UNDERSTANDING THE ROLE OF DATA AND RESEARCH IN PERPETUATING INEQUITIES IN HUMAN SERVICES

Introduction

Data and research can be useful tools for understanding inequities in interactions with human services systems for Black, Indigenous, and people of color (BIPOC); lesbian, gay, bisexual, transgender, queer, and other (LGBTQ+) populations; and individuals affected by disability status. However, incomplete or incorrect data, poorly designed analyses, and/or biased interpretation can further inequities in human services provision by reinforcing policies and practices that perpetuate systemic racism, male and heterosexual power, ableism, and other entrenched inequities. Centering equity in data collection and use can give policymakers the information they need to address inequities in human services and meet the needs of historically marginalized populations.

This memo offers a brief description of inequities in data collection, data integration, and the larger research process for historically marginalized populations in the United States. It then suggests strategies for researchers to increase equity in data-driven research through changes to data collection and analysis.

Inequities in Data and Research

Inequities can be inserted or exacerbated in various points of the research process from data collection and defining subgroups, to data interpretation, analysis, and communication.

Data Collection

Much of the current data collection related to race and ethnicity, LGBTQ+ individuals, and disability status —if such data is collected at all —fails to capture the information needed to fully understand the inequitable experiences of these populations. Research on historically marginalized populations’ participation in human services programs primarily uses either survey data (i.e., self-reported data collected from targeted or general populations) or administrative data (i.e., data collected for programmatic reasons from caseworkers or from participants’ self-reported information and application data). Both methods of data collection have implications for the quality of subsequent research and analysis of equity in human services systems.

Key Findings

- In the context of research on human services participation, inequities experienced by historically marginalized populations are found at various levels of the research process, including data collection, data analysis, and the interpretation of results.
- Individual and systemic biases in data and research limit knowledge on the needs of, and best practices for serving, historically marginalized communities.
- Promising practices to increase equity in data and research include:
  - Centering equity throughout the data lifecycle by including a diverse research team and community stakeholders in the research process;
  - Updating data collection processes to collect data that more accurately represents and includes the experiences of historically marginalized communities; and
  - Using an equity lens when interpreting and communicating data by contextualizing and disseminating findings in ways that are accessible to different audiences including the impacted communities.

This memo was supported by Cooperative Agreement number AE000103 from the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. The opinions and conclusions expressed herein are solely those of the author(s) and should not be construed as representing the opinions or policy of any agency of the federal government.
Poorly designed and implemented surveys may lead to the underrepresentation and continued marginalization of members of certain populations, particularly for LGBTQ+ individuals and people with disabilities. Some surveys do not attempt to collect relevant data on subgroups; for example, many population surveys do not collect demographic data about sexual orientation and gender identity. This can lead to the underrepresentation of LGBTQ+ individuals in survey results and limit researchers’ capacity to understand the experiences of LGBTQ+ individuals (Baker et al., 2016; Burwick et al., 2014). Even when surveys include relevant demographic questions, LGBTQ+ populations may not feel comfortable responding, resulting in low response rates. For instance, self-report data often underreports the LGBTQ+ population, as LGBTQ+ individuals who have experienced discrimination due to their sexuality or gender identity may not respond to surveys out of fear that disclosure of their identity will result in further discrimination (Baker et al., 2016; Burwick et al., 2014; Ragg et al., 2006; Sears & Mallory, 2011). Additionally, some data collection efforts may exclude members of certain populations due to non-inclusive word choice; using binary and gendered language such as “husband” and “wife” and “male” and “female,” for example, may result in the underrepresentation of LGBTQ+ populations in survey data (Averett et al., 2017; Baker et al., 2016; Duke & McCarthy, 2009; Irvine & Canfield, 2016). Finally, some surveys may exclude individuals with disabilities who are institutionalized by recruiting participants only from residential households via phone or internet; this practice may lead to underestimations or misrepresentations of people with disabilities, as they are overrepresented in the correctional institutions and long-term care facilities that house most institutionalized people (Harrell, 2017; Wittenburg & Nelson, 2006).

Administrative data may similarly lead to the underrepresentation and continued inequities among members of historically marginalized populations. While often rich in information, administrative data is collected for human services program administration purposes and may not consistently collect information needed for research, may have inaccurate or incomplete data, and may reflect only the experiences of individuals with access to human services programs. Administrative data may also reflect systemic biases as well as the implicit and explicit biases of the caseworkers who enter such data. For example, over-policing in low-income neighborhoods, which is an indicator of systemic racism, contributes to higher rates of child welfare referrals for families of color compared to White families (Dettlaff & Boyd, 2021; Ellis, 2019; Kolivoski et al., 2018). In addition, caseworkers make explicit—often subjective—decisions in the details they choose to document and the ways in which interactions with individuals are documented (Kennedy, 2021). Further, administrative data often does not collect information on all populations who face the pernicious realities of systemic discrimination. For instance, as discussed in the “Defining Groups” section below, comprehensive data collection does not often occur across human services programs on LGBTQ+ populations, Asian Americans, and Native Americans, among other groups (Baker et al., 2016; Burwick et al., 2014; Hawn Nelson et al., 2020).

Administrative data fields may be inaccurate, missing, or may be subject to unwritten rules that affect data quality. For example, because administrative data is entered by caseworkers or by participants with their own subjective goals and motivations, the quality of administrative data is often uneven (Allard et al., 2018). Participants may also enter only the minimum needed to receive benefits or may intentionally leave fields blank if they do not feel comfortable disclosing information (e.g., gender identity, sexual orientation, or disability status) because this data is, by nature, not anonymous. Written and unwritten caseworker rules about how and when to enter data, and performance expectations for caseworkers to perform their work quickly, can also affect data quality. For example, caseworkers may omit demographic characteristics such as level of education, sexuality, or ability, and instead opt to spend their limited time on data fields that are required for a participants’ program eligibility such as income and employment status (Weigensberg et al., 2014). Additionally, staff may have different interpretations of data fields. For example, data on Temporary Assistance for Needy Families (TANF) recipients’ and applicants’ child care arrangements are particularly prone to misinterpretation by program staff (Hahn et al., 2016).
Defining Groups

Decisions on how to define groups can lead to inaccurate counts of various subgroups. For example, because there is no universally accepted definition of disability, researchers disagree over how to operationally define disability, leading to challenges in conducting research on this widely heterogeneous population (Meyer & Mok, 2013; Palmer & Harley, 2012). In studying the experiences of people with disabilities, researchers often define disability using respondent-reported disability status—usually as “yes” responses to questions about limitations or impairments—or by whether respondents receive particular benefits such as Social Security Disability Insurance (SSDI) or Supplemental Security Income (SSI) (Meyer & Mok, 2013; Palmer & Harley, 2012; Theis et al., 2019). These measures of disability may under- or mis-report the true number of people with disabilities, or certain types of disabilities, because what constitutes “disability” may vary across contexts, individuals, and cultures and because not all individuals with disabilities receive benefits such as SSDI or SSI (Burkhauser et al., 2002; Meyer & Mok, 2013; Theis et al., 2019). Additionally, data collection on LGBTQ+ individuals that fails to provide appropriate terminology and groupings may also be unreliable (Baker et al., 2016). Surveys or administrative data that do not provide culturally specific terms or open-narrative text options for respondents to indicate appropriate terminology will also fail to collect reliable and representative information (Baker et al., 2016).

The aggregation and disaggregation of data to assign or reassign group membership among historically marginalized populations can make it difficult to understand the experiences of individuals within population subgroups. Data collection methods used to identify race and ethnicity are often aggregated by oversimplified categories, such as “Hispanic/Latino” and “Asian American,” which helps ensure large enough sample sizes to include the experiences of smaller racial and ethnic groups in research but can lump together groups of culturally diverse people and obscure nuanced group identities or variations in experiences such as countries of origin, immigration statuses, and/or home languages. This practice may conceal differences and inequities experienced by diverse members of subpopulations (Byon & Roberson, 2020; Teranishi et al., 2015). For example, in aggregated data regarding the broad category of Asian Americans, researchers are unable to distinguish between experiences among immigrant and non-immigrant individuals or between Asian Americans from different geographic regions, ultimately resulting in data that have an over representation of Asian American families on both the top and the bottom of the poverty distribution and obscuring the needs of lower-income Asian American families (Byon & Roberson, 2020; Teranishi et al., 2015; Kauh et al., 2021). Another example is the aggregation of data on Native American populations, which results in a loss of nuance on both group and individual characteristics. Across the 562 federally recognized—and approximately 400 non-federally recognized—Native American nations in the United States, each has unique ethnic, cultural, and linguistic practices (Kauh et al., 2021; U.S. Government Accountability Office, 2012). In contrast, the use of disaggregated data presents other challenges for research on often marginalized populations. For example, the disaggregation of data on sexual orientation and gender identity often results in small sample sizes, likely due to underreporting, making it difficult to analyze and report on findings for specific LGBTQ+ communities (e.g., transgender populations) without risking privacy concerns (Badgett et al., 2019; Kauh et al., 2021; Schwabish & Feng, 2021).

Data Interpretation, Analysis, and Communication

Results from human services research regarding people belonging to historically marginalized groups must be interpreted with attention to the quality of the data used to inform findings and the fact that sources of data on these populations may be skewed or biased in the ways discussed above.

Even when data are complete and unbiased, researchers must be careful to identify root causes for observed outcomes using administrative and survey data (Andrews, 2020; Hawn Nelson et al., 2020). For example, researchers who observe differences in employment rates by race among TANF recipients may conclude that Black TANF recipients are less likely to work than White recipients. Such an observation may ignore the
consequences of structural racism, such as employment discrimination and unequal access to services such as childcare, work readiness programs, and education and training programs (Hawn Nelson et al., 2020; Monnat, 2010). Findings that do not explore the causal factors of observed outcomes risk perpetuating racist stereotypes and disparate outcomes in human services systems.

Some research methodologies, such as predictive analytics and algorithmic assignment to treatment conditions, also present unique challenges. These approaches have been recently adopted in some human services programs with the intent of better targeting interventions and potentially reducing individual discretion and biases in service provision. However, the use of these techniques may lead to biased results when the input data, like criminal history or prior reports of child maltreatment, are biased due to factors rooted in systemic racism (Drake et al., 2021; Hawn Nelson et al., 2020; Keddell, 2019; Kennedy, 2021).

Findings from human services research about historically marginalized groups must also be communicated in ways that are accessible to broad audiences and the impacted communities (Hawn Nelson, 2020). When communicated effectively, research can influence policies, programs, and funding (Andrews, 2020; Gross, 2020). However, without careful consideration to communication, research findings may be misinterpreted and perpetuate harmful racist stereotypes (Gross, 2020). Researchers should avoid jargon and technical terms, use person-first language, explicitly include relevant contexts, and precisely define the groups being studied to ensure that findings are not misinterpreted and are understandable by a broad audience (Gross, 2020; Hawn Nelson, 2020).

**Mechanisms to Disrupt Inequity in Data-Driven Research**

To ensure accurate and unbiased findings, equity must be centered throughout the full “data life cycle” (Hawn Nelson et al., 2020). Detailed and deliberate planning for data collection, analysis, and communication is the necessary first step in the data life cycle. At the onset of a project, a workgroup of stakeholders should be convened to develop a project plan integrating a clear mission with ethical guidelines for data use and an understanding of historical contexts of the relevant issues (Hawn Nelson et al., 2020). Stakeholders should include diverse perspectives from the community in planning meetings and presentations on research findings, including inviting individuals with lived experiences and individuals with knowledge of the community or population’s history. Additionally, using a mixed-methods approach that includes focus-groups and interviews with individuals with lived experiences in tandem with quantitative data can help researchers better understand the lived experiences of program participants (Hawn Nelson, 2020). Stakeholders should also engage community members when researching the impacts of historical systems of oppression, understanding the extent of past harms, and outlining future opportunities (Balu, 2021; Hawn Nelson et al., 2020).

Research team members should come from multiple backgrounds and cultures and project roles should be assigned equitably; diverse teams benefit from multiple perspectives and sensitivities, which can improve communication, innovation, and productivity, as well as improve engagement with historically marginalized populations (Andrews, 2020). Team members must also examine the biases they bring to the research; while researchers hope to be value-neutral and objective, everyone has values and experiences that affect the way they ask questions and conduct research (Parekh et al., 2019). It is also important for research teams—who are often paid—to be considerate about the time involved when engaging community stakeholders—who are often unpaid—with project work. When community members must take time away from paid employment and family obligations to be involved in research projects, it could potentially amplify inequities rather than alleviate them (Andrews, 2020; Balu, 2021). Finding equitable ways to engage community members, such as paying community members for their time, is important because their participation in data analysis informs more accurate conclusions and provides more useful and actionable recommendations by highlighting individual, community, social, and historical perspectives of the data (Hawn Nelson et al., 2020).
When complete, findings should be contextualized and disseminated in various ways to provide informational products that are accessible to different audiences such as briefs, interactive documents, websites, social media content, posters, and infographics. The findings should include acknowledgements of the structural racism or other biases that are embedded within the data (Gross, 2020; Hawn Nelson et al., 2020). Dissemination materials should also include individual stories, when possible, to contextualize the findings with details of lived experience, and include actions that human services agencies should take to improve the lives of the communities represented in the data (Hawn Nelson et al., 2020). Analyses should also consider potential intersectional variables across populations, such as looking at race by gender or disability status by race, to highlight the experiences of subpopulations and to further examine findings (Gross, 2020; Hawn Nelson et al., 2020).

**Strategies for Promoting Equity in Data and Research on Race/Ethnicity**

Researchers can reduce risk of continued bias by using an equity lens to understand the value of various data types (e.g., administrative, survey, interview, etc.) and considering the need to contextualize findings (Hawn Nelson et al., 2020). Disaggregating data by race/ethnicity can illuminate inequities; public institutions and agencies should be required to report and share data in a way that allows the identification of underserved groups so policymakers can enact evidence-based policies to address such inequities (Byon & Roberson, 2020; Teranishi et al., 2015). Disaggregated data can also help policymakers more effectively target resources and promote social justice (Brown et al., 2021; Byon & Roberson, 2020; Teranishi et al., 2015). Ideally, agencies should collect data that can support analyzing differences in human services program delivery and outcomes and explore analytic techniques to reliably estimate differences when existing data are insufficient (McDaniel et al., 2017). However, disaggregated data also comes with its own risks that could violate informed consent, compromise individual privacy, or lead to misleading conclusions. Releasing administrative data can disproportionally impact BIPOC communities; as such, agencies need to develop data access practices that include protections to ensure privacy and confidentiality of information (Hawn Nelson et al., 2020; Randall et al., 2021).

**Strategies for Promoting Equity in Data and Research on LGBTQ+ Populations**

Data collection must utilize gender- and relationship-neutral language to collect representative data regarding LGBTQ+ populations (Baker et al., 2016). Neutrality in gender and relationship language (e.g., “spouse” or “partner” rather than “husband” or “wife”), in surveys and administrative data allows researchers to collect more and better data about LGBTQ+ populations and highlight support for LGBTQ+ populations (Baker et al., 2016; Averett et al., 2017). In addition, information fields that include an option with open-text response (i.e., fill-in-the-blank) options allow individuals to use terms they are comfortable with to describe themselves if such terms are not already listed (Baker et al., 2016). Inclusive language options on forms, such as collecting “preferred” name and giving respondents the opportunity to express their gender pronouns, support the inclusivity of LGBTQ+ populations and provide useful data for research that identifies the uniqueness of LGBTQ+ experiences while interacting with human services programs (Baker et al., 2016).

 Continued normalization of diversity in LGBTQ+ identities will likely result in more reliable data on LGBTQ+ populations because individuals may feel more comfortable self-reporting their identity. To be ethical stewards of personal information, data collection systems involving sexual orientation and gender identity status should be careful to ensure privacy of any disclosed information and to be transparent regarding the mechanisms in place to do so (Baker et al., 2016). Data fields regarding sexual orientation and gender identity should also be voluntary to protect those who are not yet ready to disclose (Ragg et al., 2006).
Strategies for Promoting Equity in Data and Research on People with Disabilities

Improving data collection and increasing research on (and with) people with disabilities will likely result in more reliable data on these populations. Disability-oriented planning, policymaking, and public health programming could be improved if federal data collection on people with disabilities used a standardized definition of disability (Palmer & Harley, 2012; Theis et al., 2019). Additionally, better data on children with disabilities in the child welfare system (e.g., prevalence, characteristics, out of home placements) could be collected if caseworkers had more and better training on disability-related concerns, specifically in identifying disabilities in children (Lightfoot et al., 2011). To promote a better understanding of families’ disability-related concerns, the federal government should consider establishing an Interagency Committee on Parents with Disabilities1 (National Council on Disability, 2012). Congress and the federal government should also consider further funding research on parents and children with disabilities and their families and consider developing initiatives for federal agencies to collect effective and comprehensive data on parents and children with disabilities and their families to better understand their characteristics and needs (National Council on Disability, 2012). In addition, state disability agencies, mental health providers, and other systems serving people with disabilities should aim to collect data on parental status of participants (Johnson et al., 2012; National Council on Disability, 2012).

Conclusion

Reliable data and thoughtful research methods are critical to understanding the current inequities experienced by historically marginalized populations participating in or trying to access human services programs. However, inaccurate, inconsistent, or biased data, along with flawed subgroup definitions and uninformed data interpretation and analysis that does not identify root causes for observed outcomes, limit researchers’ ability to recommend mechanisms for disrupting entrenched inequities. Federal actions to collect more and better disaggregated data and combat systemic biases in data collection will help to ensure the visibility of these populations in research and policymaking.

---

1The committee would include the National Council on Disability, Health and Human Services/Administration for Children and Families, Department of Labor, Department of Justice, Social Security Administration, Department of Transportation, Housing and Urban Development, Department of Education, and Veterans Affairs department (National Council on Disability, 2012).
References


