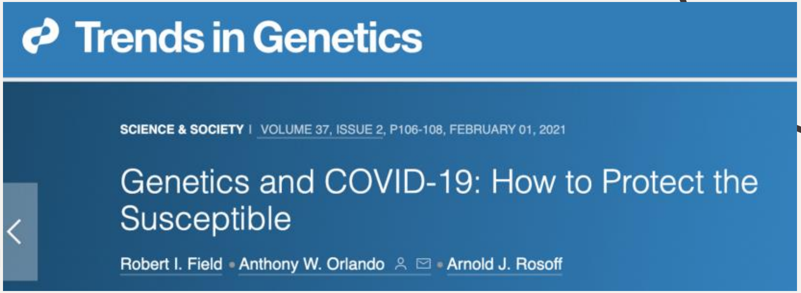
ancestry

All of Us Research Hub

23andMe

Significance

- Genetic Databases and Genetics Research
    - Privacy Concerns
    - Underrepresentation of certain subgrou
        - **Negative Implications**
            - **Less variation limited robustness of genetics research**
            - **Limited applic⇨ility of medical advancements**

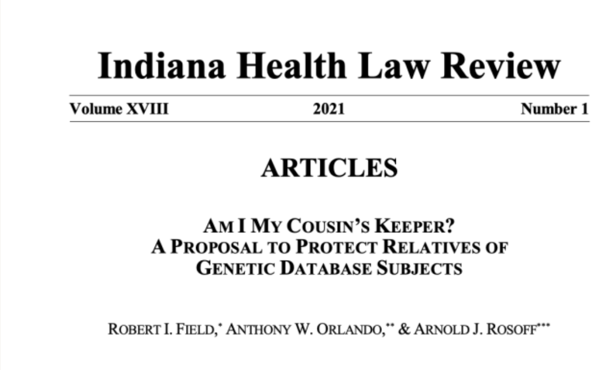


Trends in Genetics

SCIENCE & SOCIETY | VOLUME 37, ISSUE 2, P106-108, FEBRUARY 01, 2021

Genetics and COVID-19: How to Protect the Susceptible

Robert I. Field • Anthony W. Orlando • Arnold J. Rosoff



Indiana Health Law Review

Volume XVIII 2021 Number 1

ARTICLES

AM I MY COUSIN'S KEEPER?
A PROPOSAL TO PROTECT RELATIVES OF
GENETIC DATABASE SUBJECTS

ROBERT I. FIELD,* ANTHONY W. ORLANDO,** & ARNOLD J. ROSOFF***

Field, R., Orlando, A. and Rosoff, A. "Am I My Cousin's Keeper? A Proposal to Protect Relatives of Genetic Database Subjects." *Indiana Health Law Review*, vol. 28, no. 1, p.1-55, 2021.

Field, R., Orlando, A. and Rosoff, A. "Genetics and COVID-19: How to Protect the Susceptible." *Trends in Genetics*, vol. 37, no. 2, p. 106-108, 2021.

Research Questions

1

Are there specific subgroups whose underrepresentation presents a problem?

2

How can different applications of genetic databases be undermined by underrepresentation?

Literature Review

3

How common is the underrepresentation of certain population subgroups in genetic databases and genetic research?

Data Collection & Analysis

4

What are database generators, users, and regulators currently doing to address any problem?

5

What measures should/can be taken to increase the representation of underrepresented groups in genetics research?

Interview

2

Methodology & Research Process

Methodology Overview

1. Literature Review & Summary
2. Data Collection & Analysis
3. Interviews

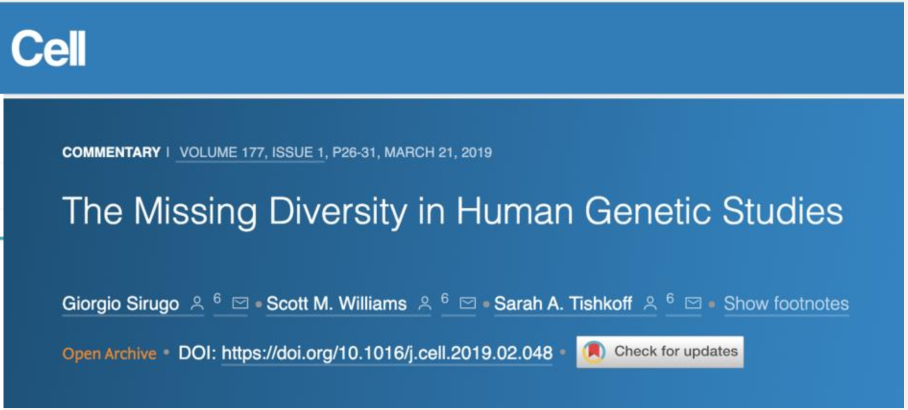
Step 1: Literature Review

- Phase 1
  - News articles
- Phase 2
  - Academic articles

LIFE & WORK | HEALTH & WELLNESS

For Some African-Americans, Genetic Testing Reopens Past Wounds

A legacy of unethical studies complicates calls for greater participation by African-Americans

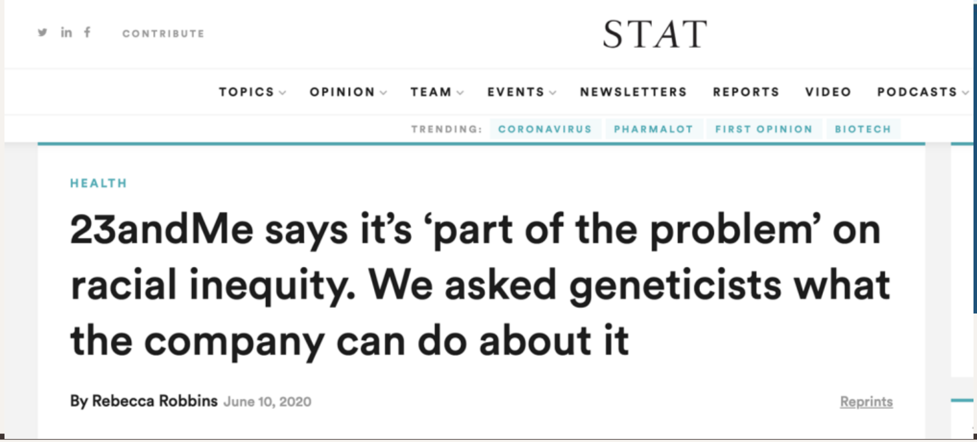


COMMENTARY | VOLUME 177, ISSUE 1, P26-31, MARCH 21, 2019

The Missing Diversity in Human Genetic Studies

Giorgio Sirugo 6 • Scott M. Williams 6 • Sarah A. Tishkoff 6 • Show footnotes

Open Archive • DOI: https://doi.org/10.1016/j.cell.2019.02.048 • Check for updates



HEALTH

23andMe says it's 'part of the problem' on racial inequity. We asked geneticists what the company can do about it

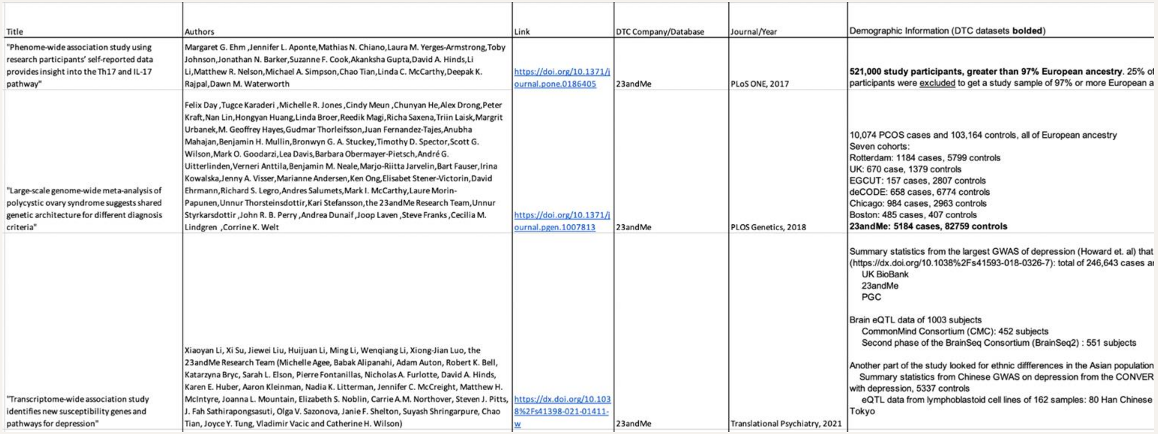
By Rebecca Robbins June 10, 2020

Reprints

Step 2a: Data Collection + Analysis

- **Phase 1**
  - Genetics research articles that specifically used DTC genetic databases
- **Phase 2**
  - Genetics research articles that used non-DTC genetic databases
- **Phase 3**
  - Identified binary questions to assess the literature
  - Ex: *“Did the article address a lack of representation as a limitation?”*

Data Collection Sheet



Title	Authors	Link	DTC Company/Database	Journal/Year	Demographic Information (DTC datasets **bolded**)
"Phenome-wide association study using research participants’ self-reported data provides insight into the Th17 and IL-17 pathway"	Margaret G. Ehm ,Jennifer L. Aponte,Mathias N. Chiano,Laura M. Yerges-Armstrong,Toby Johnson,Jonathan N. Barker,Suzanne F. Cook,Akanksha Gupta,David A. Hinds,Li Li,Matthew R. Nelson,Michael A. Simpson,Chao Tian,Linda C. McCarthy,Deepak K. Rajpal,Dawn M. Waterworth	https://doi.org/10.1371/journal.pone.0186405	23andMe	PLoS ONE, 2017	**521,000 study participants, greater than 97% European ancestry**. 25% o participants were excluded to get a study sample of 97% or more European a
"Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria"	Felix Day ,Tugce Karaderi ,Michelle R. Jones ,Cindy Meun ,Chunyan He,Alex Drong,Peter Kraft,Nan Lin,Hongyan Huang,Linda Broer,Reedik Magi,Richa Saxena,Triin Laisk,Margrit Urbanek,M. Geoffrey Hayes,Gudmar Thorleifsson,Juan Fernandez-Tajes,Anubha Mahajan,Benjamin H. Mullin,Bronwyn G. A. Stuckey,Timothy D. Spector,Scott G. Wilson,Mark O. Goodarzi,Lea Davis,Barbara Obermayer-Pietsch,André G. Uitterlinden,Verneri Anttila,Benjamin M. Neale,Marjo-Riitta Jarvelin,Bart Fauser,Irina Kowalska,Jenny A. Visser,Marianne Andersen,Ken Ong,Elisabet Stener-Victorin,David Ehrmann,Richard S. Legro,Andres Salumets,Mark I. McCarthy,Laure Morin-Papunen,Unnur Thorsteinsdottir,Kari Stefansson,the 23andMe Research Team,Unnur Styrkarsdottir ,John R. B. Perry ,Andrea Dunaif ,Joop Laven ,Steve Franks ,Cecilia M. Lindgren ,Corrine K. Welt	https://doi.org/10.1371/journal.pgen.1007813	23andMe	PLOS Genetics, 2018	10,074 PCOS cases and 103,164 controls, all of European ancestry Seven cohorts: Rotterdam: 1184 cases, 5799 controls UK: 670 case, 1379 controls EGCUT: 157 cases, 2807 controls deCODE: 658 cases, 6774 controls Chicago: 984 cases, 2963 controls Boston: 485 cases, 407 controls **23andMe: 5184 cases, 82759 controls**
"Transcriptome-wide association study identifies new susceptibility genes and pathways for depression"	Xiaoyan Li, Xi Su, Jiewei Liu, Huijuan Li, Ming Li, Wenqiang Li, Xiong-Jian Luo, the 23andMe Research Team (Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic and Catherine H. Wilson)	https://dx.doi.org/10.1038%2Fs41398-021-01411-w	23andMe	Translational Psychiatry, 2021	Summary statistics from the largest GWAS of depression (Howard et. al) that (https://dx.doi.org/10.1038%2Fs41593-018-0326-7): total of 246,643 cases a UK BioBank 23andMe PGC Brain eQTL data of 1003 subjects CommonMind Consortium (CMC): 452 subjects Second phase of the BrainSeq Consortium (BrainSeq2) : 551 subjects Another part of the study looked for ethnic diffferences in the Asian population Summary statistics from Chinese GWAS on depression from the CONVER with depression, 5337 controls eQTL data from lymphoblastoid cell lines of 162 samples: 80 Han Chinese Tokyo

Step 2b: Summary Statistics

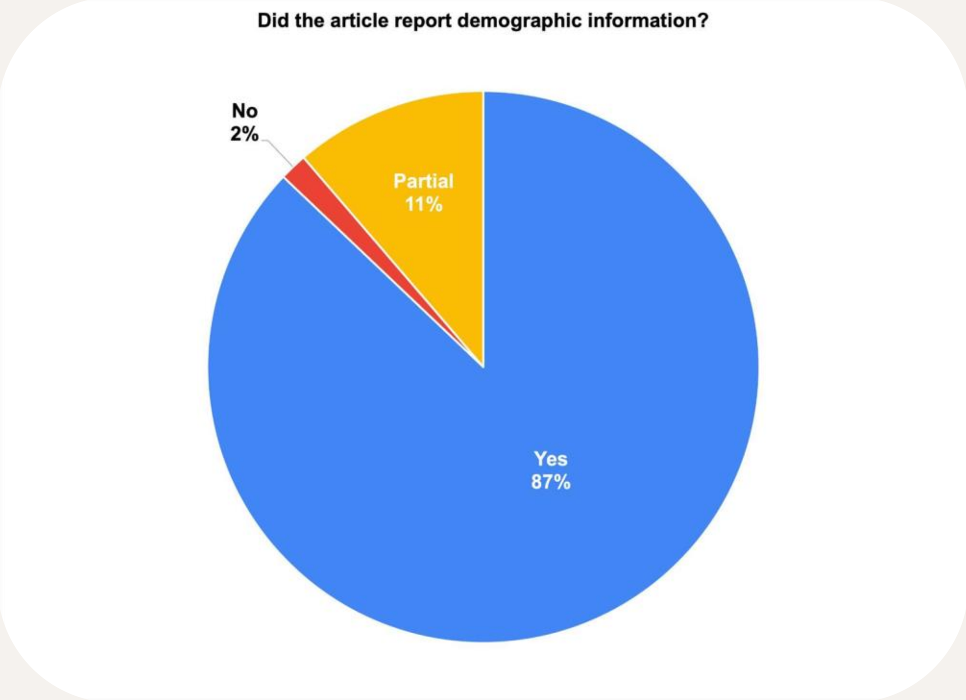
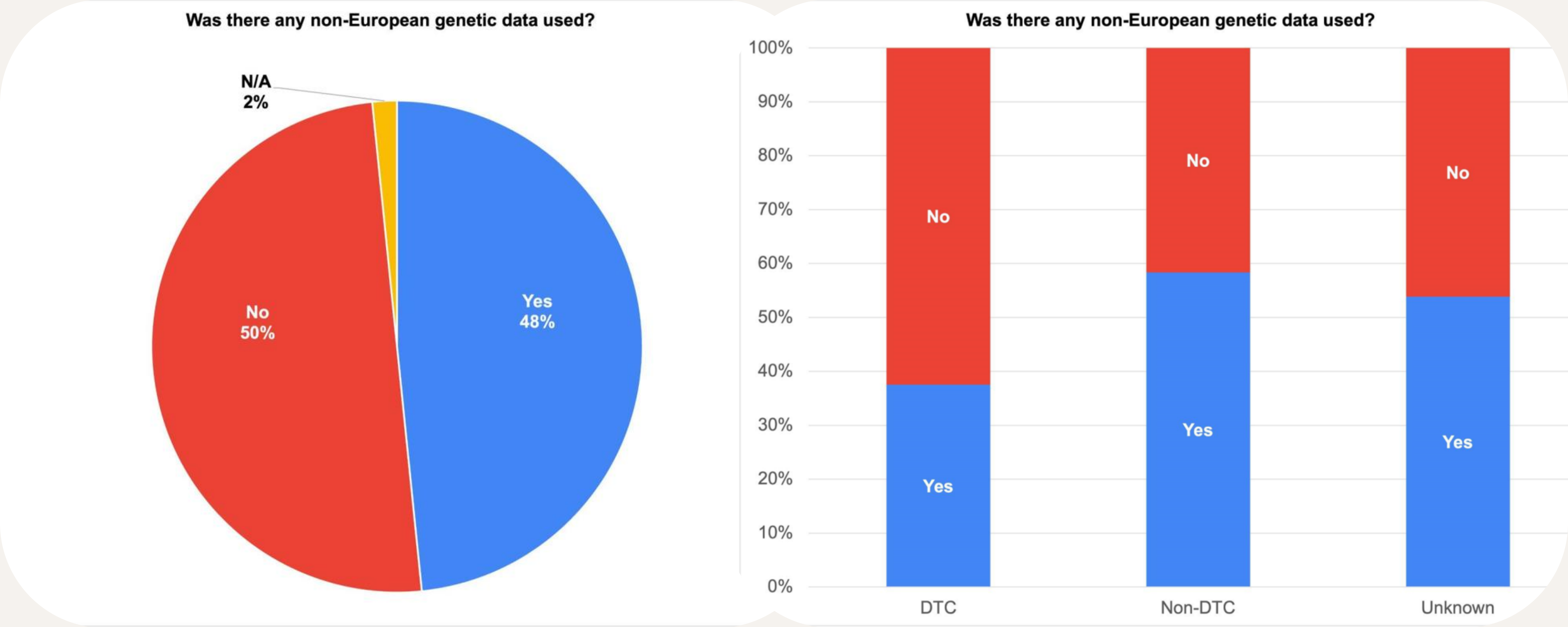
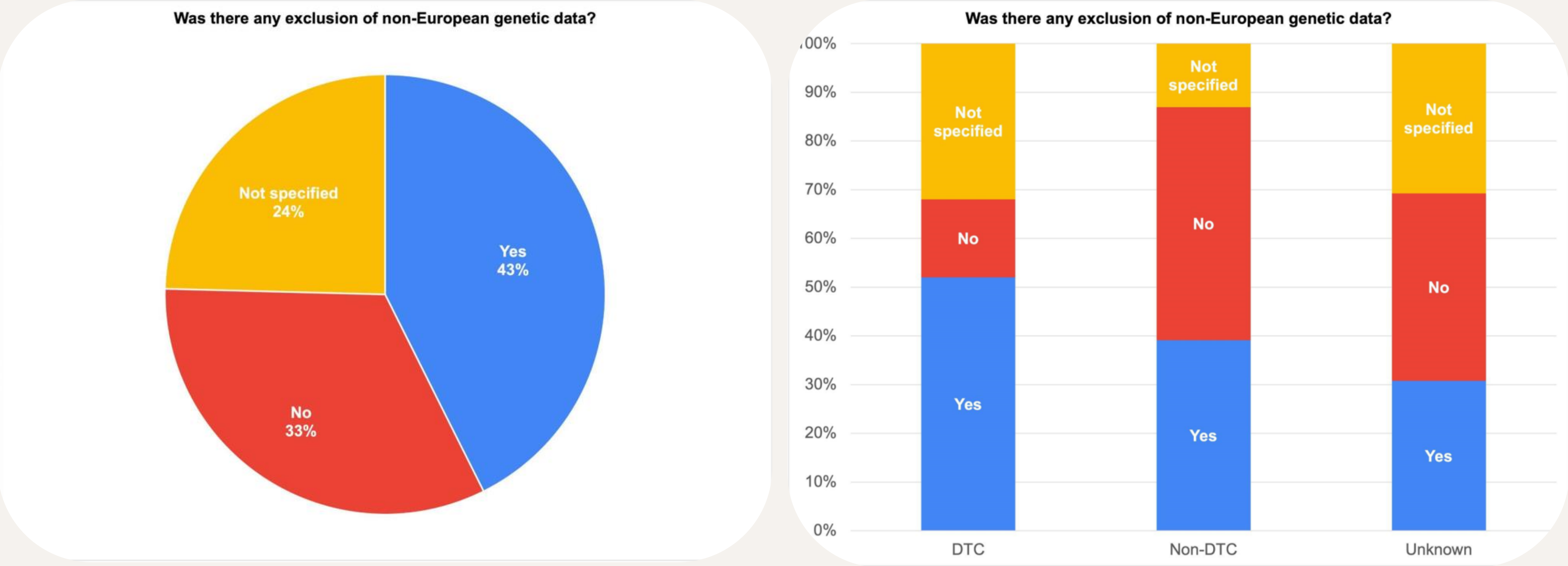


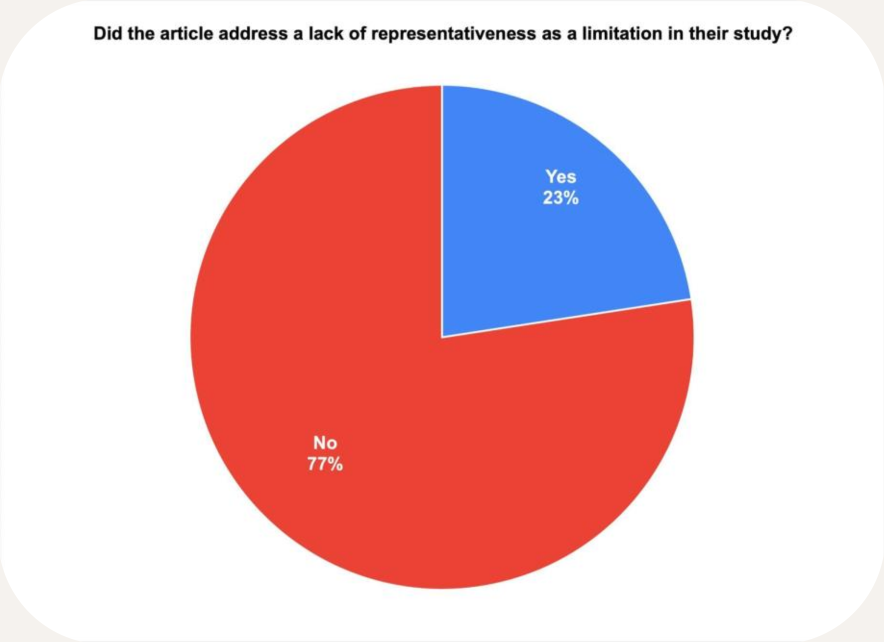
Figure 1: Reporting Demographic Information

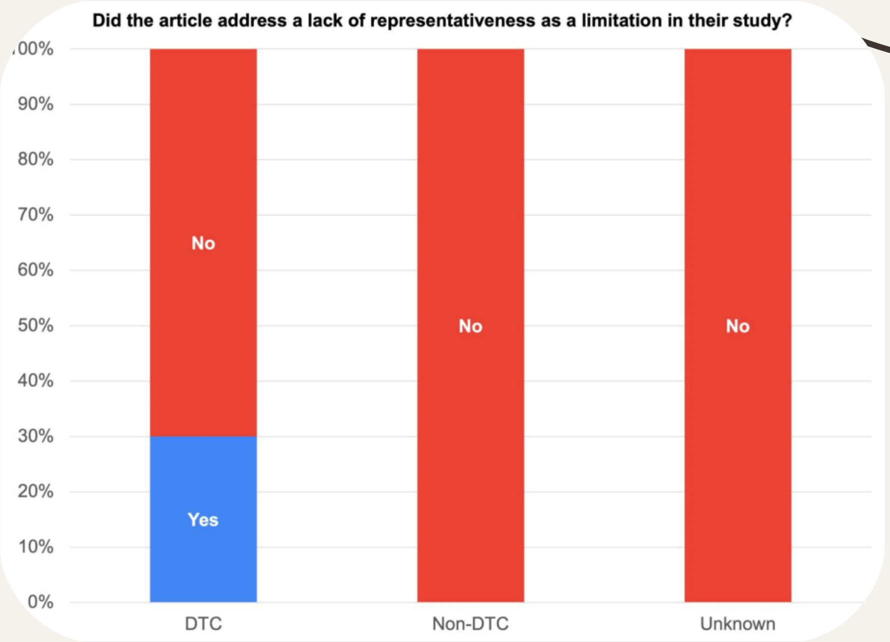


Figures 2 and 3: Use of Non-European Genetic Data



Figures 4 and 5 : Exclusion of Non-European Genetic Data





Figures 6 and 7: Acknowledging the Limitations of Underrepresentation in Genetic Databases

Step 3: Interviews

- Identified DTC officials, federal regulators, and academic researchers to interview
- Currently working on a questionnaire/interview protocol and an IRB waiver
- Sample interview questions:
  - *What is the extent of underrepresentation? What are you currently doing to address underrepresentation in your genetic database/role as a regulator?*
  - *Is there discussion within the genomics research community about reporting racial demographics in papers and/or the limitations of research given the underrepresentation of certain sugroups?*

3

Final Thoughts

Conclusions

Are there specific subgroups whose underrepresentation presents a problem?

1 Underrepresentation does present a problem

How can different applications of genetic databases be undermined by underrepresentation?

2 Underrepresentation and exclusion of non-European ancestry

How common is the underrepresentation of certain population subgroups in genetic databases and genetic research?

3 Mismatch between DTC consumers and general population

4 Mismatch between genetic data included in genetics studies and general population

Future Research

- Consumer/participant perspective
  - Surveys of DTC customers, underrepresented subgroups, general American population
  - Identify barriers to participation in DTC genetic testing
- Possible solutions at different levels
  - Testing different ways to encourage the participation of racial and ethnic minorities in DTC genetic testing
  - Finding out how regulators or publishers can encourage genetics research that prioritizes diversity in genetic datasets

Lessons Learned

Research:

> Literature scoping and literature reviews take time and effort, but are extremely important.

> Research is exciting and non-linear.

Personal:

> Health services research

> Research as a career

Research-wise:

> Importance of broader discussion of genetic databases and genetic research

> Health policy research is extremely interesting

Career-wise:

> Interest in intersection of policy and research

> Interdisciplinary Career

Acknowledgements

Professors Rosoff, Field, and Orlando

SUMR 2021 Cohort

Joanne Levy

Discussion

Any feedback or questions are welcome!

- What would you want to ask genetic database generators or regulators in our upcoming interviews?
- Given the information that we presented, what do you think can be done to increase the representation of minorities in genetic databases?
- Have you ever participated in DTC genetic testing?