Data Equity Policy Brief

Highlights

- The absence of accurate and comprehensive data on marginalized racial/ethnic groups limits our understanding of and ability to address health disparities.

- Addressing limitations in racial/ethnic data collection and reporting requires equity-centered data processes and data systems.

- Healthcare systems and researchers can facilitate implementation of equity-centered data processes by engaging community stakeholders in data collection efforts, from designing instruments to analyzing data.

- Current federal and state policies on reporting patient racial/ethnic data would benefit from updated guidelines on collecting disaggregated racial/ethnic categories.

Introduction

Data equity is an integral component of health equity initiatives for identifying and addressing health disparities in marginalized populations. This policy brief summarizes the problem and a use case on COVID-19 vaccination and provides concrete actions for healthcare systems, researchers, and policymakers to ensure the accuracy and representativeness of racial/ethnic data.

What is the problem?

There are significant limitations in how and what data we collect on race/ethnicity, including missing, incomplete and unknown data, misclassification of racial/ethnic groups, and aggregation of racial/ethnic groups.[1] These limitations shape what we know about each racial/ethnic group in relation to health disparities, the availability of appropriate prevention and intervention efforts to address health disparities, and how resources are allocated. The use of broad racial/ethnic categories (e.g., Asian, Black or African American, American Indian/Alaska Native) hides health needs among small, emerging and marginalized populations, and contributes to ongoing neglect of the diversity within and across these populations.

In a review of national COVID-19 data from December 2021, 34% of racial/ethnic data were missing for COVID-19 cases and 15% of COVID-19 death statistics.[2] Previous analyses of national administrative data and policy research have inferred that minoritized racial/ethnic groups, including Asian Americans and Latino groups, are most likely missing from these data.[2]

In contrast, studies that collect disaggregated racial/ethnic data and offer surveys in multiple languages are able to provide nuanced insights into COVID-19 data. An example of this is
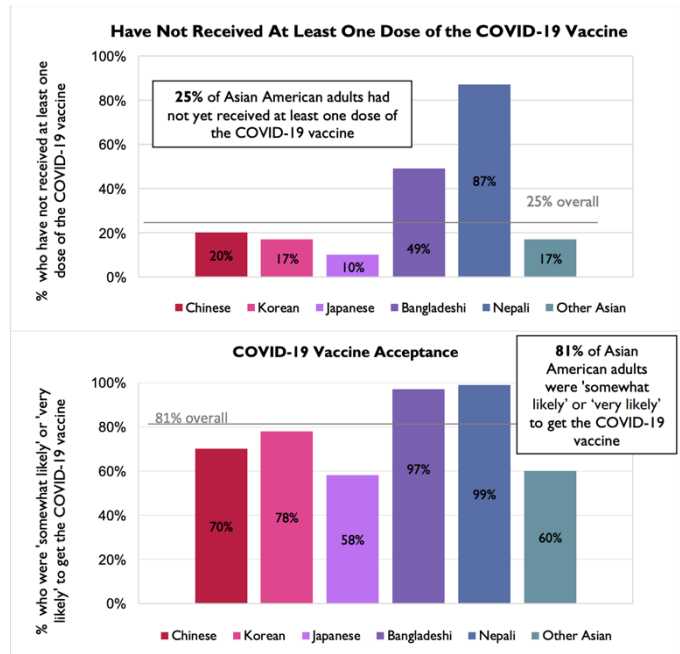

Figure 1: COVID-19 Community Survey

COVID-19 vaccine uptake among Asian American ethnic groups in New York City (NYC). A community needs assessment survey of NYC Asian American residents found low COVID-19 vaccination among Bangladeshi adults (49% received one dose) despite high vaccine acceptance (97%), which suggested access barriers, not vaccine hesitancy, may account for low vaccination rates among Bangladeshi adults (Figure 1).[3] Disaggregated data on race/ethnicity allowed for these nuanced interpretations of vaccination data and evidence to advocate for tailored outreach efforts.

What is data equity and why is it important?

Data equity, as defined by Ponce, Shimkhada, and Adkins-Jackson (2023), refers to the process of engaging marginalized communities in data collection, analysis, interpretation, and dissemination of findings.[4] Another component of data equity includes products, such as data dictionaries, sampling frames to address underrepresentation, and how groups are counted and presented in reports.[4] Incorporating data equity processes and products can facilitate the collection of robust, representative racial/ethnic data, and authentic community involvement in research. The end result, as previously stated, provides comprehensive data to examine more nuanced perspectives on health disparities by race/ethnicity and other marginalized identities (e.g. gender, immigration status, age), among communities who have been historically overlooked or missing in research.

Health disparities among marginalized communities will persist and worsen without intentional efforts to redress the significant limitations in how we collect, interpret, and disseminate data on racial/ethnic groups. Revealing hidden health disparities provides data to advocate for tailored resources and prevention

efforts (e.g. health services, community outreach) to eliminate disparities and address the root causes of health disparities. The health and well-being of marginalized communities cannot be achieved without data equity. These efforts must be conducted in partnership with community organizations and members to facilitate the use of culturally informed methods.

What can we do?

First, acknowledging limitations in current data practices and processes and having ongoing conversations to evaluate how data practices may be harming the communities is critical to advancing data equity. These conversations must involve multidisciplinary stakeholders in the data and research process, as well as community members and the populations who we are collecting data for, to build trust and strengthen relationships. These discussions will then inform why we are collecting data, how best to collect data to reach the populations we want, what data should be collected, and how the data will be used to eliminate health disparities. Second, it is important to understand evolving racial/ethnic identities and which categories and terms are used by participants to identify themselves. Incorporating this information into data collection efforts will ensure participants see questions and response options that reflect their identities. Third, minority racial/ethnic populations have valid concerns regarding how racial/ethnic data will be used as this information could be potentially used to identify them.[5] Explaining how data will be protected to maintain the anonymity of respondents will help in earning the trust of participants. Partnering with racial/ethnic communities in implementing these recommendations will ensure culturally and linguistically tailored approaches are used, specifically appropriate language translation of data collection instruments, soliciting feedback on culturally sensitive methods for data collection, and leveraging community networks for study recruitment.[6]

Below, we present recommendations for healthcare systems and policymakers to improve racial/ethnic data collection methods.

Healthcare Systems and Researchers

- Community trust building to understand community needs and concerns around data collection.
- Include community members and organizations in efforts to improve completeness of racial/ethnic data.
- Partner with community-based organizations that serve minority racial/ethnic populations on data collection efforts and approaches.

- Train providers and staff on obtaining more complete race/ethnicity and demographic data from patients. Develop healthcare system campaigns educating patients on racial/ethnic and other demographic data collection efforts, including why these data are being collected and how this information will be used (Stanford Medicine – "We ask because we care").
- Explain to patients why you are asking about their race/ethnicity information, and how their anonymity will be protected.[5]
- Collect data in multiple languages to ensure all patients are able to understand and respond to questions.[7]
- In the race/ethnicity response options provide disaggregated and culturally relevant racial/ethnic categories. Allow patients to select multiple racial/ethnic categories.[7]
- There are a number of methods to improve the quality of secondary racial/ethnic data, each with strengths and limitations, including expert reviews, name lists, name algorithms, machine learning, and data linkage.[8]
- Oversample certain sub-ethnic populations to ensure representation in data.[7]

Policymakers

- Federal and state policies mandate collection of patient racial/ethnic data. However, the racial/ethnic categories outlined in these policies are broad and there is a need for disaggregation of these minimum reporting requirements. New disaggregated racial/ethnic categories, informed by community organizations and members, should be released by federal and state agencies.

Additional Resources

There are a number of resources for researchers, healthcare systems, and policy makers on data equity and collecting disaggregated racial/ethnic data. Below are a few resources that we have compiled:

- Gould LH, Farquhar SE, Greer S, Travers M, Ramadhar L, Tantay L, Gurr D, Baquero M, Vasquez A. Data for Equity: Creating an Antiracist, Intersectional Approach to Data in a Local Health Department. J Public Health Manag Pract. 2023 Jan-Feb 01;29(1):11-20.
- Policy Link, Counting a Diverse Nation: Disaggregating Data on Race and Ethnicity to Advance a Culture of Health
- Robert Wood Johnson, Get the Full Picture: Advocating for Data Disaggregation by Race and Ethnicity
- United Nations High Commissioner for Human Rights, Data Collection and Analysis
- Urban Institute, Five Ethical Risks to Consider before Filling Missing Race and Ethnicity Data
- National Forum on Education Statistics, Forum Guide to Collecting and Using Disaggregated Data on Racial/Ethnic Subgroups

Conclusion

Data equity is an integral component of health equity initiatives for identifying and addressing health disparities in marginalized populations. As presented in this policy brief there are concrete actions healthcare systems, government agencies, and researchers can take to ensure the accuracy and representativeness of racial/ethnic data. These efforts should be conducted in partnership with community organizations and members to facilitate the use of culturally informed methods.

Data Equity Experts

Matt Chin, MPH **Research Coordinator** **NYU CSAAH**	**Farah Kader, MPH** **Project Coordinator** **NYU CSAAH**	**Lan Đoàn, PhD, MPH** **Assistant Professor** **NYU CSAAH**	**Stella Yi, PhD, MPH** **Associate Professor** **NYU CSAAH**

Contact Us

NYU Center for the Study of Asian American Health
To subscribe to CSAAH's listserv, please email jennifer.wong@nyulangone.org
Twitter: @NYU_CSAAH
https://aanhpihealth.org/

NYU Institute for Excellence in Health Equity
iehegeneraladministration@nyulangone.org
Twitter: @nyugsom_iehe
Instagram: nyugsom_iehe

Authors: Naheed Ahmed, Matt Chin, Lan Đoàn, Lindsey Osagiede, Ida Myoung, Antoinette Schoenthaler, and Nadia Islam

Design and Layout: Lindsey Osagiede, Naheed Ahmed, and Nadia Islam

References

1.   Chin M, Doan L, Chong S, Wong J, Kwon S, Yi S. Asian American subgroups and the COVID-19 experience: what we know and still don't know. *Health Affairs Blog*. 2021;10
2.   Yi SS, Kwon SC, Suss R, et al. The Mutually Reinforcing Cycle Of Poor Data Quality And Racialized Stereotypes That Shapes Asian American Health: Study examines poor data quality and racialized stereotypes that shape Asian American health. *Health Affairs*. 2022;41(2):296-303.
3.   Doan L, Wyatt L, Cown C, et al. *New York City COVID-19 Community Health Resources and Needs Assessment*. 2022. https://aanhpihealth.org/wp-content/uploads/2022/03/Asian-American_NYC-COVID-CHRNA_Advance-Findings.pdf
4.   Ponce NA, Shimkhada R, Adkins-Jackson PB. Making Communities More Visible: Equity-Centered Data to Achieve Health Equity. *Milbank Q*. Apr 2023;101(S1):302-332. doi:10.1111/1468-0009.12605
5.   Kader F, Doan L, Lee M, Chin M, Kwon S, Yi S. Disaggregating race/ethnicity data categories: Criticisms, dangers, and opposing viewpoints. 2022.
6.   Crawford M, Flores F, Kolak M, Nelson A, Simmons M. Importance of Community Engagement in Data Decision Making. Mary Ann Liebert, Inc; 2023:75-86.
7.   Wyatt L, Russo RG, Kranick J, et al. *2012-2018 Health Atlas for Asian Americans, Native Hawaiians, and Pacific Islanders: A comprehensive look at AA and NH&PI health in the U.S.* .
8.   Chin MK, Đoàn LN, Russo RG, et al. Methods for Retrospectively Improving Race/Ethnicity Data Quality: A Scoping Review. *Epidemiol Rev*. Apr 12 2023;doi:10.1093/epirev/mxad002