Direct Care Workers Count:
Why Data Matters to Advance Workforce Equity

By Breanna Betts
I. Executive Summary

Caregiving in our nation is in crisis. The long-standing pressures straining the care labor market such as minimum or unlivable wages, few to no benefits, inconsistent schedules and part-time hours forcing caregivers to work multiple jobs, limited career advancement options, and more, were exacerbated and illuminated in the COVID pandemic. These untenable working conditions have led to unprecedented staff turnover rates in nursing homes and ever-increasing demand for home care workers. Direct care workers are essential workers whose physical and emotional labor as home care workers, personal care aides, and certified nursing assistants supports older adults and people with disabilities living with dignity. Yet data collection on the direct care workforce has not kept up with the pace of the care crisis. Accurate, robust, and accessible disaggregated national and regional workforce data (i.e., labor and job quality data broken down by worker demographics) does not exist, leaving policymakers, government officials and employers largely in the dark when trying to address the challenges facing direct care workers and improving care quality for consumers. Direct care is an intensely occupationally segregated field in which most essential long-term services and supports (LTSS) are provided by women, and disproportionately by Black women, women of color, and immigrants, due to the legacy of chattel slavery and intentional policy choices that excluded caregivers from basic labor protections and institutionalized underpay and overwork.

Low quality direct care jobs are largely financed by public dollars. In 2020, over 72 percent of LTSS were paid for by a mix of federal and state government funds. To address the care crisis, federal government agencies—especially since Medicaid is the primary payer of direct care workers’ wages—could lead the way to create a system that works with workers and communities to ensure necessary information on the essential workforce of caregivers is centralized and disaggregated for the first time ever. By envisioning data systems that promote dynamic workforce development program design and assessments, the government can maximize the impact of public funding, ensure public dollars are spent on good quality, family-sustaining jobs, and support states in course-correcting on the racist, sexist legacy of care work in America.

Recommendation Toplines:

• Create a federal standard dataset of direct care workforce data which sets a “floor” for states to build upon, and standardize direct care worker job classifications
• Coordinate the national data collection effort across federal and state agencies
• Appropriate designated federal funds to equitable workforce data collection and system maintenance
• Collect robust demographic data in standardized categories and disaggregate the federal standard dataset to facilitate cross-referencing and equity assessments
• Mandate meaningful worker engagement and participatory, community-led data collection, and utilize a Black Women Best model and “data feminism” framework
• Center workers in their own words by designing and implementing the first national worker-centered survey of direct care workers
II. Why Workforce Data Equity Matters: We Can’t Change What We Don’t Measure

In the direct care industry and across many aspects of society, data has historically been wielded as an instrument to advance white, ruling class interests, but new scholarship teaches us how data can be collected and used in a way that combats oppression rather than entrenches it. The lack of adequate direct care workforce data collection results from the intersecting systems of structural (racism, sexism, ableism, ageism, xenophobia, colonialism, and heteronormativity) that impact all aspects of long-term care. Data as a technology of racism and other systems of oppression functions under the guise that data systems, statistics, and “modern” technologies are neutral and free from bias: seemingly harmless datasets and the methods in which data is collected (or omitted) from—not with—racially “othered” subjects obscure the very real impacts of data collection and the policies that data is used to exert on marginalized communities.¹

The impacts manifest on a spectrum from excluding marginalized groups like Black, Indigenous and people of color (BIPOC), LGBTQ+, immigrants, women, people with disabilities, from public funding and resources, to outright discrimination and obstruction of civil rights and liberties to the point of lower life expectancy and untimely death.¹ Theorists in “data feminism” and anti-racism recognizes that any data that are collected in a society, and the reasons why, are not neutral but rather reflect the values of the culture and entities in power.⁷ Indeed, lawmakers and powerful interest groups have used the “imagined objectivity of data and technology” throughout U.S. history as a tool to maintain white supremacy and reproduce pre-existing conditions of racism across American society.⁸

Alternative frameworks like “data feminism” task data collectors, analysts, and users (such as policymakers and program leaders) with deconstructing how the data in question was or should be collected, and about what populations, by whom and why, to counteract sexism, gender discrimination, racism, ableism, heteronormativity, colonialism, and other intersecting systems of oppression.² For example, recent applications of a data feminism framework include pioneering work by Black feminist scholars such as Dr. Evelyn J. Patterson who applied intersectional structural analysis to existing national data on maternal mortality rates in the U.S. to reveal a much deeper disparity in death rates among women of color and white women than had been determined in other studies utilizing the same dataset.⁹

Data equity frameworks offer a critical tool in advancing direct care worker justice.

¹ Julia Ticona’s research on algorithmic bias in care-work companies’ online platforms provides one example related to how childcare workers are racialized, discriminated against and surveilled under the guise of marketing “trustworthy” workers to app-users. Ticona explains how the seemingly “data driven” profile requirements and platform algorithms seem modern but are intrinsically linked to centuries of suspicion about mostly Black and brown women performing essential care labor. See: https://www.dissentmagazine.org/article/essential-and-untrusted.

² Data feminism is a way of thinking about data, both their uses and their limits, that is informed by direct experience, by a commitment to action, and by intersectional feminist thought. Its practice tunes into how standard methods in data science serve to reinforce existing inequalities, as well as uses data science to challenge and change the distribution of power. This framework for data equity is explored in depth by scholars D'Ignazio and Klein (2020), whose book is based in power analyses pioneered by preeminent Black women scholars Kimberlé Crenshaw on intersectional feminist theory and Patricia Hill Collins on theories of knowledge production, respectively. See: D'Ignazio, C. and Klein, L. (2020). Data Feminism. Retrieved from https://data-feminism.mitpress.mit.edu/pub/frfa9szd
The longstanding limitations of data on the direct care workforce that persist today are no surprise. American labor laws intentionally excluded the majority BIPOC women and immigrant direct care workforce from the benefits and protections afforded whiter, more male workforces, as laid out in “The Racial Equity and Job Quality Crisis in Long-Term Care” and other works. Structural systems of racism and sexism institutionalized data inequity, and those systems must be confronted to envision an equitable national workforce data system. By redesigning data systems with an eye to equity, the U.S. could ensure for the first time that all workers and their experiences get counted to make equitable policy interventions that are based in the reality of workers’ lives and concretely improve working conditions.

III. Problems in The Direct Care Workforce Data System, In Brief

In-depth technical analysis of the fragmented federal workforce data system is outside of the scope of this framework paper, but we aim to summarize the core problems below. The current national direct care workforce data system is underfunded and incomplete, creating data system fragmentation and widespread negative impacts on direct care workforce planning and policymaking that have been well documented over recent decades, including the following:

- **There is no federal hub for consistent, reliable information about all aspects of the direct care workforce:** Instead, three main federal datasets are available: The Census Bureau’s American Community Survey, the Bureau of Labor Statistics (BLS)’ Occupational Employment Statistics, and BLS’ National Compensation Survey. These are typically cross-referenced by researchers and analysts to attempt to build an aggregate, or summary, picture of the national workforce based on relatively small samples. The only publicly accessible federal data on basic direct care working conditions and compensation, such as wages and benefits, is limited to averages across the sectors of samples of workers employed in covered industries, rather than individual workers’ actual compensation. Covered industries include home health agencies and nursing homes but leaves the hundreds of thousands of direct care workers who are independent home care providers out. Since federal data is available for the sector, not individual workers, disparities in wages and benefits across different race, gender and other identifying demographics of workers are obscured within the same sector.

- **Many workers are not counted or undercounted:** Direct care workers who work as independent providers in private homes whose clients take on all employer-related roles including pay and tax reporting, as well as those employed in the “gray market” (i.e., are paid cash or “off the record” in private households), are often not counted at all. The very job titles that are considered direct care occupations are not standardized across the three federal datasets, and do not encompass the full picture of caregivers employed to provide LTSS. For example, the U.S. Government Accountability Office (GAO) states in its 2016 review that the federally available data’s inclusion of independent providers employed directly by consumers to provide personal care services through state-administered Medicaid programs is “unclear” and varies by state.

---

3 This 2016 report by the U.S. Government Accountability Office provides a detailed overview of the problems.
• **The lack of a federal mandate and designated funding for comprehensive data collection and maintenance on the workforce leaves states to piece together what they can with limited resources:** The lack of a comprehensive federal dataset leaves a gap for states to fill (or not) with their own workforce data collection systems. As PHI found in 2009 in a report documenting workforce data gaps that remain today:

> “[T]he vast majority of states do not have access to basic quantitative information that describes the state of the workforce in place to deliver services and supports to people with various kinds of disabilities. Nor do they have the data needed to support a comprehensive workforce strategy across service settings. Instead, various program, agency, and departmental silos’ collect their data independently, and even then, workforce data is rarely a dedicated focus, contributing to an ad hoc or disjointed approach.”

• **Gaps in demographic data and inconsistencies across how datasets account for workers’ identities makes understanding how to address the labor market issues through targeted, equity-based policymaking challenging:** Broadly, there is no one centralized national data system that enables equity-focused analysis. The direct care workforce is one of the most diverse workforces in the country. Direct care workers’ experiences vary within the same job classifications and settings of direct care, in similar geographic areas regionally and across states, based on their intersectional identities—race, gender, sexual orientation, immigration status and country of origin, age, disability, and more—impacts the different wraparound supports that workers need to stay in emotionally and physically grueling frontline care jobs, such as reliable transportation, child care, mechanisms to prevent and respond to harassment and discrimination, and culturally responsive management. Yet the lack of demographically disaggregated data hamstrings policy and program efforts to address the challenges that workers of different identities may face, and particularly obscures inequities within the workforce, such as access to opportunities for higher wages, better benefits, career mobility, advancement, and paid training in the field.

**IV. Key Recommendations**

While a wholesale redesign of the fragmented LTSS workforce system is beyond the scope of recommendations in this framework brief, we recommend that the federal government leverage its position as the majority payer LTSS to implement a “floor” of direct care workforce data collection standards by tying public investments to equitable, worker-centered data system requirements and incentives in the states in the following ways:

• **Create a federal standard dataset of direct care workforce data which sets a “floor” for states to build upon:** A reimagined direct care workforce system would rely on the national dataset as a floor that could be tailored on a state-by-state basis to the needs of specific communities of workers and their families by seeking robust input from impacted stakeholders, including local workers, consumers, caregiver unions, employers and providers. In establishing a federal standard dataset of workforce information about providers, payers, residents and consumers, worker characteristics, and job quality information, the federal government can at last create a repository of information that policymakers, advocates,
employers—and importantly, workers and communities—can access to make informed, successful policy and program decisions. Creating a federal standard direct care workforce dataset is aligned with the Biden Administration's Executive Order 13985 (January 20, 2021) and so-ordered Recommendations from the Equitable Data Working Group.

A foundational step in creating the federal standard dataset will be establishing a standard definition of direct care and the direct care worker job classification, to align the lack of standardization in job titles across the three existing federal datasets as described above in the discussion of problems in the data system. This would create a single job classification for data management purposes that captures information about home health aides, personal care aides, and nursing assistants providing LTSS across differing state programs. The new direct care worker job classification would be used as the least specific category in the federal standard dataset for understanding data across the workforce, and more specific job titles such as CNA could still be included for analysis if desired.

- **Coordinate the national data collection effort:** Addressing decades and decades of neglecting to document direct care workers’ experiences will require a coordinated national effort. The workforce data housed in the federal standard dataset recommended in this framework will need to be collected and tracked by the states due to our nation’s fragmented LTSS system and myriad of different programs across the states. To standardize the data system, all federal agencies that touch direct care workforce issues, including Health and Human Services Agency (HHS), Center for Medicare and Medicaid Services (CMS), the U.S. Department of Labor (U.S. DOL), and the Administration for Community Living (ACL), should create regulatory “strings” to tie public investment to and federal matching and reimbursement dollars to states’ implementation of equitable data systems. Tying public investment would include both incentivizing states and providers to participate in data systems reforms through increased federal Medicaid matching rates, for example, and penalizing those who do not comply. These regulatory strings should include requiring standardized collection of basic workforce data, race, gender and other worker demographic data, and data on indicators of job quality, such as job stability, wages and benefits as compared to cost of living, and more.

- **Appropriate designated federal funds to equitable data collection and system maintenance:** Mandating data collection in standard categories will require collecting data from many actors in the workforce system, including governments acting as employers of record in LTSS systems, fiduciary management systems handling payroll for independent providers, managed LTSS health plans, and employers/service providers who are licensed by federal and/or state government agencies or are reimbursed with any public funds or through public insurance programs. To ensure that comparable data is collected across the workforce and that this significant administrative and labor-intensive data collection process is thorough, designated federal funds to on-ramp new data collection and support ongoing data management should be appropriated through HHS, CMS, U.S. DOL, and ACL in a coordinated manner. States and providers face over-stretched capacity and budgets, so it is crucial that additional federal funds that are carved out for data collection and maintenance to prevent shrinking the funds available to pay direct care workers’ wages and benefits in an already under-resourced LTSS system—and large-scale investment in the direct care workforce and Medicaid LTSS on the scale of the original “Build Back Better” plan will be key to overarching system reform.

Forthcoming tools from The Center for Equity will identify indicators for a federal “floor” dataset specific to the direct care workforce.
• Collect robust demographic data in standardized categories and disaggregate the federal standard dataset to facilitate cross-referencing and equity assessments: Currently, limited raw data on the direct care workforce is collected on a national basis containing sufficient demographic indicators for analysts and policymakers to parse out inequities experienced by people with intersecting marginalized identities, particularly women and people of color. A federal standard dataset should include collecting individualized, raw data on race, ethnicity, gender, sexuality, age, immigration status, disability status, and any additional identity categories that workers and stakeholders determine to be relevant, alongside basic job quality data on compensation rates, benefits, service duration and turnover, etcetera. Standardizing the demographic categories according to direct care workers’ preferences is critical to ensure that the resulting data collected from workers across the country, in states with different demographic population distributions and divergent ways of talking about their preferred identity and can be analyzed.

One example to look to in doing this is the recent national effort in Canada to consult the public on statistical standards being developed for data collection on gender and sexual diversity to allow for the reporting of statistically diverse groups of the population in a consistent manner nationally that better reflected how Canadians describe themselves. The responses from the public consultation on preferred self-identification resulted in a range of standardized categories to be used in data analysis, from “LGBTQ2+” as the “least detailed” category level to support comparison across the largest available sample size, to more detailed standardized categories that reflect higher specificity but comprise of smaller sample sizes. In effect, people who reported their gender or sexuality in a “less common” manner of self-identifying can be counted at scale under the LGBTQ2+ umbrella yet also be counted in their preferred self-identification category in the raw data.

To assist policymakers in developing and implementing policy solutions that will truly advance equity and address the job quality issues contributing to worker recruitment and retention issues, the raw data included in the federal standard dataset should be able to be disaggregated by as many identity categories as possible (with individuals de-identified for privacy). The disaggregated categories and measures in the federal dataset can be cross-referenced to identify targeted solutions, such as how caregivers’ earnings or poverty status differs depending on intersecting identities.

• Mandate meaningful worker engagement and participatory, community-led data collection, and utilize a Black Women Best model and “data feminism” framework: A truly equitable data framework for the direct care workforce requires intentional engagement with workers themselves, with an emphasis on consulting workers from a wide range of identities. A participatory data collection process would include designing and conducting surveys and any other data collection methods with workers, rather than “about” workers, such as conducting Feminist Participatory Action Research (FPAR). These methods center workers as “experts by experience,” with external researchers sharing power equally with workers, and ensures workers maintain ownership over their personal information and are compensated accordingly for their expertise and labor. Collection should also require informed consent from

---

participants about what the data will be used for and culturally responsive, respectful question-asking in the respondents’ preferred language and utmost consideration for workers’ privacy and security. The resulting de-identified dataset should be accessible and publicly available to all.

Furthermore, equitable direct care worker engagement means centering women and nonbinary people, particularly Black, indigenous, and women of color, who disproportionately make up the backbone of the caregiving workforce, in data collection, analysis, and any policy decisions implemented using the data. Black women are overrepresented in the direct care workforce and often endure the worst of its inequities. “Black Women Best” is a foundational economic framework that says when Black women’s economic well-being is centered in policy, everyone’s well-being is better off, and the entire economy ultimately thrives. The federal standard dataset and worker engagement strategy should be designed to ensure Black women are set up to succeed and be counted.

- **Center workers in their own words by designing and implementing the first national worker-centered survey of direct care workers:** Centering direct care workers’ lived experiences and position as the true policy experts in their own lives is another key component of solving the care crisis. In addition to standardizing data collection on quantitative and demographic indicators, a comprehensive workforce data equity strategy would include developing a national workforce experience survey to administer to workers across the country on a regular basis, for example, biannually. By engaging workers and stakeholders in developing such a survey, rich experience data can be collected on a large scale for the first time to better understand the challenges facing BIPOC and women direct care workers in their own words and support the advancement of equity and job quality through targeted policy changes and investments.³

Forthcoming worker-centered, participatory research by The Center for Equity will identify sample racial equity and job quality metrics that could be embedded in the direct care workforce system, for assessing baseline working conditions and setting “north star” employment standards.

While collecting comprehensive data on the home care workers employed in the gray market or “off-the-record” in private homes is outside of the scope of the federal standard dataset suggested in this framework given that the requisite data must be provided by government agencies and providers that are already “on record” as a part of the workforce system, other available pathways to count gray market workers should be pursued. Government agencies should spearhead this effort, especially since gray market workers are disproportionately immigrants who face isolating and unprotected working conditions. As PHI (2021) argues, BLS, the U.S. Census Bureau, and other agencies should integrate additional questions.

³ The California Health Interview Survey (CHIS) offers an example of a comprehensive, disaggregated population-based survey: https://healthpolicy.ucla.edu/chis/about/Pages/about.aspx. It is the largest state health survey in the nation in which more than 20,000 Californians in each county are interviewed each year to provide robust data that is statistically representative of the state's diverse population, and renowned for its data on ethnic subgroups and gender & sexual orientation.
related to immigrants in relevant federal surveys, and HHS should also fund new studies on immigrants working in the gray market, with ethical standards for conducting this research without threatening the safety and security of respondents.\textsuperscript{xx}

V. Conclusion

Despite the direct care field being among the fastest growing in the country and in high demand that will only increase as the nation's population ages, the lack of national data standards and wide variance in state data collection results in a spotty picture of the direct care workforce in the United States that seriously undermines efforts to pinpoint policies that would reduce the “workforce shortage”—which is not only a labor shortage, but a shortage of quality direct care jobs that can recruit and retain caregivers.\textsuperscript{xxi}

Data matters, and equitable data systems (or the lack thereof) showcase our nation’s values. The United States should collect robust, worker-centric data on the essential direct care workforce because caregivers matter, as do the millions of people with disabilities, older adults, and families who rely on LTSS each day. We must create a just data system to assess the current challenges facing the direct care workforce and equip leaders with the information needed to design caregiving systems to advance equity and job quality and benefit everyone.

More About Us

The Center for Advancing Racial Equity and Job Quality in Long-Term Care (The Center for Equity) is a national hub for developing innovative workforce policy and fostering narrative change in long-term care (LTC). Our work centers the majority Black, Indigenous and people of color (BIPOC) women and immigrant caregiving workforce and confronts the links between systemic racial and gender inequities and poor job quality. We unite diverse stakeholders in service of building the equitable caregiving infrastructure our nation needs.

The Center for Equity is an initiative of the Healthcare Career Advancement Program (H-CAP), a national labor/management organization which works with employers, unions and workers to raise standards across the healthcare industry. The initiative is generously funded by the W.K. Kellogg Foundation.

Acknowledgements

We would like to express our thanks for the thoughtful comments and recommendations from our external reviewers, Kendra Bozarth, Kate Causey, Gayle Chan, Sarah Gonzaga, Ahimsa Luthuli, Rakeen Mabud, Mariana Morante, Bea Rector, and Allicyn Wilde. We appreciate your insights. Thank you, Kezia Scales, for your initial thought partnership on this topic.
Endnotes

1 "The Racial Equity and Job Quality Crisis in Long-Term Care." (2021 May). The Center for Advancing Racial Equity and Job Quality in Long-Term Care. https://centerforltcequity.org/policy-briefs/


5 See Evelyn Nakano Glenn's seminal 2012 text, Forced to Care: Coercion and Caregiving in America, from Harvard University Press.

6 Ibid.


13 Ibid.

14 Ibid.


