Race Equality Foundation

JRF

insight infrastructure

Exploring Ethnicity Data Use and Gaps in Health Care

Jabeer Butt and Halima Cooper Stubbs

December 2025

Contents

Executive Summary

This report addresses the persistent concerns about the quality, availability and use of ethnicity-disaggregated data, particularly in UK health and care. It presents the findings from a survey and two roundtables exploring how ethnicity data is collected, accessed, and used in health and care, and recommends the changes needed to build a more inclusive, coordinated, and racially just data ecosystem. It is part of the Insight Infrastructure Convening Programme on Ethnicity Data Gaps, a project commissioned by the Joseph Rowntree Foundation (JRF) and delivered in partnership with the Race Equality Foundation (the Foundation). Insights were gathered through an online survey completed by 37 respondents, two roundtable discussions with invited speakers and participants, and a series of blogs designed to prompt wider engagement and reflection.

Respondents and contributors came from academic institutions, public services, and the voluntary and community sector. They described using ethnicity data to identify and understand racial and ethnic inequalities, monitor access to services, support research, and improve service delivery. However, significant challenges persist across all stages of the data lifecycle.

KEY FINDINGS INCLUDE:

- **Purpose:** Ethnicity data is most commonly collected to monitor service access and support improvement. Participants highlighted the need for greater clarity on purpose to build trust and legitimacy.
- **Design:** Most use the 2021 Census classification as a baseline, but many adapt categories to reflect local needs or self-identification and highlights tensions between standardisation and flexibility.
- **Collection:** Concerns among the public often stem from unclear purposes, lack of engagement, or mistrust in data handling practices.
- **Access and Use:** Barriers include licensing restrictions, costs, and technical challenges, particularly for smaller VCSE organisations.
- **Infrastructure:** Fragmentation and lack of interoperability - especially within the NHS - undermine effective data use.
- **Support:** Fewer than one-third of organisations provide training to those collecting ethnicity data. Ethical, intersectional, and anti-racist approaches to analysis are often unsupported.

The report concludes that addressing these challenges is essential to tackling health inequalities. It sets out bold but practical recommendations, including calling for stronger national coordination, investment in interoperable systems, community-led approaches to data design and use, and support for anti-racist analysis. Taken together, these actions would move us from fragmented efforts to a coherent, just, and accountable ethnicity data infrastructure that is capable of driving real and lasting change.

Introduction

This report presents findings from the Insight Infrastructure Convening Programme on Ethnicity Data Gaps, a project commissioned by the Joseph Rowntree Foundation (JRF) delivered in partnership with the Race Equality Foundation (the Foundation). The programme was launched to address persistent concerns about the quality, availability and use of ethnicity-disaggregated data, particularly in health and care.

In recent years, the Covid-19 pandemic, the Windrush scandal, and broader structural inequalities have spotlighted the urgent need, for better data, to understand and respond to racial inequity. While challenges remain, there is a growing recognition that recording ethnicity is essential infrastructure for addressing racial inequality and, improvements to its collection and application have been made across sectors in recent years.

This report draws on both survey data and insights from two roundtable discussions to explore the current state of ethnicity data collection, use and governance. It identifies key challenges for data users including mistrust, fragmented infrastructure, and capacity limitations as well as examples of progress, best practice, and opportunities to build a more equitable and community-led approach to data. It finishes by offering a set of recommendations to improve future data driven action. The report is grounded in an anti-racist framework that understands ethnicity data as more than a technical tool. When gathered and used responsibly, such data can reveal the extent and consequences of racial inequality. When ignored or misused, it can entrench disparities and reinforce harm.

> **“Broader structural inequalities have spotlighted the urgent need, for better data, to understand and respond to racial inequity.**

This analysis is intended to support data producers, practitioners, policymakers and researchers in improving the collection of ethnicity data and developing more inclusive, accountable data practices. It reflects the perspectives of those engaged in front-line service delivery; community advocacy, public health, research, and systems change. The findings are illustrated with charts from the survey and quotations from roundtable participants, in addition to a small number of case studies. The intention is that this offers both analytical depth and lived insight. Throughout, we recognise and highlight both the barriers and the progress that stakeholders have made in strengthening the use of ethnicity data for social change.

Methodology

This report draws on two linked work stages: an online survey and two roundtable discussions. Together, they aimed to capture diverse perspectives and lived experiences relating to the use, collection, and analysis of ethnicity data in health and care.

Work stage one: Survey

We developed and disseminated a 54-question online survey to gather insight into how ethnicity data is collected, accessed, and used. The survey was designed around eight core themes aligned with the data lifecycle and employed branching logic to ensure relevance based on the respondent's role and expertise.

The survey was circulated in two phases over one month: first through targeted invitations to organisations within the Foundation's network, and then more widely via a snowball approach: we reached out directly to organisations and individuals to complete the survey, who then shared it with other relevant organisations, individuals or via social media such as Linkedin. Respondents represented public sector bodies, regulators, academics, and voluntary and community organisations. Responses were analysed both quantitatively and qualitatively.

Work stage two: Roundtables

Roundtable discussions were held via Zoom on 22 and 28 May 2025. These sessions brought together practitioners, researchers, and policymakers to reflect on the survey findings and explore data challenges in greater depth.

Each roundtable included presentations from expert speakers, facilitated discussion, and open reflection. Participants were prompted to respond to key themes raised in the survey and to share their own thoughts and experiences. Transcripts were then coded thematically, and quotations were selected to illustrate key points in the analysis.

The roundtables served as a model for inclusive, dialogue-based data analysis, as they enabled validation of survey themes, offered context specific insights, and revealed new priorities for action.

Supporting engagement through blogs

Throughout the project, we published four blogs authored by guest contributors. These explored issues including racial justice and data, the value and limits of census data, and innovative approaches to bridging ethnicity data gaps. The blogs supported survey participation, built interest in the roundtables, and helped raise awareness of the project's broader aims.

> **The roundtables served as a model for dialogue-based data analysis, as they enabled validation of survey themes (...) and revealed new priorities for action.**

Who participated

Recognising that ethnicity data moves through a lifecycle - purpose, design, collection, analysis, publication, and governance, the project aimed to capture perspectives from across this continuum. To achieve this, we stratified the survey outreach to engage individuals working in different parts of the data infrastructure.

Survey respondents included professionals from central, local, and regional government (including organisations such as the UK Health Security Agency and NHS England); regulators and statutory bodies; academic institutions (including the Centre on the Dynamics of Ethnicity, the Bennett Institute for Applied Data Science, and the Resolution Foundation); as well as representatives from national and local Voluntary, Community and Social Enterprise (VCSE) organisations, particularly those focused on race equality and health equity.

[1] Jabeer Butt on the case for high-quality accessible ethnicity data, Available at: https://raceequalityfoundation.org.uk/press-release/jabeer-butt-obe-on-the-case-for-high-quality-accessible-ethnicity-data/

Rethinking Data with Dr Brenda Hayanga: How methodological imagination can bridge data gaps to address ethnic health inequalities. Available at: https://raceequalityfoundation.org.uk/press-release/rethinking-data-with-dr-brenda-hayanga-how-methodological-imagination-can-bridge-data-gaps-to-address-ethnic-health-inequalities/

A critical assessment of census and survey data in ethnic group research: Insights from Dr Nigel de Noronha. Available at: https://raceequalityfoundation.org.uk/blog/a-critical-assessment-of-census-and-survey-data-in-ethnic-group-research-insights-from-dr-nigel-de-noronha/

Figure 1: Who responded to the survey

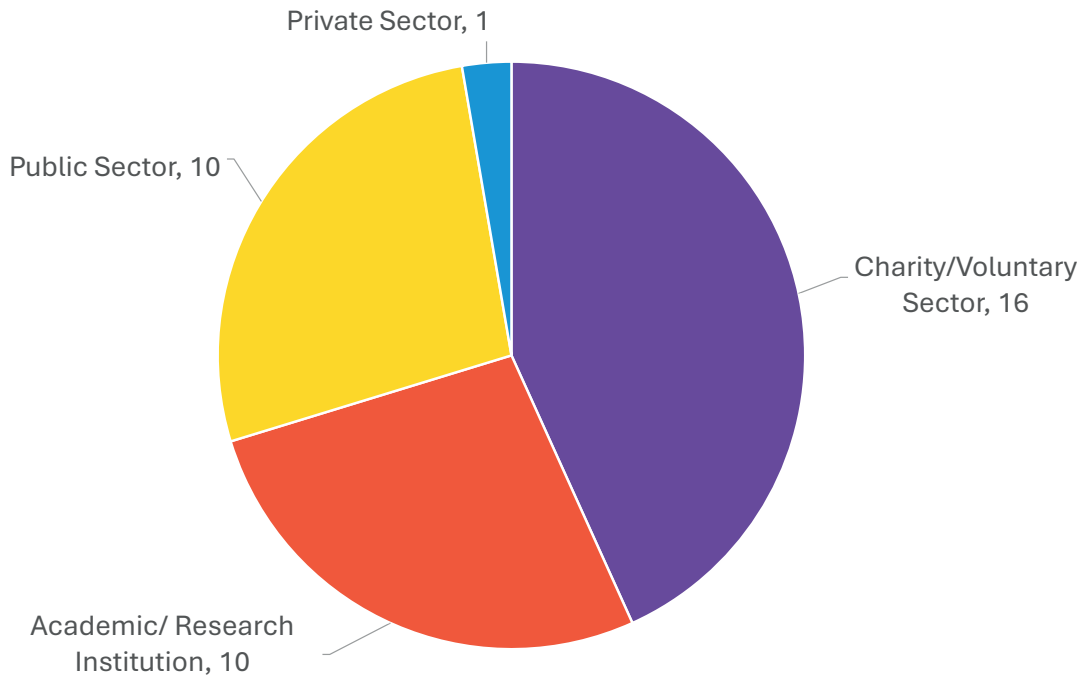
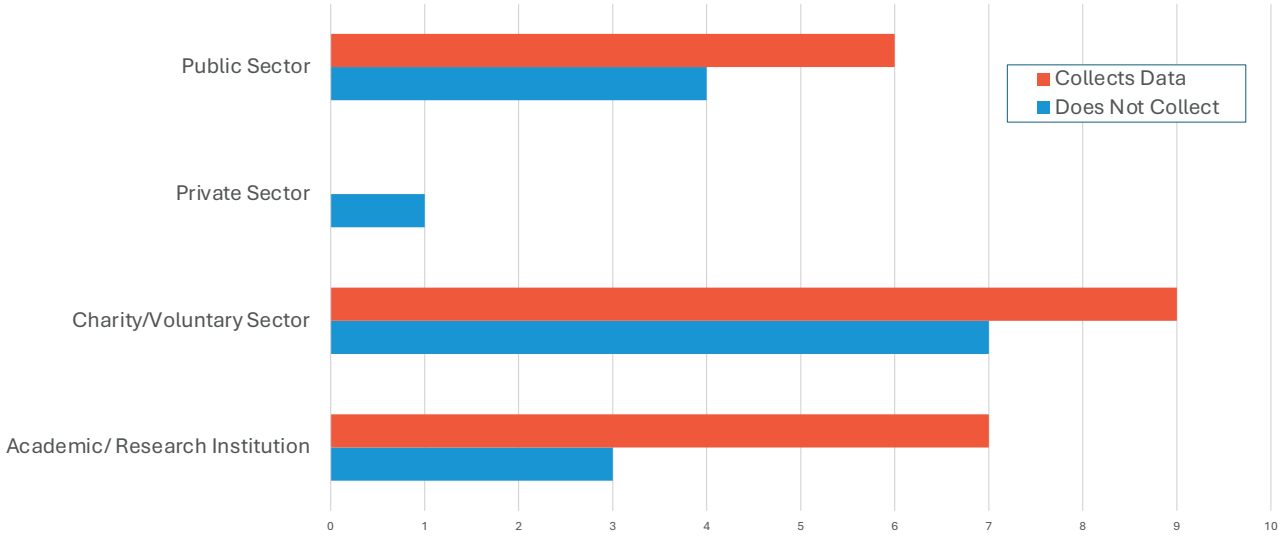


Figure 1 provides a breakdown of respondents by sector. The largest group, 43% of survey respondents, came from the VCSE sector. This weighting is likely the result of the Foundation's networks, and the challenges we encountered in securing survey responses from some public sector and research institutions, often due to their restrictions on sharing internal practices or protocols. We addressed these gaps through targeted invitations to the roundtable discussion, which drew in additional contributions from sectors underrepresented in the survey.

Figure 2: Do respondents collect data



Racism at the root: Tackling ethnic health inequities through racially-just data and policy with Professor Laia Bécares. Available at: https://raceequalityfoundation.org.uk/blog/racism-at-the-root-tackling-ethnic-health-inequities-through-racially-just-data-and-policy/

As shown in Figure 2 a large proportion of respondents were both users of ethnicity data and collectors. This meant respondents were well placed to reflect on both the strategic and practical dimensions of data collection. Notably, almost all respondents from the charity and voluntary sector reported collecting ethnicity data, offering valuable insight into frontline challenges, including how best to engage communities, explain the purpose of data collection, and improve completion rates. These perspectives add depth to the findings, highlighting the everyday realities that shape efforts to build more inclusive, accurate, and trusted data systems.

Limitations

While the project adopted a structured approach to gather findings, the number of participants remains relatively modest: 37 survey respondents and just over 30 individuals attending the roundtables. These contributions, drawn from across sectors, brought significant experience and expertise. However, the voluntary sector was especially well represented in response. While this helped to foreground the experiences of community-facing organisations, it may have also influenced the emphasis of some findings.

Therefore, the findings should be interpreted with some caution. We would not claim the sample is representative of all those engaged in the production and use of ethnicity data, and generalisations beyond the participant group are limited. Nonetheless, the depth and diversity of contributions, including from those directly engaged in collecting and applying ethnicity data, meant that the approach is valuable in generating practical, grounded insights into the barriers and enablers of effective ethnicity data use.

> “The approach is valuable in generating practical, grounded insights into the barriers and enablers of effective ethnicity data use.

What We Found

The following section presents the key findings from the survey and roundtable discussions.

The primary aim of this research was to explore gaps in ethnicity data. However, the evidence gathered pointed to wider issues: inconsistency in data quality, fragmented use, and limited institutional prioritisation across sectors. These challenges are compounded by poor communication around the purpose of ethnicity data, and unequal access to data sources.

At the same time, the research did identify examples of innovation and progress, from community-led data collection and methodological adaptations to more inclusive survey design and administrative improvements. But these advances appear uneven and often isolated from mainstream systems and infrastructure.

The findings below begin by making the case for anti-racist, purpose-driven data collection as the foundation for all improvements. We explore what this looks like in practice and why it matters.

The section follows the full data lifecycle - from purpose and design to collection, quality, use, access, infrastructure, and training, spotlighting persistent challenges for data users, and surfacing actionable solutions at each stage.

Purpose: Naming racism and enabling change

A central theme emerging from both the survey and roundtable discussions was that data should serve an explicitly anti-racist purpose. Participants stressed that data collection should not be a passive bureaucratic exercise, but a means to understand and redress structural inequality. This was seen as essential for capturing lived realities and evidencing policy changes to address inequalities.

Survey responses (see Figure 3) confirmed that the most common reasons for collecting ethnicity data were to monitor service take-up and improve service design. These were seen as activities that reflect an active, responsible use of data. Several organisations shared how ethnicity data was being used to highlight inequalities in service access, influence programme redesign, and inform strategic goals. There was also clear recognition that longitudinal datasets, which track the same data over time such as Understanding Society, sector-specific dashboards, and local authority initiatives have made tangible progress in making ethnicity data more visible and actionable.

When used well, ethnicity data supported:

- Service redesign to better meet diverse needs;
- Case-making for targeted interventions;
- Monitoring inequalities in access to care, outcomes, and standards of living.

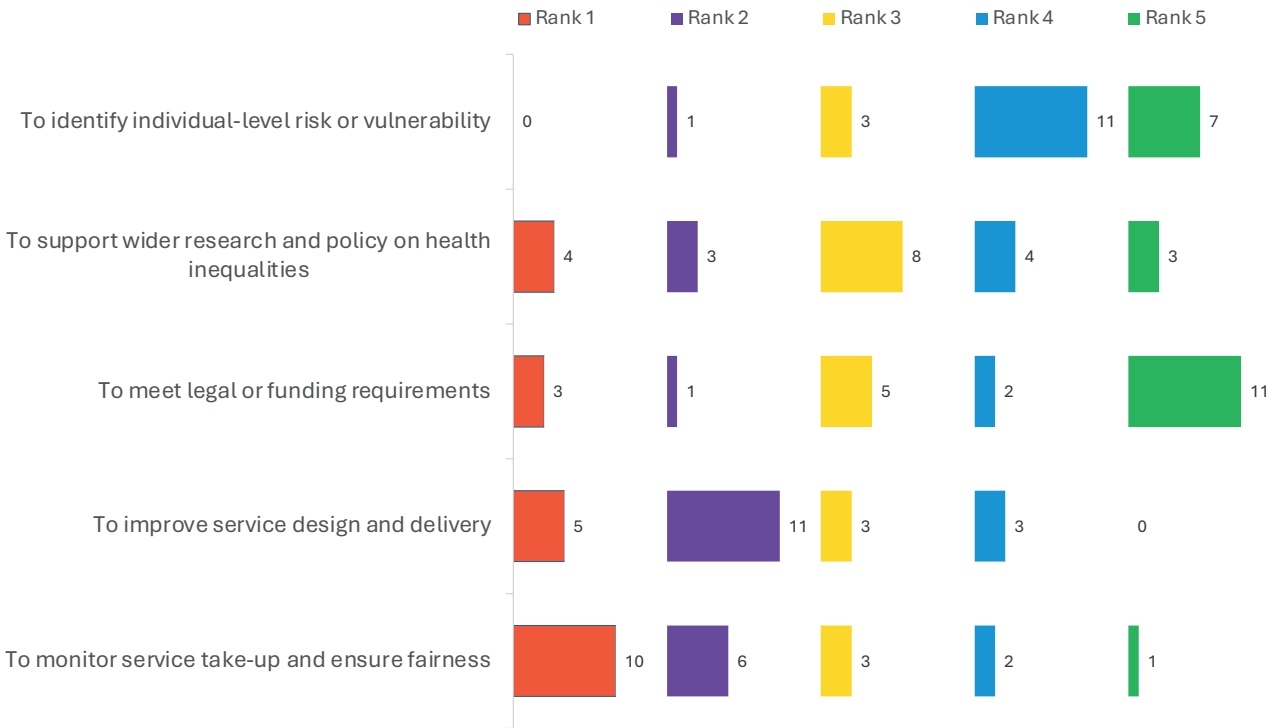
However, participants argued to further embed anti-racist principles into data practices, such as making the connections between racism, health, and socioeconomic outcomes more explicit, for example, the roundtable participants discussed how administrative and survey data can reinforce harm when divorced from community experience or interpreted without a racial justice lens.

The starting point for participants was clear: ethnicity data must have a purpose, and that purpose must be anti-racist. Rather than treating ethnicity as an individual trait or risk factor, participants urged a shift toward recognising the structural causes of racialised health and care inequalities and using data to address them.

> “We’re obsessed about collecting the data rather than what is it that we’re going to do with it.

The survey data captured in Figure 3 saw respondents rank to ‘monitor service take-up and ensure fairness’ as well as to ‘improve service design and delivery’ as the key drivers to collect ethnicity data with ‘[identifying] individual level risk or vulnerability’ ranked as the least important reason for collecting ethnicity data.

Figure 3: Ranking of reason for collecting data by those who record ethnicity



Participants noted another concerning trend in the use of ethnicity data: it is often collected without a clear plan for analysis or application. This can result in data disappearing into inaccessible systems or never being shared back to communities. Participants emphasised this reinforces distrust and scepticism with data collection.

> **People don't have confidence that their data is going to be kind of meaningfully used (...) in developing policy and practice.**

To ensure ethnicity data is used to address racism and structural inequalities, participants argued that clarity of purpose is essential. It is reasonable to conclude that clearly explaining how ethnicity data will be used can help justify collecting it and build the trust that communities are calling for. There was also optimism that data users have a desire to do more with data that is collected, if supported by the right frameworks, training and infrastructure. This is evidenced in case study 1 and 2.

Case Study 1: Centring Racism in Ethnicity Data Collection and Analysis

Contributor: Professor Laia Bécares, King's College London

Professor Laia Bécares used her contribution to the Insight Infrastructure Programme to challenge an entrenched assumption in health data analysis, that ethnicity is an individual risk factor. Instead, she argued for a racially just approach, one that sees racism, not ethnicity, as the fundamental and changeable cause of health inequalities.

Drawing on her research using longitudinal data from Understanding Society, Bécares demonstrated how racism harms health directly and indirectly. The research found that experiences of racial discrimination have immediate and damaging effects on mental and physical health. Racism was also shown to undermine household income over time, poorer health. These findings centre the tole of racism in leading to poor health of minortised ethnic groups both directly, and indirectly by leading to lower socioeconomic positions and illustrated that addressing health inequalities through socioeconomic levers alone will not be effective unless the role of racism itself is confronted.

Using this research, Bécares criticised the common analytical approaches that treat ethnicity as a fixed, behavioural, or cultural variable, devoid of context. This framing, she argued, essentialises ethnic identity and obscures the structural processes that produce inequality. A racially just framework instead views ethnicity as a social construct shaped by racialisation and differential access to power and opportunity. It demands that researchers go beyond description to ask what produces inequality and who is responsible for it.

> "A racially just framework (...) demands that researchers go beyond description to ask what produces inequality and who is responsible for it.

The implications for data collection in taking forward a racially just framework are also profound. Bécares noted the lack of recent survey data with sufficiently large samples of minoritised groups, citing that the last boosted sample in the Health Survey for England was more than a decade ago. She called for better-funded, inclusive surveys designed in collaboration with communities, and for the routine inclusion of detailed life-course measures of racial discrimination. An example of such measure is that included in the Evidence for Equality National Survey (EVENs), which Bécares developed with colleagues from UK, the US and New Zealand.

At its core, Bécares's case is not just methodological, but ethical. Without frameworks that name and measure racism, data risks reinforcing the very inequalities it seeks to document. A shift toward racially just approaches in health research and policy are therefore not optional, but foundational to meaningful change.

Design: Standardisation, usability and flexibility

Respondents and roundtable participants consistently affirmed that ethnicity data collection has improved, particularly in sectors such as health and local government. The use of Census categories in administrative datasets is now widespread, and there is broad agreement that it is better to collect data, even if imperfect, than not at all.

Yet, participants noted, there is still a lack of consistency in how ethnicity-disaggregated data is collected and presented across datasets. Participants suggested this may be because data collection is poorly designed, resulting in datasets which do not capture the most useful information.

Common limitations in data sources identified by survey respondents included:

- Categories too broad (54%);
- Lack of disaggregation (23%);
- No option to reflect dynamic or intersectional identities.

Expanding on these data limitations, roundtable participants shared specific gaps in coverage such as the absence of Gypsy, Roma and Traveller ethnic categories; NHS classifications still use 2001 census categories (see case study 2) inconsistent categorisation across datasets, and missed opportunities to capture intersectional information.

To address inconsistencies across datasets and improve usability, participants advocated for greater alignment on the use of 2021 Census categories and recognised recent efforts by the Office for National Statistics (ONS) and NHS Digital to improve standardisation. Several respondents also offered examples of their efforts to clean existing datasets, train staff, and reduce overuse of "mixed" and "other" categories. These positive examples should inform broader implementation.

However, designing ethnicity data systems with greater standardisation did raise a key concern for participants. Although participants strongly supported harmonisation, especially consistent use of 2021 census categories, there was recognition of the need for flexibility in standardised categories. Such flexibility was seen as necessary for categories to reflect how individuals understand and express their identities, and because participants were certain that categories will change over time in accordance with shifting identities, and population demographics.

> "[The] best balance to strike is (...) overarching categories, but within that give the option of self ID.

To address the issue between harmonisation, and flexibility, participants pointed to a successful hybrid approach: combining standardised tick-boxes with free-text fields, to preserve comparability while allowing for nuance.

"Ethnicity is socially constructed and fluid (…) the categories that we use are not straightforward."

This tension between comparability and inclusivity underscores the need for improved purpose-driven design. A well-balanced approach, as offered by participants, should attempt to ensure both statistical robustness and meaningful representation of peoples lived identities.

Case Study 2: Under-recorded and Overlooked – Gypsy, Roma and Traveller Communities in Health Data

Contributor: Sarah Mann, Friends, Families and Travellers

Sarah Mann, CEO of Friends, Families and Travellers offered a reminder of what happens when ethnicity data fails to reflect the diversity of communities it is meant to serve. Drawing on FFT's frontline and policy work, Mann highlighted the chronic invisibility of Gypsy, Roma and Traveller communities in national datasets, even in cases where the impacts of inequality are both visible and severe.

Despite being among the most disadvantaged ethnic groups in the UK, Gypsy, Roma and Traveller populations are frequently absent from key data collections. Mann noted that while Gypsy, Roma and Traveller ethnicities are recorded in some systems, such as school-level education data, it is missing from others, including many NHS datasets and major labour market surveys. This inconsistent categorisation undermines efforts to monitor inequalities, compare outcomes, or tailor services to community needs.

Mann called for decisive action: the 2021 Census included improved ethnicity categories that allow respondents to self-identify as Romany Gypsy/Irish Traveller, or Roma. Disaggregation would be preferable, but these categories are ready to be adopted, and delaying their implementation will only deepen exclusion. Harmonisation alone, however, is not enough. Mann stressed the need for systems to apply these categories meaningfully, with sufficient granularity, visibility, and engagement with the communities they describe.

As one local authority official put it during a parliamentary hearing: "You cannot commission for what you do not know." Without full inclusion in ethnicity data, the systemic inequalities faced by Gypsy, Roma and Traveller communities remain hidden and so do the opportunities to address them.

Collection: Improving trust and communication

An area of broad agreement was the importance of training and supporting frontline workers collecting ethnicity data. Several organisations described internal initiatives to train staff in how to ask about ethnicity respectfully and how to explain why it matters. Other examples included training programmes focused on data interpretation, intersectional analysis, and the ethical use of personal data. Producing internal guidance or toolkits to support this work was also recommended.

Nonetheless, participants identified an ongoing shortfall in racial literacy and analytic capacity. It was recognised that greater investment into workforce development was required, particularly in frontline services and smaller VCSE organisations to support this training and data capacity building.

Another concern raised was about how well the purpose of data collection is communicated to service users. It was noted that even with well-designed frameworks, effective data collection depends on how it is introduced and explained. But participants explained that some data collection practices remain extractive, offering no feedback loop or explanation of use. Participants thought that this extractive approach amplifies service users' concerns, mistrust, and thus, reluctancy to provide personal information. Training on racial literacy and communication were, therefore, linked with improved reporting, quality, and completeness of collected data.

> **"The mistrust from residents in terms of why the data has been collected (...) comes up again and again.**

"Still common for poor explanation of importance of recording (…) therefore continuing low reporting."

These points were reflected in the survey responses. Respondents were asked to report common reasons for public resistance, those most frequently chosen included:

- Privacy concerns (58%);
- Not identifying with categories (53%);
- Lack of transparency or concern about misuse (26%).

Participants did, however, identify solutions to these data collection barriers, such as:

- Co-producing surveys with community groups;
- Embedding data collection at trusted touchpoints;
- Training staff to explain the purpose and importance of data collection.

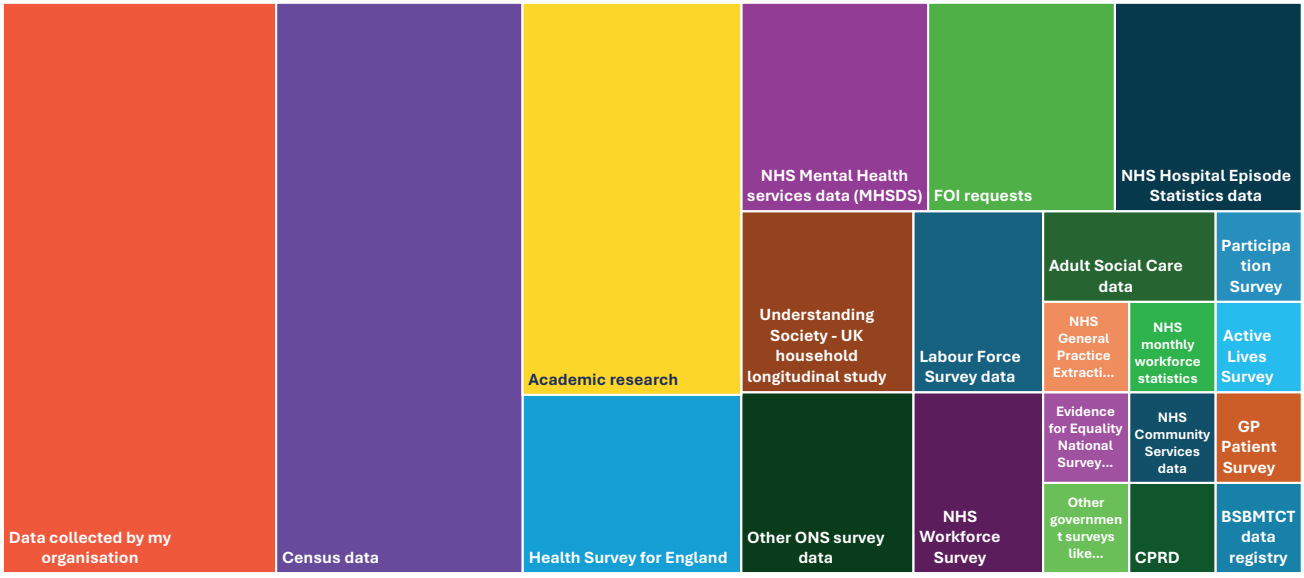
It is possible to surmise from the discussions, that building trust at the point of data collection is not optional, it is fundamental to improving completeness and quality. Not surprisingly, the view was that when communities understand how their data will be used to tackle inequality, they are more willing to share it.

Quality: Usability, accuracy and intelligent use

Figure 4, below, illustrates the range of ethnicity data sources used, revealing both the diversity, and fragmentation of the current ethnicity data environment. The most used source was data collected directly by respondents' own organisations, followed by Census data. Many also relied on NHS datasets, academic research, ONS surveys, and other administrative sources such as the Health Survey for England and Hospital Episode Statistics. This breadth of sources demonstrates the value placed on ethnicity data across sectors. Yet, for participants, it also pointed to a key challenge, different datasets have variations in quality, format, ethnicity categories, and completeness. This variability limits the data's usability, like the ability to link, compare or triangulate across datasets.

Participants did suggest measures to overcome these inconsistencies by blending sources, applying caveats, or supplementing quantitative data with qualitative insights. These suggestions emphasise the importance of transparent, and context-aware data analysis.

Figure 4 Ethnicity data sources used



Several other significant limitations in data and data sets, used by participants, were also mentioned. Most frequently cited limitations included:

- Missing/incomplete data;
- Overuse of “Other” as a catch-all category;
- Inconsistencies between datasets or over time.

Nonalignment with the 2021 census and non-disclosure or suppressed data were also referenced as concerning persistent limitations.

Some respondents shared strategies to overcome these challenges, such as:

- Triangulating ethnicity data with other administrative records;
- Supplementing with qualitative research;
- Publishing findings alongside caveats and known gaps (“intelligent transparency”).

> “Rather than waiting for the perfect data (...) use the data which is available but be transparent.

The survey and roundtables indicate that despite limitations in available ethnicity data remains a vital tool for understanding and addressing racial inequality. It was recognised that frustration with poor data quality could not be used as an excuse for inaction, as Case Study 3 illustrates. However, the challenges in leveraging imperfect data were acknowledged, as it requires sufficient training, funding and technical skills to be done effectively.

Case Study 3: Methodological Imagination in Ethnicity Data Analysis

Contributor: Dr Brenda Hayanga, City St George's, University of London

Dr Brenda Hayanga made the case for what she termed "methodological imagination": the ability to work creatively and rigorously with imperfect data to uncover ethnic inequalities in health. Her reflections, grounded in personal research experience, offered an illustrative example of how limitations in ethnicity data need not prevent meaningful analysis, but rather demand a more resourceful and critical approach.

Hayanga acknowledged, while ethnicity data recording in the UK had improved over the years, data challenges faced by many researchers still persist such as routine health and care datasets riddled with inconsistencies, particularly in the recording of ethnicity. These gaps, like the overuse of 'Other' categories or inconsistent coding practices, disproportionately affect minoritised ethnic groups and risk obscuring patterns of inequality . The absence of detailed or reliable ethnicity data in sectors like social care further narrows the field of inquiry .

Yet rather than waiting for the perfect dataset, Hayanga described how researchers can draw on multiple data sources, combining quantitative analysis with qualitative follow-up or cross-validating survey and administrative data. In her own study on social isolation and loneliness among older people from minoritised ethnic backgrounds, she revised her methodology to incorporate both survey data and a mixed-methods synthesis. Despite the challenges she was able to identify significant disparities in social connectedness between white and minoritised ethnic older adults. Her work, therefore, reinforces a central message of the Insight programme: data use must be guided not just by technical capacity, but by purpose, ethics, and imagination.

However, Hayanga also cautioned that this kind of methodological adaptability comes at a cost. It demands time, funding, and interdisciplinary expertise, and may delay the translation of research into policy and practice.

Use: From describing to explaining inequality

Survey respondents described multiple ways in which ethnicity data is currently being used, from equality impact assessments to internal audits, needs assessments, and strategy development. Some organisations reported publishing regular breakdowns of ethnicity data, such as Anthony Nolan, Age UK and the Centre for Ageing Better, or contributing to cross-sector data platforms such as NHS England and the Greater London Authority.

The benefits of co-designing and co-producing data collection were also noted. For example, the EVENS study, co-produced with voluntary organisations was thought, by those involved, to have enabled deeper insight into experiences of racial discrimination. Participants thought EVENs was a useful model for both better co-produced data collection, and the effective use of ethnicity data to address racial inequality.

However, participants emphasised that ethnicity data is, often underused and on occasion misinterpreted. One key concern was the reduction of ethnicity to a risk factor, without contextualising the role of structural racism.

> **"It's not enough to identify disparities (...) we need to interrogate the conditions that create them.**

For example, case studies were shared of the consequences of data analysis that overlooks the role of racism in health outcomes. One such case was the limitations of the ONS's experimental data on life expectancy estimates published from 2011 to 2014. It was noted that these estimates, though stated as "in the testing phase and not yet fully developed" and shown to be methodologically flawed, were widely cited without caveat, including in official policy documents.

Participants also shared examples where incomplete data sets revealed systemic issues with collection methods, engagement, analysis or data presentation:

"If significant gaps present (…) the reason may be as simple as: we're just not collecting the data."

[2] Scobie S, Spencer J, Raleigh V. Ethnicity coding in English health service datasets [Online] Available from https://www.nuffieldtrust.org.uk/files/2021-06/1622731816_nuffield-trust-ethnicity-coding-web.pdf Last accessed 14th June 2021. 2021. https://www.nuffieldtrust.org.uk/files/2021-06/1622731816_nuffield-trust-ethnicity-coding-web.pdf. Last accessed 20th August 2025

[3] Raleigh V, Glodbatt P. Note To NHSEI On Ethnicity Recording In Health And Care Records [Online]. 2020. https://www.instituteofhealthequity.org/resources-reports/ethnicity-coding-in-health-records/ethnicity-recording-in-health-and-care-records.pdf. Last accessed 20th August 2025

Access: Democratisation and ethical safeguards

To make use of existing ethnicity data, respondents and participants called for greater democratisation of, and the lowering of access barriers, to data sources. Indeed, the survey data showed ethnicity data, even when collected, is often inaccessible:

- 64% of respondents faced access challenges;
- Around 70% cited lack of publicly available data;
- Licensing restrictions and technical barriers particularly affected VCSE organisations.

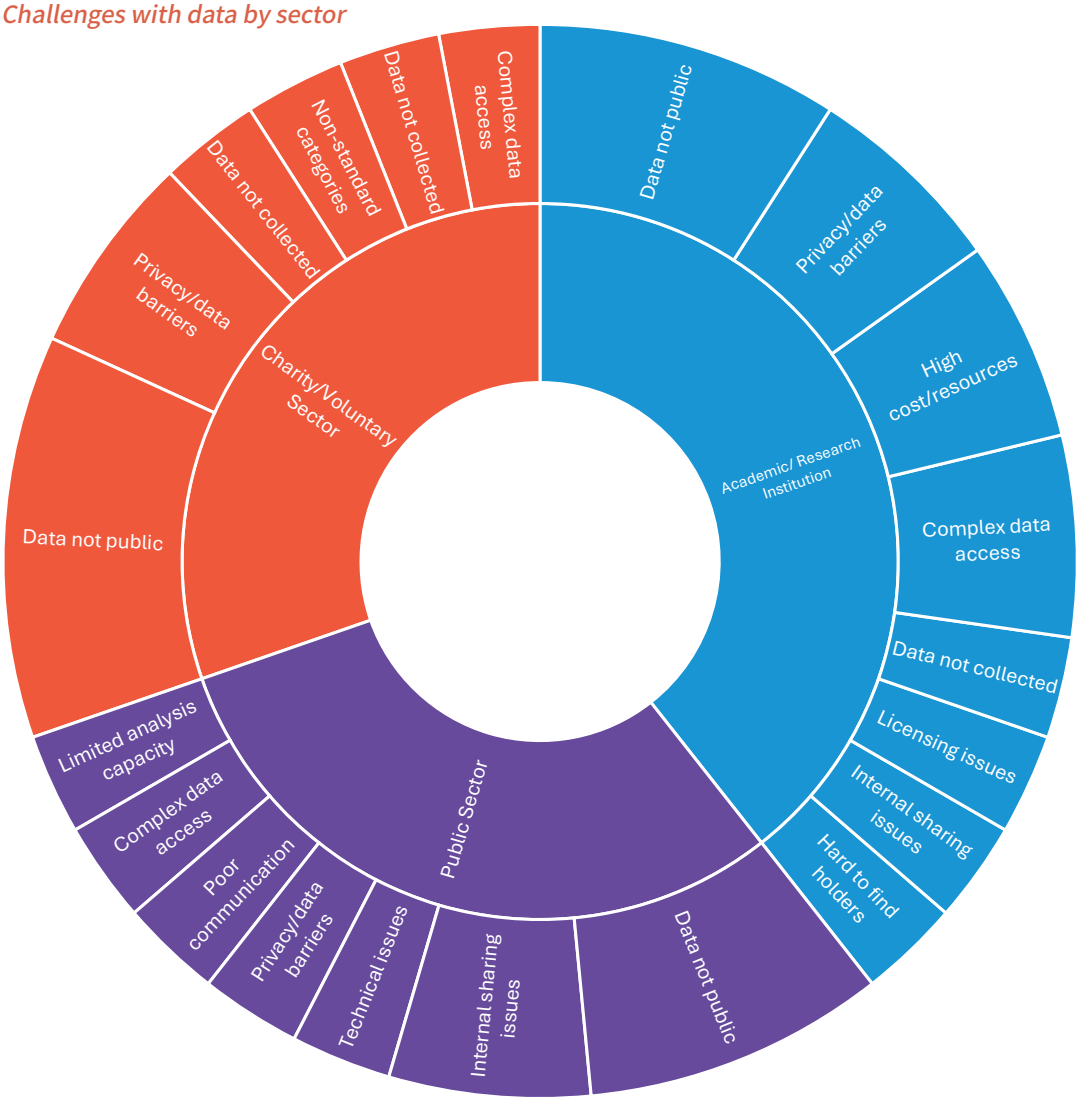
> "If you're a smaller community organisation (...) it can be very difficult to get the intersectional data that you need.

Figure 5, illustrates the types of challenges reported by respondents, broken down by sector. Academic and research institutions cited the widest range of challenges, including privacy concerns, licensing issues, and resource constraints, reflecting the complex regulatory environment in which they operate. Charity and voluntary sector organisations most frequently reported difficulties accessing publicly available data and navigating privacy restrictions. Public sector respondents pointed to internal barriers such as technical limitations, poor communication, and difficulties sharing data within and across teams. These findings indicate that where data disaggregated by ethnicity does exist, it is not always being used effectively because it is not well known or widely available.

Reinforcing the survey findings participants at the webinars highlighted several access barriers:

- Data not published in disaggregated form;
- Datasets not linked across services, which limits the ability of analysts to paint a full picture of someone's experience across different services for an analysis;
- Paywalls or restricted licences limiting analysis. For example, the UK Data Service has larger datasets allowing for intersectional analysis but requires specific organisational licenses, while the cost of a single study license for CPRD data costs at least £15,000.

Figure 5 Challenges with data by sector



Participants called for:

- Democratised access to public data;
- Removal of paywalls for small VCSE organisations;
- Greater transparency about what data exists and how to access it.

These findings reflect a broader recognition that without access, the power of data to drive equity is fundamentally limited.

Yet, this call for democratised access was balanced against the concern that, as a protected characteristic, ethnicity data should be used safely and ethically, with the correct guardrails to protect against misuse. Participants acknowledged the risk that data could be misinterpreted, particularly when taken out of context or applied in ways that ignore structural drivers of inequality.

The way forward, then, lies in building infrastructure and protocols that enable broad, equitable access to high-quality data, while also embedding safeguards that ensure data is used in a way that centres lived experience, protects individual privacy, and challenges racial injustice.

Infrastructure: Fragmentation and fatigue

The lack of coordinated infrastructure across data systems and sources was a major concern, especially within the NHS. Participants described disjointed systems, preventing data linkages between trusts, services, and boards. This can lead to repeated requests from services users for the same data to be supplied at different touchpoints not only exhausting them but undermining trust in the value and use of the data provided

Key infrastructure gaps identified:

- No national framework for data sharing across sectors;
- No interoperability between local, regional and national datasets;
- Over-reliance on outdated platforms and coding systems.

> **“We need a way to share ethnicity data for an individual across trusts and primary care.**

Participants called for:

- Investment in secure, interoperable systems;
- National coordination to link ethnicity data across public services;
- Tools for community organisations to analyse and visualise data.

Improving the way ethnicity data is shared and managed within the NHS is essential to tackling health inequalities. A nationally coordinated, interoperable system would support earlier identification of disparities, reduce duplication, and build trust by showing that data is used meaningfully. With the right infrastructure, the NHS could move from fragmented collection to proactive use of ethnicity data, enabling more equitable care and better outcomes.

Training: Skills, confidence, and racial literacy

Finally, participants emphasised that none of the above improvements are possible without sustained investment in workforce training.

Survey data revealed that:

- Only around 30% of organisations collecting data had mandatory training;
- Just under 60% reported no formal training on how to analyse or interpret ethnicity data.

Participants identified key skill gaps and needs for training to address:

- Racial literacy training for all staff collecting or analysing data;
- Practical toolkits for non-specialist users;
- Senior leadership engagement to embed data use in operational processes.

There is a growing awareness of the need, across sectors, to train and support those working with, recording and using of ethnicity data. Examples were also shared of training focused on data interpretation, data cleaning intersectional analysis, and the ethical use of personal data. For some this was accompanied internal guidance or toolkits to support this work.

However, for many, the available training to support ethnicity data collection and use was not enough. There is still a need for broader investment in workforce development, particularly in frontline services and smaller VCSE organisations. Calls were also made for nationally coordinated training and technical assistance programmes, as well as funding to support data roles within community organisations. These investments into training and workforce development, were thought to be at the core of improving better data use like improving racial literacy, leverage methodological and analytical skills to deal with data gaps and overcoming technical barriers.

Conclusion: Towards a racially just data ecosystem

This research paints a nuanced picture of where we are and where we need to go. Ethnicity data is being used in important ways across sectors including to uncover health inequalities, monitor service access, inform research, and improve practice. Academic researchers, public bodies, and voluntary and community organisations are each bringing distinct strengths to this ongoing work.

However, our findings clearly reveal that the potential of ethnicity data remains constrained by persistent challenges in quality, categorisation, infrastructure, and access. Participants across the survey and roundtables described how fragmented systems, limited training, inconsistent approaches to classification, and unclear communication continue to undermine trust and reduce the impact of data that is already being collected. Small VCSE organisations, in particular, face structural barriers to accessing and using data, despite often being best placed to act on what the data reveals.

At every stage of the data lifecycle; purpose, design, collection, access, use, and governance, this report has offered both critical insights and practical solutions. It has highlighted tensions that need to be navigated: between standardisation and categorisation flexibility; between individual privacy and the need for linked datasets; between urgent action and the limitations of imperfect data. Addressing these tensions requires coordinated investment, national leadership, and a commitment to supporting organisations and communities with the tools they need to interpret and apply data meaningfully.

The confirmation of a 2031 Census offers an opportunity to revisit longstanding debates about categorisation, harmonisation, and the scope of data collection. But census reform alone will not resolve the deeper challenges highlighted in this report. What is needed is a shared commitment across sectors to building data systems that are transparent, inclusive, and explicitly focused on advancing racial and ethnic equity in health.

A racially just data ecosystem will not emerge by accident. It must be built, deliberately, collaboratively, and with those most affected at its heart.

Recommendations

For Policymakers and National Bodies

- **Develop an anti-racist national data strategy:** Ethnicity data collection must be driven by a commitment to racial equity, with structures to ensure accountability, transparency, and redress.
- **Invest in digital infrastructure and interoperability:** Ensure systems used across health, care, and research sectors can collect, linking, and analysing high-quality ethnicity data.
- **Support co-production in data governance:** Embed the voices of people of Black, Asian and minoritised ethnic backgrounds in the design and oversight of data policies.
- **Commit to the 2031 Census as a minimum baseline:** Ensure census categories remain relevant and provide the investment needed to deliver inclusive, high-quality data.

For Local Systems and Public Sector Organisations

- **Standardise ethnicity categories while allowing for local nuance:** Use national standards (e.g. ONS) but engage communities to ensure relevance and respect.
- **Use data to inform resource allocation and service redesign:** Shift from descriptive reporting to action-oriented analysis that addresses structural disparities.
- **Strengthen data access and transparency:** Create clear, equitable routes for researchers and VCSE partners to access anonymised datasets.
- **Prioritise staff training:** Provide mandatory, co-developed training for those collecting or analysing ethnicity data, including on ethical use and anti-racism.

For Voluntary, Community and Social Enterprises

- **Build capacity for community-led data collection:** Fund and support trusted community organisations to gather and analyse ethnicity data in ways that reflect lived experience.
- **Use data to challenge and advocate:** Strengthen the use of data in influencing public policy and service provision, especially through partnerships and coalitions.

For Researchers and Funders

- **Fund equity-centred research:** Prioritise studies that address racialised health inequalities and ensure community benefit is a condition of funding.
- **Challenge data exceptionalism:** Avoid treating ethnicity data as inherently sensitive or difficult—recognise the ethical risks of not collecting or using it well.
- **Invest in methods development:** Support research that improves how ethnicity is conceptualised, measured, and understood across diverse communities.

Resources

Joesph Rowntree Foundation, Insight Infrastructure, Toolkit

Access the toolkit here: **https://insightinfrastructure.co.uk/**

The Insight Infrastructure programme aims to enable the use of timely and actionable insights to support social change. It works towards democratising access to high-quality data and evidence through open collaboration and innovation, improve and link up existing data, unlock new data sources and enable others to take data driven action. This has led to a number of outputs including a 'trusted toolkits' for social change-makers to guide their decisions and actions with quality data.

Blogs

Commissioned blogs for this programme of work, on understanding data limitation, improving the use of ethnicity data, and applying a racially just approach to ethnicity data collection and analysis. All four blogs are available here:

- Jabeer Butt on the case for high-quality accessible ethnicity data, Available at: **https://raceequalityfoundation.org.uk/press-release/jabeer-butt-obe-on-the-case-for-high-quality-accessible-ethnicity-data/**
- Rethinking Data with Dr Brenda Hayanga: How methodological imagination can bridge data gaps to address ethnic health inequalities. Available at: **https://raceequalityfoundation.org.uk/press-release/rethinking-data-with-dr-brenda-hayanga-how-methodological-imagination-can-bridge-data-gaps-to-address-ethnic-health-inequalities/**
- A critical assessment of census and survey data in ethnic group research: Insights from Dr Nigel de Noronha. Available at: **https://raceequalityfoundation.org.uk/blog/a-critical-assessment-of-census-and-survey-data-in-ethnic-group-research-insights-from-dr-nigel-de-noronha/**
- Racism at the root: Tackling ethnic health inequities through racially-just data and policy with Professor Laia Bécares. Available at: **https://raceequalityfoundation.org.uk/blog/racism-at-the-root-tackling-ethnic-health-inequities-through-racially-just-data-and-policy/**

Improving life expectancy data disaggregated by ethnicity

Research carried out by a team at the University of Manchester, King's College London, and the University of York tested ONS experimental life expectancy data between 2011 –2014. Despite, consistent evidence showing the people from Black, Asian and minoritised ethnic backgrounds experience worse health outcomes than the White British Population, these ONS estimates suggested that minoritised ethnic groups in England and Wales have higher high life expectancies.

The subsequent research revealed how data shortcomings led to errors in the conclusions on life expectancy. Concerningly, these statistics, despite experimental, directly informed public policy and health service provision. Read the research here:

- Taylor, H., Becares., L., Kapadia, D., Nazroo, J., Stopforth, S and White, C. (2024) 'Examining assumptions and missing data concerns around experimental life expectancy estimates.' Available at: **https://kclpure.kcl.ac.uk/portal/files/299821661/Ethnic_Inequalities_in_Mortality_in_England_and_Wales_MPO_FINAL.pdf**
- Summary of the research also available here: **https://nhsrho.org/blogs/life-expectancy-estimates-are-affected-by-missing-data-and-methodological-assumptions-why-we-should-not-rely-on-experimental-estimates/**

Methodological imagination in practice

Dr Hayanga's research leveraging and triangulating multiple ethnicity datasets and methods illustrates the methodology required to overcome limitations is available here:

- Hayanga, B., Kneale, D. and Phoenix, A. (2021). Understanding the friendship networks of older Black and Minority Ethnic people living in the United Kingdom. Ageing and Society, 41(7), pp. 1521–1540. doi:10.1017/s0144686x19001624.
- Hayanga, B., Stafford, M., Saunders, C.L. and Bécares, L. (2024). Ethnic inequalities in age-related patterns of multiple long-term conditions in England: Analysis of primary care and nationally representative survey data. Sociology of Health & Illness, 46(4), pp. 582–607. doi:10.1111/1467-9566.13724.

Intelligent Transparency: Openly communicating limitations in datasets

The Office for Statistics Regulation defines 'Intelligent Transparency' as working in an open way when referring to data and statistics in the public domain. Transparency and clarity support public confidence in analysis and the organisations that produce analysis and minimise the risk of misinterpretation

Further guidance on how to use intelligent transparency is available here: **https://osr.statisticsauthority.gov.uk/transparency/**

The EVENS Survey

Evidence for Equality National Survey (EVENS), Centre on the Dynamics of Ethnicity: the largest and most comprehensive survey to document the lives of ethnic and religious minorities in Britain during the pandemic.

The full EVENS dataset including measures on health, social cohesion, socioeconomic characteristics, and racism, available here: **https://www.ethnicity.ac.uk/research/projects/evens/code-research-projects-evens-data/**

Further details on the unique design of the EVENs study, and collection process including non-probability survey approach and asset-based approach; co-creating the survey with community organisations, is available here: **https://www.ethnicity.ac.uk/research/projects/evens/code-research-projects-evens-about/**

Accessible data on ethnicity, ageing and inequality

The Centre for Ageing Better produced a set of statistics collating evidence from the English Longitudinal Study of Ageing and Understanding Society survey data, on inequalities in health, wealth and life circumstances for Britian's older generations. Available here: **https://ageing-better.org.uk/resources/ageing-inequality-ethnicity-evidence-cards?page=4#:~:text=Older%20generations%20are%20becoming%20more,statutory%20statistics%20and%20data%20monitoring**

More recent reports on the state of inequalities for ageing populations, produced by the Centre for Ageing better, can also be found here: **https://ageing-better.org.uk/the-state-of-ageing-2023-4**