



RESEARCH REPORT

# Ethics and Empathy in Using Imputation to Disaggregate Data for Racial Equity

## Recommendations and Standards Guide

*K. Steven Brown*

*Lesleigh Ford*

*Shena Ashley*

*with Alena Stern and Ajjit Narayanan*

*July 2021*



## ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

# Contents

<b>Acknowledgments</b>	<b>iv</b>
<b>Ethics and Empathy in Using Imputation to Disaggregate Data for Racial Equity</b>	<b>1</b>
What Is Imputation and How Can It Be Approached with Ethics and Empathy?	2
Who Should Be Involved in Developing and Applying Ethically Imputed Disaggregated Data?	7
Deciding When Imputation Is the Right Approach for Disaggregation	14
How to Take an Ethical Approach to Imputation	17
Conclusion	22
<b>Notes</b>	<b>23</b>
<b>References</b>	<b>24</b>
<b>About the Authors</b>	<b>26</b>
<b>Statement of Independence</b>	<b>27</b>

# Acknowledgments

This report was funded by Robert Wood Johnson Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute’s funding principles is available at [urban.org/fundingprinciples](http://urban.org/fundingprinciples).

The authors would like to thank Megan Randall for excellent project management over the course of this project; Graham MacDonald and Kathy Pettit for reviewing and providing helpful feedback; and Jahnavi Jagannath, Jacqueline Rayfield, and Yipeng Su for their broader contributions to this project. We also would like to thank the interviewees for this project for their time and insights and for informing the thinking in this report. The list of interviewees and additional thanks are provided in the associated and separately published landscape scan.

# Ethics and Empathy in Using Imputation to Disaggregate Data for Racial Equity

Too frequently, public datasets that are regularly used to inform policy and practice have published aggregated data that do not allow for racial and ethnic disparities in access and outcomes to be made visible. To better understand communities' policy needs and concerns, we argue that data collectors and data owners, researchers, and policymakers should increase data disaggregation, which is when data report on key variables and outcomes by detailed subgroups. In this report, we focus on the need to disaggregate data for racial and ethnic subgroups including Black, Latine, Asian American, Pacific Islander, and Native American people, although data would ideally provide even more specificity, including national and ethnic origin.

There are strong ethical and practical reasons for data disaggregation. People can use disaggregated data to provide clarity in places and policy areas where disparities have been suspected but not identified, can allow people to better see themselves in the data so they can understand their challenges and opportunities and advocate for themselves, and can promote a more just and equitable distribution of resources.

The lack of appropriately disaggregated data has harmed communities of color. The COVID-19 pandemic has brought greater and long-overdue attention to the need for real-time disaggregated data.<sup>1</sup> Consequently, more data scientists and researchers, like ourselves, are focusing on data disaggregation (Markle Foundation 2021). As this practice grows,<sup>2</sup> we believe it should be accompanied by a robust discussion of the ethical risks associated with the application of different methodological strategies, one that focuses on using imputation to better understand racial and ethnic differences in data where racial and ethnic information is largely missing or absent. To that end, we have developed, through our experience and interviews with the field and a case study (Stern and Narayanan 2021), a set of recommendations and standards for data analysts and data consumers around how to impute information on race and ethnicity ethically and with empathy.

This effort is part of a project being led by the [Racial Equity Analytics Lab](#), which seeks to equip today's change agents with data and analyses to advance social and economic policies that help remedy persistent structural racism. We believe more thoughtful, race-conscious policies and practices have

the power to forge new avenues of opportunity and prosperity for Black, Latine, Native American, Asian American, Pacific Islander, and other communities, and that timely, reliable data are essential for designing race-conscious solutions—as well as for holding decisionmakers and institutions accountable for choices that perpetuate oppressive systems.

This report has four main sections. The first provides an overview of imputation, why it is an important and needed tool for disaggregated data and race-conscious policymaking, and how to approach it with ethics and empathy. The second explores several key questions researchers, government officials, community leaders, and other stakeholders should ask when considering and applying imputed data disaggregated by race and ethnicity. These questions concern who should be involved, if and when imputation is the right approach to disaggregating data and collecting more detailed information on race and ethnicity, and how to conduct an imputation for disaggregation with a more explicit focus on ethical application. The third considers how to determine whether imputation is the right approach to disaggregation, and the fourth outlines recommendations and standards for the field based on lessons we learned from the landscape scan and the case study. Lastly, though this report focuses on using imputation to better understand disparities in race and ethnicity, the lessons here could be more broadly applied to other areas with limited visibility or disaggregation for gender, age, disability, and other characteristics.

## What Is Imputation and How Can It Be Approached with Ethics and Empathy?

In the field of statistics, imputation has long been used to fill in information missing from a dataset (Rubin 1996). To avoid generating statistically biased estimates that can produce invalid, inaccurate results, researchers can use different approaches to solve for missing data by either removing the data completely under certain conditions or using imputation as a way to develop reasonable proximates to fill in information for the missing data (Donders et al. 2006). Although a range of specific methodologies are used throughout statistics and data science (e.g., single versus multiple imputation, hot deck versus classification and regression trees), they involve the same overall approach: taking information observed with confidence and relative completeness in a dataset or sample population and using those data to make inferences about other people or units in the data who otherwise look similar, but are missing some critical information.

## Using Imputation to Disaggregate Data

Although imputation is primarily used to fill in data for incomplete variables within a dataset, it can also be used to integrate or append datasets or even create new variables—like race and ethnicity. For instance, to produce a set of probabilities that estimate a person’s likely race or ethnicity, a team led by the RAND Corporation designed an approach called Bayesian Improved Surname Geocoding and drew people’s surnames and information about the racial and ethnic compositions of their residential neighborhoods from the census (Elliott et al. 2009; Fremont et al. 2016).<sup>3</sup>

Imputation can be a relatively efficient way to address missing data or add new variables, such as race and ethnicity. For such important identifying characteristics, having complete self-reported information is the gold standard, and collecting new data to obtain that information when it is missing or otherwise absent would be ideal. But collecting new data is often not practical because of the financial costs or the time constraints and urgency of informing an active policy debate or proposed legislation. And in some cases, collecting or obtaining data on race and ethnicity faces even higher barriers, due either to long-standing, institutionalized practices (e.g., race not being collected in IRS/tax data) or to valid legal or regulatory barriers to capturing that information (e.g., race not collected in credit data). Given researchers often work with secondary and administrative data sources they neither collected nor have full control over, imputation, data integration, and matching practices more generally may often be not only the fastest and most cost-effective option, but the only option for identifying and understanding potential racial and ethnic disparities in data.

Datasets that can provide important insights on the presence and complexity of disparities or inform policy may contain no information about race. Given the relationship between race and ethnicity and outcomes across many policy domains, one could argue that information on race and ethnicity is “missing” even in datasets for which variables of race and ethnicity were not originally included in the forms or questionnaires. Imputation methods such as Bayesian Improved Surname Geocoding and multiple imputation that fill in missing information are powerful tools for providing access to key data sources where information on racial and ethnic identification is incomplete or altogether absent. Such methods have been applied successfully by researchers in the federal government and other institutions, including for collecting data about COVID-19. (Anson-Dwamena, Pattah, and Crow 2020; CFPB 2014; Fremont et al. 2016). When applied accurately and ethically, these methods and the data they provide can help researchers, policymakers, and advocates better understand the depth of racial and ethnic disparities and develop policies and practices to address them. Importantly, though, these statistical and data-science techniques involve risks, and although the literature discusses ethical concerns within advanced data analytic approaches generally and for algorithmic applications more

specifically (AECF 2020; Lee, Resnick, and Barton 2019), it provides much less guidance on addressing these concerns when using and applying imputed race and ethnicity data for policy application.

## **Approaching Imputation Ethically and with Empathy**

Ethical imputation requires weighing the risks and benefits of adding new information to existing data, and engaging with the relevant communities to consider the harms that applying these data could have for groups and people whose identities are more represented and made more visible in the data. To be clear, in discussing ethical imputation, we do not mean to imply that researchers who have used imputation to improve or incorporate information on race and ethnicity have not weighed the ethical implications of their work. Indeed, many have. Rather, we argue that all researchers who use imputation to develop variables to analyze group-based differences should consider its ethical impacts and the need for making engagement and accountability as central to their work as data preparation and methodological rigor.

Two of the things we emphasize are the need to address possible harms to people and communities and the need for empathy toward their concerns. Researchers, data scientists, and government officials have been increasingly identifying and coalescing around a set of key potential risks (detailed below) associated with data-integration efforts and tools and with attempts to achieve data equity more broadly.

Ethical concerns around data integration and data analysis primarily involve the following (these risks are also commonly mentioned in literature on data matching and integration):

- misrepresenting communities by using inappropriate input data and/or methodology
- using data for purposes that harm people of color
- violations of privacy and risk of reidentification, particularly for smaller or less-data visible groups of peoples (such as Native Americans)
- not accounting for or not providing informed consent
- excluding people and communities of color from ownership of their data and from decisions on research process and methods

Sophisticated methods for generating or appending racial and ethnic identifiers—such as machine learning and data linkage and integration—share many of these broader challenges but also pose distinct risks. Although the body of work examining harms and harm-mitigation approaches for

algorithms and big data in general is growing,<sup>4</sup> it lacks specific guidance on the ethical risk areas that data scientists, statisticians, and researchers encounter when using imputation, matching, or related methods to fill missing race and ethnicity data (AECF 2020; Hawn Nelson, Jenkins, Zanti, Katz, and Berkowitz et al. 2020). In our research attempting to highlight the ethical checkpoints of imputation (Randall, Stern, and Su 2021), we have seen a lack of practical guidance in the literature and in the field about how to address the following issues:

- an increased chance of the findings being inaccurate, given the imputed data are estimates, rather than directly observed or first-person-reported data
- balancing accuracy with fitness of the imputation for the use case
- difficulty engaging people in explaining the complexity of the methods and analytic process
- a lack of accountability structures for researchers and data scientists outside of institutions with review boards (imputations can be done with secondary data without checkpoints or data use agreements)

There are also gaps in the field around the implications of using imputation when trying to embed empathy for the people the imputation is intended to represent. The goal of imputation is to fill a data gap. If imputation is being used to better understand racial distributions or disparities (beyond just filling missing data), then how have the researchers or analysts considered who is looking to fill these data gaps? How have they engaged these people and incorporated their interests and goals into the analytic process?

Alongside the more commonly voiced ethical risks (e.g., violation of privacy), community leaders, advocates, and a subset of researchers have also called for increased and more deliberate community engagement and community influence, with some calling for full community ownership over the data and entire research process (Milner and Traub 2021). Researchers are using community-engaged methods more to involve community members directly in the development of research questions and use cases and provide regular checkpoints between quantitative analysts and community representatives to talk about progress and improve processes as necessary. In community-engaged methods, involved and empowered community members help interpret the results and decide on the appropriate applications.<sup>4</sup>

Researchers who rely on quantitative methods, particularly those who use secondary data sources, can complete their analyses and make recommendations for policy change without having ever consulted with the people who would be impacted by the work or those policies. And even researchers

who do engage with community members for feedback and insight are under no obligation to alter their analyses or recommendations, especially if they are not in violation of any laws, regulations, or ethical concerns as laid out by their hosting institution or institutional review board (IRB). Advocates and community partners, like Chicago Beyond, have argued that research-driven policy should not only be driven by expressed community needs with community input throughout the research process, but also result in policy changes that can be experienced within a person's lifetime, so that the people who seek the policy change also get to benefit from it (Chicago Beyond 2018). Mention of these risks is common in literature on data matching and data integration (Gaddy and Scott 2020). Whether researchers can empathetically consider community needs and aspirations is not typically included among the common ethical risks and goes beyond the concerns typically covered by accountability systems like IRBs.

Community-engaged methods represent a powerful potential solution to addressing concerns around empathy, but they are more easily applied for researchers working primarily at the local level where key partners may be more readily identifiable and potential use cases more specified. That said, imputation and other data-analytic efforts to fill data gaps may focus on national datasets without focusing on engaging any specific racially or geographically defined community.

We would argue that the need to orient the imputation toward a clearly defined and requested need is no less important, and that representatives from organizations focused on racial equity (e.g., Government Alliance on Race and Equity) and policy-specific interest groups (e.g., Change Machine for credit and financial well-being) can either proxy for communities or facilitate connections to more appropriate representatives. How this could work in practice would depend on whether the findings from the imputation are meant to be broadly used or whether the data are being generated for a specific place .

For a broad use case, the process could start by involving key interest groups—interest groups with a focus on race and equity and interest groups focused more specifically on the policy area—early in the analysis process, as researchers and analysts consider imputing race and ethnicity onto a dataset. These interest groups could provide high-level feedback and direction and could convey concerns about issues that could arise from the imputation process. After the initial feedback and direction, the research team could perform the imputation, working to resolve any concerns raised by the interest groups, institutional accountability systems (e.g., IRBs), and other researchers and stakeholders involved in peer review.

For a specific place or party (such as a state, a city, or a network of hospitals), the research team could work with the client or primary stakeholder to be connected with community organizations or

interest groups that have experience working in the affected communities and understand the policy needs well. Similarly to the broader application, researchers could solicit feedback from these groups around the needs and how imputation could be applied, making sure to acknowledge and incorporate the interest groups' contributions and to compensate them for contributions if requested. Based on the initial feedback, the research team could perform the imputation with the necessary quality checks and ethical review, and once the dataset is complete and ready for use, reengage community partners to consider thoughtful and useful applications. These approaches would have researchers and analysts not only consider ethical risks but include the perspectives of communities of color and other people affected by the application of data with imputed race and ethnicity. Taking empathy into account in imputation means representing the personhood of the people affected by the imputation beyond their representation as a part of distribution, and considering and thoughtfully addressing their concerns, needs, worries, and aspirations. In addition to ethical risks, empathy means considering the agency of the people affected and how they could use the data to represent their own interests, which may differ from those of the researchers and data creators.

## Who Should Be Involved in Developing and Applying Ethically Imputed Disaggregated Data?

When evaluating which people, organizations, and stakeholders to engage with when imputing data disaggregated by race and ethnicity, we recommend that data analysts and researchers consider where the research process would most benefit from stakeholder review and input and engage with three groups—researchers, community partners and representatives, and end users—in the imputation process. Each of these groups comprises the following actors and organization types:

- **Researchers, practitioner peers, and institutional accountability partners** include people and organizations working within and outside of required accountability mechanisms, such as university faculty, graduate students, and researchers and analysts at research/policy organizations and think tanks, private consulting firms, and corporations.
- **Community partners and representatives** include people and organizations working in the context of communities at the local level and/or in partnership with state- or national-level coalitions and networks; community-based researchers who advocate on behalf of historically marginalized communities; community-based organizations working at the intersections of service delivery, advocacy, and/or community engagement; community leaders and advocates with deep connections to Asian, Black, Indigenous, and/or Latine communities and have

expertise or an interest in how data are collected from residents and community members; and community members from whom data was collected.

- **Additional end users** comprise changemakers across industries and sectors, including people and organizations using imputed data to answer research questions and explore new areas of inquiry. These include federal, state, and local government entities and agencies, university faculty, graduate students, data scientists and analysts at research/policy organizations and think tanks, community partners, community-based organizations, service providers, private consulting firms, and corporations.

There is some overlap across these categories. For example, university faculty members can be responsible for imputing race and ethnicity data and also end users of the imputed data. Similarly, community partners are well positioned to provide input about the communities they represent or belong to, lift up potential risks or harms of the data used in an imputation process to particular communities, and use the data and/or translate the findings and implications of the analyses conducted with imputed data back to the community.

## How to Engage Research and Practitioner Peers

There is increasing interest in expanding access to disaggregated data and identifying best practices for appending race and ethnicity data to existing datasets among scholars, data advocates, funders, policymakers, and research organizations. This provides a tremendous opportunity to build synergy and cross-sector collaboration among diverse data experts, analysts, subject matter experts, and end users. And there is also momentum to address long-standing disparities across key outcome areas, including health.

Recently, researchers, scholars, and activists from organizations like [Data for Black Lives](#) and the [American Medical Association](#) called for the public release of state-level data on COVID-19 infections and deaths by race and disparities in these data to target interventions and resources to the communities hit hardest. Moreover, according to the Centers for Disease Control and Prevention, race and ethnicity data were missing from nearly 50 percent of the vaccine records in the first month of vaccine rollout. To address this challenge in Virginia, the Virginia Department of Health has used the Bayesian Improved Surname Geocoding methodology developed by the RAND Corporation to impute the missing race and ethnicity data from vaccine records (Anson-Dwamena, Pattah, and Crow 2020).

To help create a community of thoughtful, engaged researchers and practitioners, data analysts should consider engaging with researchers, data scientists, and data equity advocates in the following ways:

- Coordinate with and among researchers and data scientists conducting research on a particular topic or policy area of interest. For instance, they could talk to public health researchers or epidemiologists about projects related to COVID-19 or to labor economists about employment-related questions.
- Have researchers with expertise in imputation, survey analysis, and privacy-preserving analytics review the imputation methodology. Acknowledge the tensions in sharing data and code with the need to preserve privacy and confidentiality. Reviewers should make recommendations or propose research alternatives to improve the imputation approach and check the imputation code used for accuracy, either through direct review (if possible) or through conversations with the researchers doing the imputation about their process.
- Engage with researchers familiar with the communities of focus to provide any input on considerations related to the concerns that might be raised from within those communities.

Identifying and engaging with a rich network of researchers and data scientists can help build a body of knowledge, technical expertise, and documentation on imputation for a particular area of research. Engage other researchers and data scientists familiar with the method or the topic/policy area of interest. Engage them by having them review methodology and check code, and ask them to review for blind spots, potential improvements, or research alternatives. Moreover, in addition to engaging with researchers and practitioners, analysts and data creators should engage with institutional accountability partners and other support staff to weigh potential legal and ethical ramifications.

## How to Engage Accountability Partners

Successfully addressing ethical concerns requires researchers and analysts doing imputation to solicit and incorporate external feedback, ideally from institutional or community partners whose mission is to increase the benefits and mitigate the harms that can come from research. Institutional review boards serve in this capacity in many places with formal research communities, such as universities, nonacademic research centers, hospitals, and government agencies. Engaging with an IRB or other research-focused ethics committees provides necessary protection for human research participants through independent reviews of the ethical acceptability of human research proposals. An IRB can approve, request modifications of, or reject research proposals involving human subjects. Although

IRBs are often thought of as most responsible for mitigating risks, especially for vulnerable populations, the 1979 [Belmont Report](#) has helped inform the work of IRBs for decades (Gaddy and Scott 2020). The report highlights that the ethical principles of human subjects research should include respect for persons, beneficence (which suggests that the research should not just mitigate harm, but support well-being), and justice. Particularly in its reference to justice, the Belmont report highlights that the ethical application of human subjects research should consider “fairness of distribution” and “what is deserved.” It also acknowledges that applying justice through research requires an understanding of historical harm, including racialized harm (e.g. Tuskegee syphilis experiment) (Belmont Report 1979).

For researchers using an imputation process to append race and ethnicity data, the IRB is well positioned to review draft proposals, encourage researchers to gather as much information as possible on the group of people they are interested in learning more about and recommend strategies to ensure data privacy and protect sensitive information, which is especially important for already vulnerable people. The IRB can also advise on any legal implications associated with appending data in ways that might make it easier to reidentify people or entries in the dataset. Although additional reviews or checkpoints will require additional resources and labor, these opportunities for feedback will also help ensure that researchers take the necessary steps to familiarize themselves with the groups on which the research is focused and increase community autonomy and buy-in throughout the research process. Researchers have recently called for IRBs to adopt racial-equity considerations by more explicitly considering and integrating racial and ethnic communities’ perspectives and ideas about how work might impact them (Andrews, Parekh, and Pecktoo 2019). Specifically, collecting and considering background and contextual information on the groups of people involved in the study, ensuring that information about the research is written and translated in a way that is accessible to community members and study participants, and using data collectors who reflect the impacted communities are just a few of the recommendations.

Many researchers and analysts work in places where an IRB does not exist and where there may be no formal accountability mechanism to review research for ethical concerns beyond legal liability. In these cases, analysts could work other key partners and stakeholders to set up their own data governance systems to help provide the necessary accountability systems. In the related field of data integration, researchers at the University of Pennsylvania’s Actionable Intelligence for Social Policy center have argued for and provided frameworks for data governance structures that could apply well to imputation work (Gibbs et al. 2017; Hawn Nelson, Jenkins, Zanti, Katz, and Burnett et al. 2020). Because governance plans are formal arrangements between partners, they can be established by any group of researchers to provide accountability. A data governance plan provides direction and goals for a project and provides guidance for who should be involved in the data integration process, including

who owns and has access to the data, who gets to analyze and manipulate the data, and who is served by the data. A governance plan also establishes who holds authority over direction and who needs to be kept informed throughout the process.

Ethical oversight and community improvement—including work done to dismantle the effects of structural racism in communities of color—are also key goals of good data governance systems (Hawn Nelson, Jenkins, Zanti, Katz, and Burnett et al. 2020). Actionable Intelligence for Social Policy suggests that legal liability is the baseline concern for data integration and that ethical risks and whether a project is a “good idea” are critical factors for deciding whether and how to move forward with data integration—in our case, imputation—to inform policy (Gibbs et al. 2017; Hawn Nelson, Jenkins, Zanti, Katz, and Burnett et al. 2020).

## How to Engage Community Partners and Representatives

Power dynamics that often benefit researchers and disempower community members and other stakeholders whose data are collected should be acknowledged and accounted for as a part of the imputation process. In addition to providing accountability, community-engaged methods can be used to bring affected community members into the research process to decide which questions to focus on, how to interpret results, and how to apply findings (Falkenburger et al. 2020). To the extent possible, researchers working with large datasets should prioritize including the perspectives of affected communities in the imputation process. Increasingly, researchers are making the case that deep community engagement—being mindful not to engage so much as to be an undue burden—can and should be integrated into quantitative research, noting the benefits to both the research process and to the communities of focus (Gaddy and Scott 2020; Gibbs et al. 2017; Hendey et al. 2020). Community engagement should be made an essential component of planning for imputing race and ethnicity data, and researchers should invite community partners to work with them, provide input, and outline any concerns about the imputation approach. What follows are specific recommendations about how to engage community partners and representatives when planning on imputing data:

- **At the beginning of the imputation process, analysts should engage with community members when deciding whether imputation is the right approach to fill the data gap.**
  - » This is a critical opportunity for learning and transparency, where data analysts can articulate the motivation for imputing race and ethnicity data. Analysts should seek answers to important questions around who is “the community” and who are “community representatives,” especially when using national-level data, third-party data, or large-scale

administrative data. In these cases, there will likely be multiple or loosely defined communities and there may be a lack of consensus or clarity around the applications of disaggregated data. Once identified, analysts and researchers should be prepared to share with community partners and members the ways that the lack of disaggregated data mask long-standing inequities and make it difficult for researchers to know how and to what extent racism influences outcomes for different racial and ethnic groups. Researchers should also outline the potential for data analysis to encode racial biases in the input data sources.

- **During the imputation process, analysts should check in with community members to weigh risks and discuss possible mitigation strategies**
  - » During the research process, data scientists and researchers should include checkpoints where they assess methodological decisions that could increase the risk of producing inaccurate results that could harm people or communities. Researchers should engage community members to assess the extent of this risk and determine whether the risk is outweighed by the potential community benefit of producing disaggregated data. Presenting methodological decisions transparently and clearly is key to facilitating meaningful community engagement at these checkpoints.
  
- **As imputation concludes, analysts should consult with community partners when applying the imputed data and deciding whether to release results from the analysis.**
  - » One of the goals of disaggregated data is to show where aggregate data are masking disparities. Once these disparities are identified among specific subgroups, researchers, policymakers, and funders can decide to support and implement appropriate programs, endorse evidence-based interventions, and direct targeted resources where they are most needed to address the disparities. Before these decisions are made, community partners and members should be made aware of the results of the imputation process and should be partners in interpreting and contextualizing the findings. Researchers should also engage community members to ensure that the language and terminology associated with the imputation and the corresponding results are accessible to and can be understood by community members and other stakeholders who speak languages other than English and have varying levels of knowledge and understanding of statistics and data science. Most importantly, as researchers finalize the new dataset that includes race and ethnicity data, they should determine the level of risk associated with making the data public. It may be necessary to take additional steps to protect vulnerable communities and groups—such as

by releasing data at acceptably coarse aggregates, generating synthetic data that preserve privacy, or only releasing data to certain trusted actors—or in some instances forgo publishing the disaggregated data altogether.

## Who Are End Users and How Should They Be Engaged?

To identify the best-case application of and uses for imputed data and to ensure those data are accessible to the intended audience, data analysts should engage with end users. These include community representatives and other stakeholders in potentially affected groups, including the original data developers and owners, individual community members, community partners, researchers, data analysts, policymakers, and other decisionmakers, at various points in the research process. Moreover, analysts should engage with a broad constituency of practitioners, policymakers, advocates, funders, and other analysts who would use the imputed data, and people who represent affected communities (be they geographic communities, like neighborhoods, or racial/ethnic communities, like Pacific Islanders) who are in a position to engage the researchers to discuss findings and the approach. Analysts should also consult with end users who may have insights about what details the imputation should highlight or make more salient. If possible, it is important to engage them before releasing results to ensure concerns about potential impact—positive and negative—are heard and addressed. Community partners and end users should also be engaged after results are released to ensure ongoing risks are minimized and data are having their intended impacts. Engagement should occur in the following stages:

- **At the beginning of the imputation process**, peer researchers, community representatives, policy area experts, accountability partners, and potentially data owners (if using proprietary data) and funders should be **invited** to provide feedback on the project's approach and goals to check that the aims of the project are aligned with the end users' needs and expectations.
- **During the imputation process**, peer researchers, accountability partners, data owners (if using proprietary data), and potentially policy area experts should be **consulted** to provide feedback on the appropriateness of the data, the rigor of the analytic approach, and whether the resolution of concerns around ethics and empathy support the end users' expressed need for the data.
- **After the imputation is complete**, all potential end users should be **informed** of the results and potential analytic value of the final data through broad dissemination strategies (like public reports, guides, and data tools) and through specific and targeted outreach to the key

stakeholders who have been involved throughout the process and for whom the data provide the most potentially impactful use case. Community partners and representatives should be partnered with to help translate and communicate findings in accessible ways and provide access to the results in ways that these communities can use to advocate for policy changes that best represent their interests.

## Deciding When Imputation Is the Right Approach for Disaggregation

Whether researchers and other actors should pursue research involving imputation at all given potential risks and other alternatives for obtaining the relevant information is always a relevant question. As with any research involving risks, actors engaging in this kind of research should not only carefully weigh the pros and cons, but seek out accountability structures to help them determine whether the benefits merit conducting the research. Institutional review boards play this role in many universities, hospitals, and research organizations, and community review boards are increasingly playing an active role in guiding and advising research projects. Review boards are not the only possible accountability structures, particularly given they focus more on preventing individual harm than group harm or opportunities for individual and group benefit. Critically important here are the perspectives and feedback of affected communities, which can identify and weigh the potential risks and benefits. Community-engaged methods can provide best practices for helping researchers involve communities in this capacity.

### Should Researchers Impute Race and Ethnicity?

Below, we provide considerations that may compel researchers to pursue or decide against imputation of race and ethnicity.

#### THE CHECKLIST FOR “YES”

Imputation can be a valuable tool for filling data gaps and for answering critical research and policy questions on racial disparities. Directly collecting missing data can be difficult because of time and financial constraints, or prohibitive because of regulations preventing the collection of data explicitly on race. The potential opportunity costs of using resources on imputation rather than strengthening data collection systems are low. The biggest opportunity for benefit is to provide findings back to communities looking for better data to help answer questions about key disparities. In addition,

clarifying the end use case is important. Data application is not neutral, and ethical considerations emerge depending on when and how researchers and other actors use the data. The parties deciding whether to impute information on race and ethnicity should weigh the considerations that follow.

**Ask whether imputation will sufficiently fill the data gaps.** Researchers and policymakers need to ask whether imputation (or another advanced method) is this the right tool for the question at hand, and which methodological choices they should select considering that use case. Researchers and data scientists should not impute race and ethnicity, or otherwise fill data gaps, in a vacuum divorced from an actual use case or research question. They should also consider potential downstream uses by others. Without some anticipation of how others not involved in the initial development and analyses might use the data, researchers will not have all the information they need to make sound choices grounded in the full slate of ethical implications.

**Be clear about which racial and ethnic categories can be imputed accurately.** Remain honest with researchers on the team, accountability partners, and community members and representatives about methodological limitations, especially regarding the groups of people for which imputation is a valid method and those for which it is not.

**Engage and collaborate with community members or representatives.** Drawing on community-engaged methods, develop a process to engage either with community members or interest-group representatives as partners in choosing questions of interest and interpreting findings.

**Clearly explain the method to people with varying levels of technical expertise.** As methods become more complex, they become more difficult to communicate transparently. This makes it more difficult to obtain necessary community feedback on those methods and their outputs. Researchers should explain processes and methods in clear terms that communities can easily understand.

**Implement a data-governance plan and data protections.** Researchers and policymakers should put the necessary protections and accountability mechanisms on data usage in place to prevent abuse. They should recognize, however, that making data accessible to those who want to use it for progress can also expose the information to actors who may want to use the information in bad faith to target a community whose interests and benefit they do not share. Even with excellent protections, if and as access expands, the possibility exists that people or organizations acting in bad faith or with predatory intent could use the data in ways that cause more harm than benefit.

**Consider whether the benefits outweigh the risks.** Community members should feel that the benefits of the identified use case for data imputed for communities of color outweigh the risks. Where policy

decisions based on data without detail on race risks developing poorly targeted solutions and exacerbating racial disparities, part of the risk calculation will involve the risk of not collecting any disaggregated data. As long as the imputation is rigorously executed, engages community thoughtfully and collaboratively throughout the process, submits to accountability measures, and protects privacy, even with other needed improvements, imputed information on race and ethnicity can be used to the benefit of people of color in developing more equitable policy.

#### THE CHECKLIST FOR “NO”

Although imputation can help address gaps in data and research on race and ethnicity, the data we have often limit its value as a tool. At times, there are statistical criteria related to reliability and sample sizes that aren't easy or possible to overcome. It's important to see where datasets and algorithms fall short and are themselves biased. Using flawed methodologies and algorithms simply reinforces existing patterns and biases in the data. The "scientific" or "empirical" nature of a statistical analysis encourages people to accept findings, even when those findings should be questioned. **Researchers should not decide to impute data on race or ethnicity if they make the determinations that follow.**

***The risks outweigh the benefits.*** If an imputation cannot be rigorously executed, submit to accountability measures, successfully incorporate feedback and insight from community members, or preserve people's privacy and confidentiality, then that lack of data governance offers insufficient protection. Not only are there increased risks that the data may be unrepresentative or inaccurate, but actors may more readily use the data or findings from an analysis to target, discriminate (intentionally or unintentionally), or harm communities of color and alienate them further from data and data-driven policymaking.

***There are more optimal alternatives for filling the data gaps and needs.*** Analysts and data creators should carefully consider potential alternatives. They should engage end users and community-level stakeholders, taking their feedback and expressed needs into account. Recognizing that few, if any, analytics processes are perfect, all stakeholders should reach a consensus about need and usefulness before using any imputed data for policy or decisionmaking or releasing it for wider use.

***Communities preferences for data governance cannot be met.*** Different communities may have different standards for data governance and data sharing—which may extend to different subgroups within a community (e.g., [across tribal communities](#))—and what steps should be taken to protect community and personal information. In some cases, the benefits might not outweigh the risks.

***Communities' historical trauma with previous research and data collection cannot be sufficiently reckoned with.*** It is critical to understand communities' relationships with research and data, including the harms and exploitation that different industries and fields have inflicted on communities of color while seeking data or performing research. Researchers should understand historical harms that data and research have had on communities and should avoid imputation if they lack that understanding and do not have a plan to take that history into account in the new analysis. Affected communities often lack the power to control the data they provide and the resulting research narrative. Given the history of technological advances, "big data" approaches may lead to further oppression (Milner and Traub 2021).

***Particular communities of interest for the research cannot be sufficiently represented in the imputed data.*** For communities with smaller populations or less visibility outside of large national data sources (like the census), it is not advisable to try and impute data. This is particularly true for Asian American, Pacific Islander, and Indigenous communities. It may be possible to collect new self-reported race and ethnicity data from communities missing data. Though the process may take longer, the data may be more accurate than in an imputation involving more errors or uncertainty. If direct data collection at scale is not possible, pulling in qualitative perspectives through interviews and focus groups or including findings from other closely related research could be helpful.

From our research and landscape scan, interviews, and case study, our view is that imputation is a powerful tool for increasing visibility into and for communities of color. In cases where imputation can be done with methodological rigor and representative accuracy, the benefits to understanding and addressing disparities are generally robust. That said, engaging with communities to respond to their requests for policy change requires considering alternative forms of data collection and analysis when those communities raise serious concerns about imputation that cannot be resolved.

## How to Take an Ethical Approach to Imputation

Assessment of the risks and benefits of imputation does not end with the decision to proceed with the imputation process; every decision throughout that process has the potential to inject bias and inaccuracy into the imputation results that could risk harm to communities of color. Accordingly, analysts should establish checkpoints before, during, and after the imputation process to ensure that each decision is made with ethics and empathy at the forefront. At each checkpoint, the researcher should mitigate any risk identified as best as possible, clearly communicate limitations on ethical data use imposed by bias that cannot be fully mitigated, and engage with community members to determine

whether the unmitigated risk outweighs the potential benefits from imputation and warrants terminating the process. The checkpoints at each stage are as follows.

### **Checkpoint 1: Before Imputation, Audit Input Data for Bias**

Data ethics advocates have produced considerable evidence of the potential for data analysis to encode racial biases in input data sources. This risk can be harder to detect in complex methodologies, like machine learning or imputation, because of the layers of transformation between the input data and the analytical output—in our case, the imputed race/ethnicity variable. Accordingly, when gathering datasets before imputation, researchers should examine each dataset for potential bias by asking the following questions:

- Does the dataset accurately represent the underlying population that it aims to measure? How might structural racism drive unrepresentativeness?
- Do all of the datasets being used in the imputation and benchmarking represent the same population?
- Are data missing? If so, does that missingness disproportionately affect certain racial or ethnic groups? Is it correlated with other variables of interest?

### **Checkpoint 2: During Imputation, Examine Where Bias Could Be Introduced at Each Step**

Input data is not the only potential source of bias in data analysis; every methodological decision has the potential to bias the outcome. This is a particular concern when imputing race and ethnicity because a decision may not change overall accuracy, even though it may differentially reduce the accuracy of imputations for smaller racial and ethnic groups. At each step of the methodology, researchers should examine whether their methodological choices could be differentially effective by racial and ethnic group. Researchers should clearly communicate limitations resulting from any of these choices as well as the uncertainty of the resulting estimates.

### **Checkpoint 3: After Imputation, Assess Whether Imputed Data on Race and Ethnicity Are Accurate Enough to Be Used Ethically for the Analytical Purpose**

Even with robust efforts to mitigate bias in the input data and imputation methodology, some amount of inaccuracy in the imputation is inevitable, just as any prediction or estimation from data is unlikely to be perfect. The question when imputing race and ethnicity onto datasets is whether the resulting data are accurate enough—in terms of overall accuracy and differential accuracy across racial and ethnic groups—to be responsibly used for a given analytic purpose. The answer to this question will vary based on the intended use. Researchers should assess the accuracy of disaggregated by racial and ethnic group by comparing estimates calculated from the imputed data against trusted disaggregated statistics. They should also examine whether the data are precise enough to ethically support their intended analytical use case by examining the potential for the imputed data to yield inaccurate conclusions relative to the “next best” data source. (For information on how we implemented these checkpoints when imputing race and ethnicity onto credit bureau data, please see our companion case study, Stern and Narayanan [2021].)

### **Standards for Ethical Imputation**

In table 1, we present the following standards for producers and users of imputed data on race and ethnicity in the policy context: relevance, interpretability, coherence, accuracy, privacy, and institutional environment. We developed these using lessons learned through our landscape scan, interviews with experts, and through our own imputation case study. The standards are intended to help data scientists, analysts, researchers, and policymakers evaluate the appropriateness and accuracy of the imputation process.

TABLE 1

**Standards for Imputing Data Disaggregated by Race and Ethnicity: Relevance, Interpretability, Coherence, Accuracy, Privacy, and Institutional Environment**

Standard	Description of standard for evidence	Recommended actions for analysts
Relevance	<p>The degree to which the disaggregated data meet the needs of the user in regard to data level, timeliness, etc.</p> <p>How suitable the data are for the research purposes.</p>	<p>Be clear about use cases. Who needs these imputed data and why?</p> <p>Using accessible language, be honest and transparent about the benefits and weaknesses of your imputation or data integration effort. It's possible that the imputation solved a critical missing data problem, or that it's a short-term patch in a situation where additional on-the-ground data collection is what is needed most.</p>
Interpretability	<p>The clarity of information to ensure that the data are used appropriately.</p>	<p>Be explicit about how the data are represented, what model was used, and why. Consider how the data might be limited or biased.</p> <p>Ensure that the data do not reinforce or enhance any biases. Be intentional about not just reporting what the results are but also describing why they are what they are, which can help researchers better understand the underlying or contributing racialized problems or challenges.</p> <p>Accurately calculate and clearly communicate uncertainty of any estimates derived from imputed data.</p>
Coherence	<p>Making sure the datasets used in the imputation represent the same population as the dataset being imputed.</p>	<p>Describe in detail the reasons for selecting the data source being used to append race and ethnicity data. Be clear about how the data source is similar to or different from the dataset of interest and why it was selected to draw on to impute missing race and ethnicity values.</p> <p>Also provide detail about how the data were imputed, clarifying how the imputation process using these data fit the intended purpose of the imputation.</p>
Accuracy	<p>The closeness of the imputed data to their unknown true values.</p>	<p>Use data analysis best practices, including making every effort to be rigorous and careful and documenting each step and decision point used in the imputation process. Make the methodological approach available for other analysts and researchers to review.</p> <p>At each step of the imputation process, examine the potential for methodological choices to produce inaccurate results—including differential accuracy by racial and ethnic group.</p> <p>Mitigate risks of inaccuracy where possible, communicate any unmitigated risk, and engage community in determining when unmitigated risk warrants terminating the imputation process.</p> <p>Benchmark estimates calculated from the imputed data against trusted statistics disaggregated by race and ethnicity. Estimates can also be benchmarked against the same dataset used to create the imputation, but at different levels of aggregation (e.g., national versus state).</p>

Standard	Description of standard for evidence	Recommended actions for analysts
Privacy	The ability of the imputed data and any output or analyses that could come from the data to preserve the privacy and not risk identifying people.	<p>Establish clear data governance so ownership and management of the data (and any sensitive information that may be part of the imputation, particularly surnames) so that data are preserved and not seen or handled by anyone without proper protocols and protections in place.</p> <p>Establish guardrails so outputs and results won't risk exposing anyone's identity or otherwise violating privacy.</p> <p>Work with community members from smaller and less data-visible groups to agree on approaches to balance releasing information that is helpful for providing information on disparities or policy impact, while minimizing individual-level and community-level privacy risks and other potential harms.</p>
Institutional environment	The credibility of the researchers for producing high-quality and reliable data, and quality standards in the organization for assuring adherence to ethical practice.	<p>Build in accountability throughout the process—both on the technical side and the potential (policy/ethical) impact side. As part of this, develop a structure and process for consensus building and decisionmaking, which can include deciding not to do the imputation.</p> <p>With the aid of key stakeholders, develop a set of policies around release and restriction. Who gets access to the data when it's done and under what conditions?</p> <p>Data-privacy laws limit the processing of sensitive data categories including race, ethnicity, and sexual orientation. Data analysts should be aware of these restrictions and to the extent possible engage with their institutional review board or other ethical-protections entities to ensure the necessary steps are taken to ensure data privacy.</p>

Source: Urban research team.

For each of these standards, data analysts should seek to build and strengthen connections with the people and communities who would be most impacted by the application of the imputed data. These researchers and analysts should create processes as part of the research design that allow community partners and representatives to learn about the research and provide input at each point in the study, particularly at the beginning of the project to identify and address common goals and concerns and toward the end of data development to help disseminate findings and shape applications. Integrating community engagement supports and strengthens each of these standards to ensure the imputed data on race and ethnicity are developed and deployed in ways that are ethical, empathetic, and work toward racial equity.

## Conclusion

To address racial disparities, changemakers must be able to identify them, but in many important policy domains, such as health and financial well-being, data disaggregated by race and ethnicity are not widely and consistently available. The COVID-19 crisis has shown that in addition to obscuring disparities, the absence of these data can lead to inadequately and poorly targeted policies and can even exacerbate disparities.

Collecting self-reported data could make datasets complete and therefore provide accurate information on race and ethnicity, and would avoid many of the ethical risks discussed in this report. But collecting new data comes with tradeoffs, and doing so is not always possible or even allowed and can be expensive and time consuming. Data-integration and data-appending efforts like imputation can be more efficient alternatives to getting racially disaggregated information and addressing important and timely policy questions (Hawn Nelson, Jenkins, Zanti, Katz, and Burnett et al. 2020).

Approaches like data integration and imputation in particular do not typically require the input of the people whose data are being combined or augmented. As such, these approaches are subject to greater ethical and empathy-related risks. These risks—which include inaccurate representation, risk of reidentification, and not fully acknowledging the personhood of people impacted by the imputation—are important considerations, but they can be understood and mitigated. Researchers, data scientists, and other data analysts can work to better understand what constitutes ethical imputation. They can also work with other key parties—especially institutional partners and impacted communities—to consider when imputation is the right approach to addressing disaggregation-related challenges and to apply lessons to effectively address key concerns.

An especially important takeaway from our research is that community-engaged methods should be used whenever possible to incorporate impacted communities' insights into how efforts to use imputation to disaggregate data are conceived and applied. Community engagement can not only mitigate risks, but help researchers better focus their efforts to develop data sources that can be leveraged to address sought-after policy solutions. If researchers and data analysts ask themselves the questions posed in this report and apply the standards to their own work, we believe imputation can be an effective and ethical tool for empathetically addressing critical gaps in data disaggregated by race and ethnicity.

# Notes

- <sup>1</sup> Ibram X. Kendi, “What the Racial Data Show,” *The Atlantic*, April 6, 2020, <https://www.theatlantic.com/ideas/archive/2020/04/coronavirus-exposing-our-racial-divides/609526/>; Leana S. Wen and Nakisa B. Sadeghi, “Addressing Racial Health Disparities in the COVID-19 Pandemic: Immediate and Long-Term Policy Solutions,” Health Affairs blog, July 20, 2020, <https://www.healthaffairs.org/doi/10.1377/hblog20200716.620294/full/>.
- <sup>2</sup> For instance, see President Biden’s [executive order](#) on racial equity in the federal government released earlier this year.
- <sup>3</sup> Though the estimated race and ethnicity is less reliable at the individual level, this approach can produce accurate estimates of racial distributions and racial disparities across larger datasets or samples, especially when used in conjunction with education, income, and other variables.
- <sup>4</sup> “Community Engagement Methods at Urban,” Urban Institute, accessed July 2, 2021, <https://www.urban.org/research/data-methods/community-engagement-methods-urban>.

# References

- AECF (Annie E. Casey Foundation). 2020. *Four Principles to Make Advanced Analytics Work for Children and Families*. Baltimore: AECF.
- Andrews, Kristine, Jenita Parekh, and Shantai Pecktoo. 2019. *How to Embed a Racial and Ethnic Equity Perspective in Research: Practical Guidance for the Research Process*. Bethesda, MD: Child Trends.
- Anson-Dwamena, Rexford, Priya Pattah, and Justin Crow. 2020. "Imputing Missing Race and Ethnicity Data in COVID-19 Cases." Richmond: Virginia Department of Health.
- CFPB (Consumer Finance Protection Bureau). 2014. *Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity: A Methodology and Assessment*. Washington, DC: CFPB.
- . 2015. *Data Point: Credit Invisibles*. Washington, DC: CFPB.
- Chicago Beyond. 2018. *Why Am I Always Being Researched?* Chicago: Chicago Beyond.
- Donders, A. Rogier T., Geert J. M. G. van der Heijden, Theo Stijnen, and Karel Moons. 2006. "Review: A Gentle Introduction to Imputation of Missing Values." *Journal of Clinical Epidemiology* 59: 1087–091.
- Elliott, Marc N., Peter A. Morrison, Allen Fremont, Daniel F. McCaffrey, Philip Pantoja, and Nicole Lurie. 2009. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." *Health Services Outcomes Research Methods* 9: 69–83.
- Falkenburger, Elsa, Nora Hakizimana, Ebonie Megibow, and Eona Harrison. 2021. *Promoting Adolescent Sexual Health and Safety*. Washington, DC: Urban Institute.
- Fremont, Allen, Joel S. Weissman, Emily Hoch, Marc N. Elliott. 2016. "When Race/Ethnicity Data are Lacking: Using Advanced Indirect Estimation Methods to Measure Disparities." *RAND Health Quarterly* 6 (1): 16.
- Gaddy, Marcus, and Kassie Scott. 2020. "Principles for Advancing Equitable Data Practice." Washington, DC: Urban Institute.
- Gibbs, Linda, Amy Hawn Nelson, Erin Dalton, Joel Cantor, Stephanie Shipp, and Della Jenkins. 2017. *IDS Governance: Setting Up for Ethical and Effective Use*. Philadelphia: University of Pennsylvania, Actionable Intelligence for Social Policy.
- Hawn Nelson, Amy, Della Jenkins, Sharon Zanti, Matthew Katz, Emily Berkowitz, T. C. Burnett, and Dennis Culhane. 2020. *A Toolkit for Centering Racial Equity throughout Data Integration*. Philadelphia: University of Pennsylvania, Actionable Intelligence for Social Policy.
- Hawn Nelson, Amy, Della Jenkins, Sharon Zanti, Matthew Katz, T. C. Burnett, Dennis Culhane, K. Barghaus, K., et al. 2020. *Introduction to Data Sharing and Integration*. Philadelphia: University of Pennsylvania, Actionable Intelligence for Social Policy.
- Hendey, Leah, Kathryn L. S. Pettit, Jake Cowan, and Marcus Gaddy. 2020. "Investing in Data Capacity for Community Change." Washington, DC: Urban Institute, National Neighborhood Indicators Partnership.
- Lee, Nicol Turner, Paul Resnick, and Genie Barton. 2019. *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*. Washington, DC: Brookings Institution.
- Markle Foundation. 2021. *Unlocking Responsible Access to Data to Increase Equity and Economic Mobility*. New York: Markle Foundation.
- Milner, Yeshimabeit, and Amy Traub. 2021. "Data Capitalism and Algorithmic Racism." New York: Demos.
- Randall, Megan, Alena Stern, and Yipeng Su. 2021. "Five Ethical Risks to Consider before Filling Missing Race and Ethnicity Data." Urban Institute: Washington, DC.

Rubin, Donald B. 1996. "Multiple Imputation after 18+ Years." *Journal of the American Statistical Association* 91 (434): 473–89.

Stern, Alena, and Ajjit Narayanan. 2021. *Ethics and Empathy in Using Imputation to Disaggregate Data for Racial Equity: A Case Study imputing Credit Bureau Data*. Washington, DC: Urban Institute.

# About the Authors

**K. Steven Brown** is a senior research associate in the Center on Labor, Human Services, and Population and associate director of the Racial Equity Analytics Lab at the Urban Institute. He regularly leads and collaborates on research and policy-advising projects that explore sources of and solutions to racial inequalities in economic opportunity. His primary research focuses on employment, particularly career pathways, earnings, and workplace protections; financial well-being; and economic mobility. Previous work includes federal evaluation and academic research on segregation and homeownership, access to affordable housing, and educational equity. Brown received his bachelor's from Princeton University and his master's in sociology from Harvard University, where he was also a fellow in the Multidisciplinary Program for the Study of Inequality and Social Policy at Harvard Kennedy School.

**LesLeigh Ford** is a policy program manager in the Center on Nonprofits and Philanthropy at the Urban Institute, where she examines philanthropic and nonprofit trends and racial disparities in economic opportunity. Her areas of expertise include philanthropic and nonprofit management, covering topics from racial equity and justice to participatory and equitable grantmaking, funder collaboratives, and nonprofit technical assistance. Using in-depth interviews, focus groups, and surveys, Ford evaluates philanthropic and government-funded grant programs. Ford earned her BA in political science and English from the University of Michigan, her MEd in education policy and management from the Harvard Graduate School of Education, and her PhD in sociology from Duke University.

**Shena Ashley**, a vice president at the Urban Institute, has expanded the Center on Nonprofits and Philanthropy's policy research and programmatic initiatives through visionary leadership and strong financial management while maintaining an active portfolio of rigorous academic scholarship.

Before joining Urban, Ashley held academic positions in the Department of Public Administration and International Affairs in the Maxwell School of Citizenship and Public Affairs at Syracuse University and in the Andrew Young School of Policy Studies at Georgia State University, where she taught courses on the nonprofit sector and quantitative research methods and advanced a research agenda focused on grantmaking effectiveness and nonprofit management. At the Annie E. Casey Foundation's Atlanta Civic Site, Ashley was director of research and policy and developed innovative tools for community engagement in performance measurement.

## STATEMENT OF INDEPENDENCE

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.



500 L'Enfant Plaza SW  
Washington, DC 20024

[www.urban.org](http://www.urban.org)