City Research Online

City St George's, University of London

Citation: Manzur, H., Blom, N. & Barbosa, E. C. (2024). (Mis)Representing Ethnicity in UK Government Statistics and Its Implications for Violence Inequalities. Social Sciences, 13(5), 235. doi: 10.3390/socsci13050235

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/32850/

Link to published version: https://doi.org/10.3390/socsci13050235

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

social sciences

MDPI

Article

(Mis)Representing Ethnicity in UK Government Statistics and Its Implications for Violence Inequalities

Hannah Manzur [1,*], Niels Blom [1] and Estela Capelas Barbosa [2]

[1] Violence and Society Centre, City, University of London, London EC1V 0HB, UK; niels.blom@city.ac.uk
[2] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 1QU, UK; e.capelasbarbosa@bristol.ac.uk
* Correspondence: hannah.manzur@city.ac.uk

Abstract: The question of how we measure, categorise, and represent ethnicity poses a growing challenge for identifying and addressing ethnic inequalities. Conceptual critiques and qualitative studies highlight the complexities and challenges of measuring ethnicity, yet there remains a lack of quantitative studies investigating the implications of these complexities for inequalities research. This paper addresses this gap by scrutinizing methodological processes and analysing the implications of measurement and categorisation in the Crime Survey for England and Wales (CSEW), critiquing the UK's standardised measurement of ethnicity in national survey data and government statistics. Based on our comparative quantitative analysis of standardised ethnicity categories and regional origins and our evaluation of the CSEW and census' methodologies, we propose an alternative categorisation of ethnicity, focusing on the 'Mixed', 'Asian', and 'Latinx/Hispanic' ethnic groups. Using adjusted crosstabulations and logistic regression models, we found variations in ethnic patterns of violence based on standardised measures and our alternative recategorisation, particularly relating to the distinction between 'Asian' sub-groups, the recategorisation of 'Mixed' ethnicities, and the inclusion of 'Latinx/Hispanic' as a distinctive ethnic group. Our findings reveal valuable insights into the implications of ethnic categorisation for understanding violence inequalities, with significant implications for further policy and research areas.

Keywords: ethnicity; inequalities; measurement; categorisation; violence; survey data

check for updates

Citation: Manzur, Hannah, Niels Blom, and Estela Capelas Barbosa. 2024. (Mis)Representing Ethnicity in UK Government Statistics and Its Implications for Violence Inequalities. *Social Sciences* 13: 235. https://doi.org/10.3390/socsci13050235

Academic Editor: Thomas McNulty

Received: 2 January 2024
Revised: 2 April 2024
Accepted: 5 April 2024
Published: 24 April 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In 2021, a petition to the UK Parliament calling for 'Latinx/Hispanic' to be included in the UK Census highlighted the importance of representation in national survey data, asserting, "We are not white, black, Asian, and certainly not 'other'" (UK Parliament Petition 2021). The measurement and categorisation of ethnicity is often taken for granted in mainstream research and policy, yet it represents a core underpinning logic in how systems of racial and ethnic inequalities operate, change, and persist in society. They not only shape our understandings of these systems and how to tackle them through the production of knowledge for policymakers and service providers, but often signify a source of discomfort, alienation, and misrepresentation among minoritised ethnic groups (Woozeer 2022).

As societies become increasingly ethnically diverse and heterogeneous in a post-colonial and globalised world, the question of how we measure, categorise, and represent ethnicity poses a growing challenge for data producers, researchers, and policymakers seeking to identify and address ethnic and racial inequalities. Conceptual critiques and qualitative studies have well-documented the complexities of measuring ethnicity and the inadequacies of standardised approaches (Burton et al. 2010; Khanna 2010; Aspinall 2000; Ballard 1997), yet there remains a lack of quantitative studies which investigate the implications of these complexities and inadequacies for understanding inequalities

in society. The limited existing evidence from health revealed discrepancies between self-identified and observed race, changing racial categories and identification, and challenges in categorising and counting 'mixed' race groups (Williams 1996). More recently, increased interest in ethnic health disparities following the Covid-19 pandemic highlighted gaps in ethnicity data availability, consistency, and quality (Office for Statistics Regulation 2021; Office for National Statistics 2023a). Besides these, there remains a lack of critical attention to how the measurement of ethnicity may shape our understanding of ethnic inequalities produced through these measures, which in turn shape policy and practice, across a plethora of research and policy areas. The UK produces a wealth of nationally representative survey data across a range of research and policy areas which measure ethnicity using a standardised approach derived from the national census (UK Data Service 2023; Office for National Statistics 2022). These surveys play a significant role in evidencing ethnic disparities in socio-economic (e.g., Family Resources Survey, see Department for Work and Pensions 2022; Annual Population Survey, see Office for National Statistics 2023b), health (e.g., Health Survey England, see NHS Digital 2022; Adult Psychiatric Morbidity Survey, see Rhead et al. 2022), housing (e.g., English Housing Survey, see Garrett et al. 2014), employment (e.g., Labour Force Survey, see Powell and Francis-Devine 2023), and crime research and policy (Crime Survey for England and Wales, see Office for National Statistics 2021). In order to address ethnic inequalities, it is necessary to collect data which reveal the often hidden, yet persistent, ways in which individuals' experiences are shaped by their membership of particular ethnic groups (Williams and Husk 2013). How ethnicity is measured is therefore of significant relevance.

Minoritised ethnic groups have long been subject to elevated levels of institutional and interpersonal violence, and ethnic inequalities in the distribution of violence victimisation and fear persist (e.g., Finney et al. 2023; Kumar et al. 2020; Bruce and Roscigno 2003). Violence in society is perpetrated and regulated through systems of inequality (Walby 2009), which may affect minoritised groups directly, through elevated risk factors and victimisation rates, and indirectly, through a heightened sense of fear and anticipation of violence even when not directly subjected to physical violence. Both victimisation and fear of violence are significant in capturing inequalities in the way people's lives are shaped by violence (Bjørnholt and Hjemdal 2018).

Whilst the links between how we measure *violence* and our understanding of inequalities in violence have been investigated in several recent studies (Pullerits and Phoenix 2023; Walby et al. 2016; Cooper and Obolenskaya 2021), far less focus has been placed on how we measure *inequality groups*, particularly ethnic groups. A central challenge in measuring inequalities is whether and how nationally representative and comparable data on inequalities can be produced without inadvertently excluding or misrepresenting the complexities and variations in self-identities and experiences (Burton et al. 2010; Nerenz et al. 2009). Using quantitative victimisation and attitudinal data on violence in the Crime Survey for England and Wales (CSEW), this paper explores the complexities and challenges of measuring ethnicity in UK national survey data, and analyses the implications of different measurement approaches for understanding ethnic disparities in violence experiences (victimisation and fear). The CSEW derives its approach to measuring ethnicity from the UK decennial census (in line with the UK's GSS standardised approach, UK Government Equality Hub 2023), which undertakes considerable evaluation of its approach to capturing UK demographic data (Office for National Statistics 2009). As such, the methodological critique, statistical findings, and recommendations provided by this study not only apply to the CSEW, but more broadly to the census' approach to measuring ethnicity which sets the standardisation of ethnic measures for most nationally-representative surveys and government statistics in England and Wales (e.g., Office for National Statistics 2009).

Investigating and addressing the measurement of ethnicity is essential to both improving the collection, interpretation, and use of these data, as well as revealing how the underpinning logics, methodologies, and interpretations involved in translating complex self-identities into broad categories shape our understanding of inequalities in society.

This paper investigates the central question—How does the way we represent ethnicity matter for understanding inequalities in violence?—by first investigating what exclusions, categorisations, and interpretations are involved in translating complex, multifaceted self-asserted identities into neat, mutually-exclusive categories in UK survey data, before comparing how alternative measurement approaches could address these challenges through quantitative data analysis of the CSEW. This research seeks to bridge the gap between conceptual deliberations and practical applications by scrutinising methodological processes and revealing the implications of measurement in the CSEW as part of a wider critique of the UK's standardised approach to capturing ethnicity in national survey data and government statistics. This research thus shows how the measurement of ethnicity plays a significant role in our understanding of violence inequalities by demonstrating how alternative categorisations reveal different and hidden patterns of ethnic disparities in violence outcomes.

2. Debates and Challenges in the Meaning and Measurement of Ethnicity

The concept and meaning of ethnicity have been long debated in both academic (e.g., Anthias and Yuval-Davis 2005; Anthias 1992, 2007; Fenton and Bradley 2002; Callister et al. 2009) and public discourse (Bush 2022; Rose and Rose 2005; Schneider 2023; Prewitt 2013), with varying definitions and perspectives. Perspectives differ in their framing of ethnicity as a 'practice' or 'articulation' of ethnic group membership (Anthias 2007), as an embodied identity (Alcoff 1999), or as a system of inequality which structures society (Walby 2009). Drawing on perspectives from Anthias and Yuval-Davis (2005, also Anthias 2007), this paper approaches ethnicity as a complex and multi-dimensional social construct used to define socio-cultural differences and group membership based on a variety of characteristics, including culture, religion, nationality, and language. It acts to position individuals within systems of inequality, a core component of which is the structuring, organising, and categorising of individuals and groups through identities, behaviours, and social relations.

2.1. Ethnicity: Race and Other Dimensions

Whilst ethnicity is considered distinct from 'race', the two are interconnected and often incorrectly conflated in language and measurement. Historically, the measurement of race was widely derived from 'scientific racism' which promoted deeply flawed assertions of biological differences between groups of people based on skin colour and physical characteristics (Saini 2019). Racial measurements and categorisations were used to impose hierarchal structures of racial superiority and have been a central tool in justifying and implementing colonial occupation and exploitation, genocide and ethnic cleansing, and systems of apartheid and segregation (Hoover 2007). The imposition of 'race' as a means of categorising human beings has been extensively critiqued (see, for example, Barot and Bird 2001), and despite concerns of a resurgence of scientific racism (Saini 2019; Morning 2014), the notion of biological race has been widely rejected among social scientists (Wagner et al. 2017). National data from most countries focus on ethnic groups (or comparable proxies) rather than race (the US being a notable exception, Morning 2015). Yet 'ethnicity' is often used as a proxy measure for racial difference and inequalities as 'race' represents a dimension of ethnicity. Whilst ethnicity measures may benefit from capturing aspects of racial inequalities without re-imposing flawed racial categories and logics, the uneven distinction and conflation of ethnicity and race also poses challenges.

Alongside race, further dimensions which form the bases of ethnic groups include language, culture, religion, country or region of origin, ancestry (including parentage), shared histories, and others (Burton et al. 2010). Ethnic groups often draw on multiple dimensions, creating a patchwork of shared experiences and communal bonds which inform ethnic identities, practices, and group memberships. Whilst the multifaceted nature of ethnicity as a concept is reflective of the complex ways such groups are formed and hold meaning, this complexity poses a significant challenge for statistical measurement.

Ethnicity cannot be reduced to a single dimension (such as race or ancestry alone), yet broad categorisations of ethnic groups tend to unevenly prioritise different dimensions across ethnic categories, and survey design recommendations advise against using a single variables to measure multiple indicators (e.g., regional origin, race, singular/mixed ethnicity) (Kautt 2011).

2.2. Complexity versus Practicality

Much conceptual and theoretical scholarship emphasises characteristics of ethnicity which are often at odds with conventional forms of measurement (Burton et al. 2010; Nerenz et al. 2009; Kautt 2011). Whilst specialised surveys offer more nuanced and theoretically grounded approaches to ethnicity measures (Modood et al. 2002; see Burton et al. 2010), standardised measures in national surveys prioritise stability, comparability, and methodological practicality (Burton et al. 2010; Nerenz et al. 2009). Standardised categorisations enable consistent comparison and aggregation of ethnic differences across multiple data sources and research areas, yet must also standardise ethnic groups in a way which meets diverse research needs, is broadly reflective of ethnic variation in the population, and is practical for data collectors (ibid).

Quantifying and analysing ethnic variation involves processes of interpretation and simplification of complex, fluid, and subjective experiences and perceptions of individuals' sense of ethnic identity and group membership. This process of 'fitting' ethnicity into neat, mutually exclusive categories necessarily entails a simplification of the complexities of ethnicity and the specificity of individuals' ethnic identities. Herein lies one of the central challenges in measuring ethnicity: how to retain the specificity and meaning of ethnic groups in a way which is reflective of individuals' asserted identities whilst producing data which are reliable, comparable, and useable in quantitative analyses of ethnic variation.

This paper seeks to untangle the specific challenges in balancing representativeness, whereby respondents recognise themselves in ethnic categories; conceptual coherence, whereby ethnic categories offer consistent and meaningful information for both respondents and researchers; and methodological practicality, whereby ethnicity data are practical for quantitative analyses and align with methodological standards (e.g., sufficient sample sizes, comparability over time and across data sources).

3. Evaluating Standardised Measures of Ethnicity in the UK: Changes, Challenges, and Complexities in UK Survey Data

This study first assesses the challenges, inconsistencies, and limitations of the current approach to measuring ethnicity through a critical evaluation of standardised ethnicity measures in UK survey data (particularly the CSEW), before analysing quantitative data in the CSEW to compare different approaches to measuring ethnicity and its implications for violence-related outcomes.

3.1. National Variations

Official approaches to measuring ethnicity vary widely across countries, based on national context, time period, data source, and mode of collection. A study of 141 countries' ethnicity measures in national censuses found diverse conceptualisations of ethnicity and significant variation in language used, dimensions captured, response options and structures, and question design (Morning 2015). Respondents may be asked about 'nationality', 'indigenous status', 'race' or directly about 'ethnicity', and may respond via open-ended free text answers, multiple-choice or tick-box options, or a composite of multiple questions, with closed-ended check-boxes or code lists using single-selection of pre-set categories among the most common approaches (ibid). This variation restricts researchers' ability to compare ethnic variation across countries and over time, often limiting evidence of ethnic inequalities to country-specific and time-bound contexts.

The European Social Survey, for example, collects data on whether respondents "feel [they] are part of the same race or ethnic group as most people in [their country]" and

the country of origin of their parents (European Social Survey 2020). Whilst this captures self-identified ethnic minoritisation and national heritage, it fails to capture the multiple dimensions of ethnicity, particularly race, which may instead be inferred based on parental origin, but provides an unreliable and vague measure of ethnicity. Early measures in other surveys and administrative data often relied on 'assigned' rather than 'asserted' ethnicity, whereby a person's ethnicity was 'observed' and attributed to them by a researcher, state actor, or other authority, rather than based on the individual's asserted identity and sense of belonging to a particular ethnic group (e.g., Wood 1984, p. 46).

3.2. UK Standardised Measures and the Crime Survey for England and Wales

The UK adopts a standardised approach to measuring ethnicity which enables cross-comparison across different national surveys and analysis of trends over time. The UK uses a two-tier categorisation of high-level (5 groups) and low-level (19) ethnic groups which seeks to capture multiple dimensions of ethnicity. Most large-scale UK surveys use a standardised measure based on the decennial census, with minor variations, including the UK Household Longitudinal Survey (UKHLS), the UK Labour Force Survey (UKLFS), the Adult Psychiatric Morbidity Survey (APMS), the Annual Population Survey (APS), the CSEW, and others. In analysing the implications of the current standardised approach to measuring and categorising ethnicity in relation to violence outcomes, we focus on the implementation of these measures in the CSEW.

The CSEW (formerly the British Crime Survey) was introduced in 1982 as a biannual (annual since 2001) household victimisation survey collecting data from approximately 35,000 respondents per year on a range of crime-related subjects, including detailed information on violent crime. The CSEW was initiated as a reliable measure of criminal offences in England and Wales which, unlike police data, captured victimisation not reported to the police. It has a high response rate compared to other national surveys (70% in 2019, Office for National Statistics 2023c). Alongside police data, it acts as a central data source on crime in England and Wales for government and public services, particularly the UK Home Office.

Despite significant attention and development of the UK Census methodology for measuring ethnicity upon which the CSEW is based (Office for National Statistics 2009), several challenges persist, including ethnic group exclusion and self-identified representative groups; conceptual coherence of categories and conflation of multiple dimensions of ethnicity; accessibility and useability of the data by researchers; and longevity of ethnic groups in the face of a changing demographic landscape and research agenda.

Since its 1982 introduction, the CSEW has recorded ethnic dimensions of violence victimisation (predating the 1991 Census' ethnicity inclusion). Yet how ethnicity is captured and represented has changed over time, reflecting changing socio-cultural norms, political priorities, demographic trends, and methodological developments (see Appendix A, see also Laux 2019). Early CSEW waves used assigned rather than asserted ethnicity (1982, 1984), whereby interviewers recorded 'observed' ethnicity, with the 1982 questionnaire including a category for 'mixed/uncertain' (Wood 1984). Ethnic categories of 'White', 'Black (West Indian or African)', 'Indian/Pakistani/Bangladeshi', and 'Other'/ 'Other non-white' reflected socio-cultural norms around ethnicity language, the UK's political and research priorities, and a demographic landscape shaped by post-colonial migration flows (Kautt 2011).

Standardised ethnic measures in the UK have benefited from stability over time, with the high-level ethnic groups ('White', 'Black', 'Asian', 'Mixed', and 'Other') remaining broadly similar over 40 years of CSEW data, despite growing ethnic diversity during this period, and low-level groups (currently 18 categories) providing more detailed breakdowns of these high-level groupings (Appendix A). This approach balances the need for stable categorisation over time and categories which reflect the growing ethnic diversity in the UK, yet it is not without its limitations. Whilst capturing all ethnic groups in the UK would be impractical for quantitative data collection, the current response options are insufficient in capturing non-specified ethnic groups, who are otherwise categorised as 'Other'. Ethnic

groups including Latinx/Hispanic, Somali, Kashmiri, Jewish, and others have all been raised for potential inclusion due to ethnic specificity and demographic changes, yet remain excluded from measures largely due to sample size (Office for National Statistics 2009). The exclusion of ethnic groups carries implications for representativeness, whereby respondents are unable to recognise themselves in the categories provided and forced to select either an ill-fitting ethnic group or the miscellaneous 'Other' ethnicity which fails to adequately represent respondents' identity or provide researchers with meaningful data.

Although the CSEW has benefitted from the continued demographic relevance of its high-level ethnic categories, the emphasis on stability over time fails to balance the need to anticipate emerging ethnic groups, such as Latinx/Hispanic which is one of the fastest growing ethnic groups in the UK (Blair 2019). The introduction of 'Arab' in 2012/13 was welcomed, although earlier inclusion would have enabled researchers, for example, to capture experiences of violence against these groups during key periods of rising discrimination and hostility against Arabs during the Global War on Terror period. Capturing ethnic diversity and disparities in a population should anticipate rather than try to keep up with demographic trends, in order to avoid frequent and lagging updates to ethnic categories and improve comparability over time.

Aspinall (2009) observes a shift from 'colour' to 'culture' in how ethnicity is defined and categorised in England and Wales, with increased inclusion of regional and national dimensions of ethnicity. The language used in survey questions has also varied, with respondents asked to identify their ethnicity based on self-description ('which group . . . best describes you?'), group belonging ('which of these groups do you consider to belong?'), cultural background ('indicate your cultural background'), or explicit ethnic group ('what is your ethnic group?') across several decades of questionnaires (see also Appendix A). In 2012/13, following 2011 Census changes, the CSEW renamed 'Black' and 'Asian' ethnic categories to 'Black/Black British' and 'Asian/Asian British', which addressed Kautt's 2011 critique of the association of 'British' with 'White' ethnic groups only, whereas 'Black' and 'Asian' groups were associated with non-British regions and nationalities (i.e., 'White-British', 'Black-African', 'Asian-Indian'). This update sought to reflect a more modern approach to nationality and ethnicity which reflected asserted values of diversity and multiculturalism in the UK.

Due to the multi-dimensionality of ethnicity, which combines characteristics of culture, nationality, language, religion, and race, single measures of ethnicity typically conflate dimensions inconsistently (Kautt 2011). Race is captured in high-level 'White' and 'Black' categories, although the former is then disaggregated by nationality (e.g., 'White-British') and the latter by region ('Black-African', 'Black-Caribbean'). 'Asian' is used to denote both race and continental origin and is only disaggregated by South Asian nationalities (Indian, Pakistani, Bangladeshi) in the CSEW and UK standardised measures.

Additionally, the conflation of the multiplicity of ethnicity (i.e., single or mixed/multiple ethnicities) with ethnic group membership, which represent different dimensions of ethnicity, leads to conceptual incoherence in what the categories are measuring. By 'lumping' together respondents from different ethnic groups into a singule category on the basis that they all identify with more than one ethnicity, and further categorising this alongside 'singular' ethnic groups, this approach undermines the meaningful content of the 'ethnic group' variable (Kautt 2011). The UKHLS aligns with the UK standardised approach, although it included additional questions in 2009/10 for coding 'Mixed' respondents, asking "Do you come from, or have parents or grandparents from any of the following ethnic groups?". This approach would enable a greater specificity in multiple ethnicities and a more accurate estimate of 'Mixed' ethnic populations, which studies have shown are likely to be underestimated in survey data (Ahlmark et al. 2015).

As well as issues associated with the categories themselves, the accessibility and useability of ethnicity data were also identified as posing several challenges. Following the 2011 Census which implemented several changes to ethnic categorisation, the 2012/13 CSEW introduced new ethnic groups for 'Arab' and 'Gypsy or Irish Traveller' and recate-

gorised 'Chinese' from 'Chinese/Other' to 'Asian'. Since the release of the CSEW 2012/13, the detailed ethnicity variable (now 18 groups) has been moved from the main dataset to the secure access datasets, significantly reducing researchers' access to ethnicity data and potentially making researchers more inclined to use the five-grouped ethnic measure.

A further challenge in the useability of ethnicity data relates to the need for sufficient sample sizes of ethnic minorities, who not only each represent small sub-sets of the population but also have higher non-response rates and barriers to accessing surveys and disclosure routes (Ahlmark et al. 2015), particularly in relation to violence (Femi-Ajao et al. 2020). In response to researchers' need for sufficient sample sizes for ethnic analyses and increased interest in ethnic disparities, the CSEW introduced an ethnic boost sample in 2001/02, yet challenges in comparability with the main population limited its useability and it was discontinued in 2007/08. Reduced ethnic sample sizes resulting from the discontinuation of the ethnic boost sample in 2007/08, as well as the reduction in the overall sample size in 2012/13, restricted researchers' ability to analyse ethnic variation.

4. Materials and Methods

We used 11 years of pooled data from the CSEW, combining the survey waves 2004–2007 and 2013–2019[1] and combining non-secure victim form and non-victim datasets with secure non-victim form datasets[2] (Office for National Statistics 2023d; Blom 2023). We used face-to-face interview data, with a high (approx. 70%) response rate. Our sample of 426,136 respondents (see summary statistics, Table 1) included all adult (16+) respondents, excluding those with missing data for our main variables (see Appendix B). Five survey waves (2008–2012) were excluded due to missing variables required for recategorisation (country of origin).

Table 1. Summary statistics: frequencies and percentages of individual characteristics, pooled data (waves 2004–2007 and 2013–2019).

	N	*Unweighted %*	*Weighted %*		*N*	*Unweighted %*	*Weighted %*
Respondent Ethnicity (5 groups—ONS categories)				**(Any) violence victimisation**			
White	391,191	91.80	88.58	None	415,361	97.47	97.09
Mixed	3587	0.84	1.06	Violence victim	10,775	2.53	2.91
Asian	17,247	4.05	6.17	Total	426,136	100.00	100.00
Black	9610	2.26	2.78	**Fear of stranger violence**			
Other	4501	1.06	1.41	Not very/at all worried/NA	169,440	69.21	67.59
Total	426,136	100.00	100.00	Fairly/Very worried	75,374	30.79	32.41
Ethnic regrouping				Total [1]	244,814	100.00	100.00
White	390,739	91.69	88.44	**Occupational Class**			
Black/Black British	11,332	2.66	3.27	Higher managerial	12,095	2.84	2.77
Arab/MENA	471	0.11	0.16	Higher professional	29,088	6.83	6.88
South Asian	15,243	3.58	5.54	Lower managerial/higher technical	103,422	24.27	23.62
ESEC Asian	4969	1.17	1.58	Intermediate occupations	52,360	12.29	11.89
Latinx/Hispanic	782	0.18	0.26	Small employers and own account workers	40,901	9.60	9.26
Other	2600	0.61	0.74	Lower supervisory and technical occupations	38,324	8.99	8.74
Total	426,136	100.00	100.00	Semi-routine occupations	65,929	15.47	14.72
Mixed Ethnicity				Routine occupations	52,281	12.27	11.56
Not mixed (single ethnicity)	422,549	99.16	98.94	Never worked	14,297	3.36	3.50
Mixed/Multiple ethnicity	3587	0.84	1.06	Not classified	17,439	4.09	7.05
Total	426,136	100	100	Total	426,136	100.00	100.00
Respondent country of origin—regional groupings				**Respondent Sex (binary male/female)**			
UK	376,941	88.46	85.59	Male	193,978	45.52	48.63
Europe	18,882	4.43	5.22	Female	232,158	54.48	51.37
North America	1635	0.38	0.42	Total	426,136	100.00	100.00
Caribbean	1979	0.46	0.49	**Respondent Migrant-status (binary UK-born/migrant)**			
Latin America	1069	0.25	0.34	UK-born	376,853	88.43	85.57
Africa	8131	1.91	2.34	Migrant (foreign born)	49,283	11.57	14.43
MENA Region	2335	0.55	0.70	Total	426,136	100.00	100.00
South Asia	9729	2.28	3.31	**Continuous Variables (Age and Wave)**			
East/SE/Central Asia	3581	0.84	1.07		*N*	*Mean*	*SD*
Oceana/Other	1854	0.44	0.53	Age (16+)	426,136	51.29	18.58
Total	426,136	100.00	100.00	Wave	426,136	2011.4	5.45

[1] Note that analyses of fear of stranger violence used a reduced sample of 244, 814 respondents. See Section 4.1.

4.1. Current and New Ethnicity Variables

Variables were recoded for comparability and analytical purposes (see Appendix B), with significant variable recoding for our alternative categorisation of ethnicity (see Figure 1). For standardised ethnicity, we used the Office for National Statistics (ONS) ethnicity measures which distinguished between the high-level categorisation of five ethnic groups ('White', 'Mixed', 'Asian', 'Black', and 'Chinese/Other'[3]) and low-level categorisation, providing a detailed break-down of 16 ethnic groups (the latter only consistently available in the secure dataset). Ethnicity recategorisation was determined through cross-tabulations of standardised ethnicity and regional origins (Table 2) and the evaluation of misrepresentation risks (see Section 5). We also recoded the ONS regional categories to further distinguish regions with distinctive (high-level) ethnic origins (e.g., the Caribbean, Middle East, and North Africa, see Appendix C). Based on these findings, we used regional origin to estimate ethnicity in cases where ethnic information was otherwise unavailable (e.g., 'Other') and distinguish between specific groups (see Figure 1).

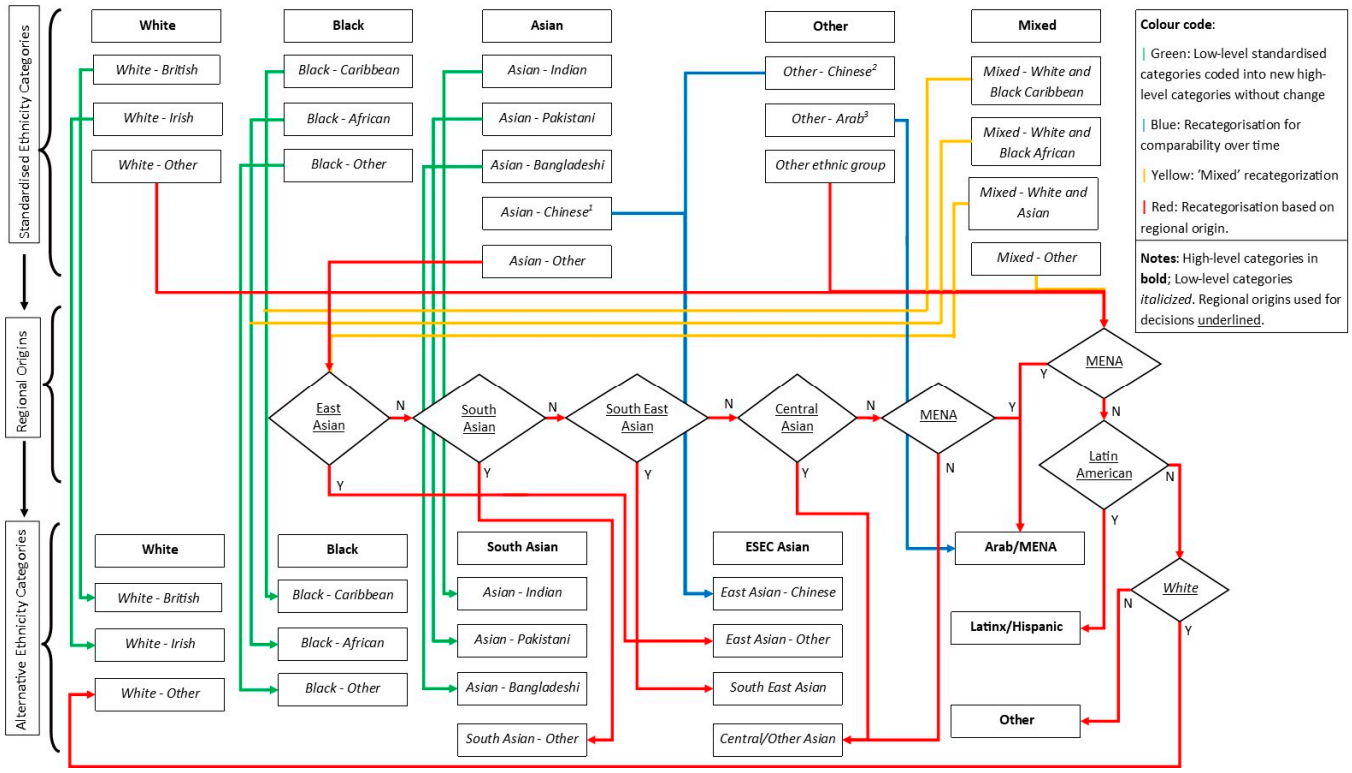

Figure 1. Illustration of recoding decisions for the alternative ethnicity measure, based on standardised categories and regional origin. Notes: [1] (next to Asian–Chinese): "Post 2012/13 data (due to recatgorisation of 'Chinese' respondents from the high-level 'Chinese/Other' category to the high-level 'Asian' category from 2012/13 onwards)". [2] (next to Other–Chinese): "Pre 2012/13 data (due to recatgorisation of 'Chinese' respondents from the high-level 'Chinese/Other' category to the high-level 'Asian' category from 2012/13 onwards)". [3] "Post 2012/13 data (due to the inclusion of 'Arab' as a low-level category within the high-level category of 'Other' from 2012/13 onwards)".

Table 2. Cross-tabulation of unadjusted percentages for respondent ethnicity by regional origin, pooled data (waves 2004–2007, 2013–2019).

	Regional Origin (Country of Origin Grouped by Geographic Region)										
	UK	**Europe**	**North America**	**Caribbean**	**Latin America**	**Africa**	**MENA Region**	**South Asia**	**ESEC Asian**	**Oceana/Other**	**Total**
Respondent Ethnicity (ONS 5 groups)											
White	96.98	95.79	91.15	4.27	53.55	24.97	41.94	4.75	16.63	75.51	91.83
Mixed	0.66	0.82	2.24	5.89	11.64	3.52	5.38	1.06	2.73	1.37	0.83
Asian	1.29	0.39	0.69	0.66	2.55	17.34	18.54	92.90	34.69	6.25	4.04
Black	0.81	0.81	2.00	88.07	9.56	51.50	5.46	**	**	10.96	2.25 **
Chinese or Other	0.27	2.19	3.93	1.12	22.71	2.68	28.69	1.29	45.95	5.92	1.05
Total	100	100 **	100 **	100 **	100 **	100 **	100 **	100 **	100 **	100 **	100 **
Respondent Ethnicity (ONS 16 groups)											
White—British	95.56	18.72	34.91	3.61	10.79	15.11	14.32	4.44	15.00	35.78	86.32
White—Irish	0.27	14.07	1.50			0.23				0.77	0.87 *
White—Other White Background	1.14	63.01	54.74	0.66	42.76	9.62	27.61	0.32	1.63	38.96	4.64
Mixed—White and Black Caribbean	0.28	0.11		4.27	1.99						0.28 *
Mixed—White and Black African	0.06	0.25				2.15	1.29				0.12 *
Mixed—White and Asian	0.17	0.12	0.94		1.14	0.26	2.11	0.89	1.97		0.22 *
Mixed—Any Other Mixed Background	0.14	0.34	1.31	1.63	8.51	1.11	1.98	0.17	0.76	1.37	0.22
Asian or Asian British—Indian	0.61	0.07			1.32	13.79	1.46	42.77	1.97	4.16	1.83 *
Asian or Asian British—Pakistani	0.47	0.17				0.62	0.82	27.44			1.06 *
Asian or Asian British—Bangladeshi	0.12							11.41			0.37 *
Asian or Asian British—Other Asian Background	0.09	0.14	0.69	0.66	1.23	2.93	16.26	11.28	32.72	2.08	0.78
Black or Black British—Caribbean	0.50	0.09	0.75	76.38	5.49	1.13				4.66	0.86 *
Black or Black British—African	0.25	0.57		9.14	1.23	48.44	4.43			5.70	1.26 *
Black or Black British—Other Black Background	0.06	0.15	1.25	2.54	2.84	1.92	1.03			0.60	0.13 *
Chinese	0.08	0.06	0.69			0.20			36.97	0.60	0.39 *
Other Ethnic Group	0.19	2.13	3.24	1.12	22.71	2.49	28.69	1.29	8.98	5.32	0.66 *
Total	100	100 *	100 *	100 *	100 *	100 *	100 *	100 *	100 *	100 *	100 *

Notes: empty cells denote supressed percentages based on cell counts below 10 to adhere to the Statistical Disclosure Control (SDC) requirements. * = total percentages derived from adjusted N based on the sum of all non-suppressed values. ** = the top part of the table (5 ethnicities and regional grouping) is based on the bottom part of the table (16 ethnicities and regional regrouping), thus excluding suppressed values in the bottom part of the table to adhere to the SDC.

4.2. Violence Victimisation and Fear

Physical and sexual interpersonal violence victimisation was based on ONS-defined offence codes for violent offences (serious wounding, other wounding, common assault, attempted assault, robbery, attempted robbery, snatch theft from the person), plus sexual violence offence codes (rape, serious wounding with sexual motive, other wounding with sexual motive, attempted rape, indecent assault) using data from respondent victim forms recording crime victimisation (see Appendix B). Fear of violence was measured using the survey question "How worried are you about being physically attacked by strangers?", whereby respondents reporting being 'fairly' or 'very' worried were coded as fearful (coded 1), compared to those who were 'not very' or 'not at all' worried (coded 0). As this base question was given to approximately half of respondents, we used a reduced sample (missing N = 244,814) in our analyses on the fear of violence. Both violence experience measures (victimisation and fear) were binary variables. All regression analyses controlled for respondent age, occupational class, migrant status, and sex (see Appendix F for uncontrolled regression results). We used ONS-provided population weights, recoding only for comparability over time. See Table 1 for all variables' summary statistics (frequencies and un/weighted percentages).

4.3. Analyses

We first performed unadjusted cross-tabulations to compare standardised ethnicity (5 and 16 ethnic groups, respectively) and the new regionally grouped country of origin. We identified which regional origins were over-represented as 'Other' ethnicities and assessed the homogeneity of each region by ethnicity (Section 5). Since UK-born respondents did not provide additional indicators of ethnicity, we primarily focused on migrant respondents in our interpretation of findings as the basis for later recategorisation.[4] Second, drawing on existing literature, our evaluation of the CSEW and census standardised approaches (Section 3.2), and comparative analysis of ethnicity and regional origin, we proposed an alternative categorisation of ethnicity, with justifications for decisions on specific ethnic groups (Section 5). Finally, we used cross-tabulations and a series of controlled logistic regression models to compare the effect of standardised and alternative ethnicity measures on estimates for violence outcomes (victimisation and fear). To capture significant differences between multiple ethnic groups within a single measure, we used rotating referent groups in our models. In regression analyses with alternative ethnicity categorisation, we additionally controlled for 'Mixed' ethnicity (binary variable indicating single or mixed/multiple ethnicity). In order to additionally test our findings based on like-for-like comparisons, we also ran our regression analyses using a migrant-only sample (see Appendix E).

5. Recategorising Ethnicity

Using cross-tabulations of standardised high (5 groups) and low-level (16 groups) ethnicity and regionally grouped country of origin (Table 2), we identified which regional origins had higher proportions of 'Other' ethnicities and assessed the ethnic homogeneity of each region (see Del Pinal et al. 2007 for a comparable approach used in the US Census).

Based on our ethnicity-regional origin cross-tabulations (Table 2), we found that respondents born in the UK (non-migrants), Europe, North America, the Caribbean, and South Asian regions were largely ethnically homogeneous (largest ethnic group >85% of the total regional population), with low rates of 'Other' ethnic identification (<5%). UK, Europe, and North America-originating respondents were predominantly 'White' (97.0%, 95.8%, and 91.2%, respectively), Caribbean-origin respondents were 88.1% 'Black' (76.4% 'Black-Caribbean'), and South Asian-origin respondents were 92.9% 'Asian' (81.6% identifying with a named South Asian sub-group: India/Pakistan/Bangladesh). Respondents from these regions had low rates of 'Other' ethnic identification (between 0.2 and 3.2%), indicating a high degree of alignment with the ethnic categories provided in the survey. However, we found higher rates of 'Other' ethnicities among respondents from East, South East, and Central (ESEC) Asia (9.0%), the Middle East and North Africa (MENA, 28.7%), and

Latin America (22.7%). We also found high levels of ethnic heterogeneity among MENA (41.9% 'White', '28.7% 'Other', 16.3% 'Asian-Other') and Latin America-originating respondents (42.8% 'White-Other', 22.7% 'Other', 10.8% 'White-British', 8.5% 'Mixed-Other'), as well as African-origin respondents, with 51.5% identifying as 'Black' (including 48.4% Black-African), 25.0% as 'White', and 17.3% as 'Asian'.[5]

Based on these findings, and also drawing upon our evaluation of the CSEW and census approach to ethnicity (Section 3), we propose an alternative measure and categorisation of ethnicity which distinguishes between South Asian and ESEC Asian groups; recategorises 'Mixed' ethnicities; distinguishes 'Arab' groups; and introduces 'Latinx/Hispanic' as a distinct ethnic group. Figure 1 illustrates the recoding process for our alternative ethnicity categorisation, using standardised low-level ethnic groups and regional origins to recategorise and (in cases where ethnicity data are missing for recategorisation) estimate respondent ethnicities for this alternative measure.

5.1. 'Asian'

The broad categorisation of 'Asian' fails to account for variation in both asserted and assigned ethnicity. Individuals subject to discrimination based on their assigned 'Asian' ethnicity typically face different ethnic stereotypes and forms of discrimination based on their perceived affinity to 'East Asia' or 'South Asia', shaped by legacies of colonial and racial systems of inequalities in British history (e.g., Brooks 2019; Oh 2021). As Nerenz et al. (2009) note, "the Asian category blurs ancestry [i.e., ethnicity] distinctions and vast cultural and geographic diversity", potentially alienating respondents from various Asian-origin ethnicities (see also Holup et al. 2007; Laws and Heckscher 2002). Whilst ethnic categories often align with national demographics (i.e., sub-group specificity depending on Asian population sizes), it is notable that the 'Asian' ethnic group encompasses over half the global population, involving substantial internal ethnic variation compared to other standardised ethnic groups and regions. Demographic and migratory histories also influence perceptions of what 'Asian' means, for example, as more associated with South Asians in the UK and East Asians in the US.

The ONS category of 'Asian' was introduced following migration from former British colonies, particularly India, Pakistan, and Bangladesh, to enable government tracking of demographic changes. The specification of 'Asian' ethnicities as 'Indian', 'Pakistani', 'Bangladeshi', or (from 2001) 'Asian-Other', and the pre-2012/13 categorisation of 'Chinese' as 'Chinese/Other' rather than 'Asian', supports the argument that the category 'Asian' was designed to primarily capture South Asian rather than all Asian ethnicities. This is also reflected in respondent ethnic identification (Table 2), with those from ESEC Asian countries having the third highest rate of 'Other' identification (9.0%).

To address the variation in the character of discrimination, and the relatively high rates of misrepresentation as 'Other', we propose differentiating between (1) South Asian and (2) ESEC Asian groups. We coded respondents as 'South Asian' if they identified as a specified South Asian ethnicity (Indian/Pakistani/Bangladeshi), or identified as 'Asian-Other' and originated from South Asia. The 'ESEC Asian' group includes respondents identifying as 'Chinese' (any regional origin), or as 'Asian-Other' or 'Other' (if originating from ESEC Asian countries). Whilst further differentiation between ESEC Asian groups would be preferable, the proposed approach seeks to balance the need for sufficient subsample sizes for high-level group analysis.

5.2. 'Latinx/Hispanic'

Our analysis of ethnic variation by regional origin also revealed a high proportion of respondents from Latin American regions identifying as 'Other', corresponding with the absence of 'Latinx/Hispanic' as an ethnic group. This aligns with similar findings in the US, where 47.9% of Hispanic respondents were classified as 'White' and 42.2% as 'Some other race', with most Hispanic respondents classified as the latter also giving 'Hispanic' as their race (Del Pinal et al. 2007). This further supports the inclusion of 'Latinx/Hispanic' as an

ethnic group, since ethnicity is more than simply 'race', yet racial groups do not adequately capture Hispanic respondents without additionally specifying them as a group. It also demonstrates how perceptions of ethnicity, origin, and race are perceived as overlapping, whereby 'Hispanic origin' is prioritised over pre-designated racial groups (ibid). The introduction of 'Latinx/Hispanic' as an ethnic group in the UK has also been supported by civil society efforts (UK Parliament Petition 2021; Coalition of Latin Americans in the UK 2023).

We therefore propose the inclusion of 'Latinx/Hispanic' as a defined ethnicity group in data collection. Due to the absence of data collected on 'Latinx/Hispanic' as an ethnicity category, we coded a proxy measure for this group using Latin American migrants who identified as 'Other' or 'White-Other'. Whilst this approach fails to capture UK-born Latinx/Hispanic respondents, including second-generation migrants, it provides the closest estimate of 'Latinx/Hispanic' respondents based on the available data.

5.3. Arab and Middle East and North African (Arab/MENA)

MENA-originating respondents had the highest proportion of 'Other' ethnicities (28.7%). This finding supports the 2012/13 introduction of 'Arab', although noting that as this group does not 'fit' into any high-level category and represents a small proportion of the population, it remains subsumed with 'Other' in the high-level categorisation of ethnicity. In order to capture 'Arab' respondents during waves pre-dating the 2012/13 introduction of this category (required due to sample size limitations), we similarly coded MENA-origin migrants otherwise categorised as 'Other' (or also 'White-Other' and 'Asian-Other' in pre-2012/13 data) as 'Arab/MENA'. Whilst this approach encounters similar limitations to 'Latinx/Hispanic', with only post-2012 'Arab' respondents being accurately captured and pre-2012/13 data only capturing 'Arab' as an approximate migrant-only group, this enabled us to roughly distinguish 'Arab' respondents from 'Other' ethnicities for the purpose of analysing ethnic variations in violence outcomes. This proxy approach draws on previous US Census methods for reducing the number of 'Some Other Race' by recategorising them based on other available ethnicity-related data (Del Pinal et al. 2007).

5.4. 'Mixed' Ethnicities

A particularly contentious challenge we sought to address was the misrepresentation of respondents with multiple ethnicities, which has seen increased interest in the 21st century as one of the fastest growing (Song 2012) and most under- and misrepresented (Del Pinal et al. 2007; see also Goldstein and Morning 2000; Song 2012) ethnic minoritised populations.

As noted in our critical evaluation of the CSEW's methodology (Section 3), the categorisation of 'Mixed' ethnic respondents into a singular category alongside defined ethnic groups conflates two distinct dimensions of ethnicity: ethnic multiplicity/heterogeneity and ethnic group membership. Whilst other ethnic categories also conflate characteristics of ethnicity—race, nationality, regional origin—these collectively make up the bases of ethnic groups, that is, the characteristics that define a shared identity or community. The lack of conceptual coherence in using a single measure for both the multiplicity and group membership of ethnicity undermines the comparison between ethnic categories. The 'Mixed' category is measuring something entirely different to other categories within the same measure, as well as to the measure itself.

Unifying 'Mixed' ethnic respondents into a singular 'ethnic group' either produces a second 'miscellaneous' category devoid of meaningful information on ethnic group experiences, or assumes that 'Mixed' individuals share equivalent commonalities (histories, identities, belongings) to other ethnic groups. For the latter, both academic studies and public discourse (Song 2010) widely dispute the notion that "there is 'a' mixed race group" (Song 2010; Woozeer 2022). Whilst 'Mixed' ethnic groups may share some similar experiences (rejection by both ethnic groups, ambiguity in identity, intersecting oppressions), this does not represent a coherent ethnic identity or group to which they belong (Song 2010).

Further, by recategorising 'Mixed' ethnic respondents into a singular group, this approach erases the very characteristic which this variable seeks to measure: ethnic groups.

Whilst the low-level ethnic groups benefit from some distinction between combinations of ethnicities (e.g., 'White and Asian'), unlike most other countries' approaches (Morning 2015), they are recategorised into a singular 'Mixed' group at the high level, which is almost exclusively used in research. Given the importance of representation of ethnic group membership for respondents and of meaningful data on respondents' ethnic groups for researchers, the removal of this information means that 'Mixed' ethnic groups are less represented in the data than 'single' ethnic groups, creating inequalities within the data themselves.

In attempting to address these issues of conceptual incoherence, conflation, and erasure of ethnic group identity, and misplacing 'Mixed' ethnicities within systems of ethnic inequalities, we propose a recategorisation of 'Mixed' ethnicities which offers an alternative or complementary approach to representing both the multiplicity and specificity of ethnic group membership. This approach first regroups 'Mixed' ethnic respondents into ethnic groups on the basis of minoritised ethnic group membership, whereby each minoritised ethnic group (South Asian, ESEC Asian, Black, Arab/MENA, Latinx/Hispanic, Other) includes both single- and multiple-ethnic respondents. It then codes a separate variable for ethnic multiplicity which distinguishes between 'single' and 'mixed/multiple' ethnic respondents, allowing researchers to distinguish between, and explore the interacting effects of, the role of ethnic groups and heterogeneity/multiplicity. Similar 'prioritisation' approaches to categorising multiple ethnic groups were implemented in several countries around the 1990s, such as New Zealand and the US (Cormack and Robson 2010; Nerenz et al. 2009; Office for Management and Budget 2000). This approach typically prioritises minoritised ethnicities when recategorising mixed ethnic groups into single ethnicities, particularly in relation to civil rights claims and research on discrimination and inequalities (ibid).

This approach was largely discontinued following concerns regarding (1) prioritising some minoritised ethnic groups above others, resulting in lower counts for deprioritised groups (e.g., Pacific Islanders in New Zealand, see Cormack and Robson 2010); (2) applying a prioritisation which may not match the individual's own prioritisation of their ethnicity (Cormack and Robson 2010); (3) neglecting the significance of 'mixedness' as a core part of mixed ethnicity groups' experiences and identities; and (4) reinforcing racial boundaries which no longer align with a diversifying populace (Holup et al. 2007; Nerenz et al. 2009; Payson 1996; Valles et al. 2015; Didham and Callister 2012; Cormack and Robson 2010). Whilst this approach has largely been discontinued in national survey data (Cormack and Robson 2010), various studies continue to implement it as a means of increasing ethnic group sample sizes without excluding 'Mixed' ethnicities and as indicators of respondent's position within systems of racial inequalities (e.g., Greil et al. 2011).

A key challenge in this proposed approach is that it involves a prioritisation of respondent's minoritised ethnic group membership over, in many but not all cases, their white ethnic identity. It may therefore be contested that in attempting to resolve the erasure of any ethnic group identity, it incorrectly privileges one identity over the other and ignores the importance of ethnic heterogeneity among individuals. It also risks repeating and at worst legitimising racial logics of 'whiteness' as the default and 'pure' ethnic group, such as the 'one drop' rule used to discriminate against any affiliation with 'non-whiteness' regardless of the individual's asserted ethnicity or sense of group membership (Khanna 2010). However, in defence of this proposal, it may be countered that ethnic measures already apply this logic by categorising all 'Mixed' ethnicities as Black and minoritised ethnic or 'non-white' in binary measures of ethnicity which are commonly used in analysing differences in outcomes by ethnicity and is derived from the standardised five-group ethnic measure in most UK survey datasets. This existing binary measure not only groups mixed white and minoritised ethnic respondents as 'non-white', but also incorrectly groups mixed white respondents (i.e., respondents who identified as 'Mixed—Other' with both ethnicities being as white, such as British-Finnish or Irish-Croatian) separately from 'White-Other'.

By distinguishing separate variables for 'single/mixed' and 'ethnic' group, we are able to examine the effect of multiple ethnicity without erasing important information on ethnic group membership, as well as allowing for analysis of the interaction between them, rather than relying on an incomplete and inconsistent measure for both simultaneously. Rather than a prescribed replacement of the standardised approach, we use this alternative approach as a means of testing the implications of different categorisations of ethnicity, with the aim of opening discussion on how to address the complexities and inconsistencies involved in wrestling such a multi-dimensional and subjective concept into standardised measures which are statistically meaningful for researchers.

5.5. Recoding Ethnicity

In order to explore the effects of different ethnicity measures on ethnic disparities in violence experiences, we recoded the standardised ethnicity variables based on the above proposals to derive a proxy alternative categorisation for ethnicity. The alternative categorisation sought to improve upon the current standardised approach by (1) differentiating 'Mixed' ethnicity as an ethnic characteristic, rather than a group, and integrating mixed-ethnic respondents into specific ethnic groups; (2) differentiating between 'South Asian' and 'ESEC Asian' ethnic groups; and (3) introducing the ethnic group 'Latinx/Hispanic' and distinguishing 'Arab' from 'Other' ethnicities.

Our approach to recoding 'Other' ethnicities based on regional origins data draws on comparable methods to recategorise racial groups in the 2000 US Census (Del Pinal et al. 2007). The census applied the "90 Percent Rule" to recategorise responses indicating ethnicity into a standardised racial group. By cross-tabulating ancestry (similar to ethnic groups in the CSEW) and race responses, respondents who provided 'ethnic' (not racial) groups were recategorised into a racial group if 90% of respondents from the same ancestry selected this racial group (e.g., "German" was coded as 'White', "Jamaican" as 'Black/African American'). Ancestry groups with less than 90% alignment with single racial groups were designated "Some other race" (e.g., 'Mexican', 'Guyanan') (Del Pinal et al. 2007).

Our approach similarly seeks to reduce the number of respondents categorised as 'Mixed' or 'Other' which provides little useful information for researchers seeking to understand variation by ethnic group, as well as to provide a proxy measure for ethnicity which accounts for Arab/MENA and Latinx/Hispanic ethnicities and differentiates between Asian ethnicities as an indication of potential future alternative approaches. As this approach relies on regional origin for the recategorisation of some groups (Arab, Latinx, some Asian sub-groups), and therefore is only able to use data from migrant respondents, it offers an approximate recategorisation for the purpose of demonstrating variation in violence outcomes between standardised measures and a proxy alternative. It serves as a tool to investigate the implications of different ways of measuring violence and propose recommendations for methodological improvements. As such, it uses a mix of asserted ethnicity by respondents and quasi-assigned ethnicity where asserted ethnicity and regional origin are used to form newly assigned categories. These quasi-assigned categories aim to provide an indication of the need for changes which would improve the accuracy and representativeness of asserted ethnicity. The process of estimating otherwise excluded ethnic groups faces limitations in terms of the data it relies upon (i.e., 'Latinx/Hispanic' and 'Arab' groups made up of migrants only) and is designed to provide a proxy estimate preceding methodological improvements which may implement more detailed, inclusive, and representative measures and categories for ethnicity.

6. Implications for Violence Victimisation and Fear

In analysing the implications of different ethnicity measures on inequality patterns in violence-related outcomes, we ran cross-tabulations (Table 3, showing frequencies and un/weighted percentages) and a series of logistic regression models (Table 4, see Appendix D for detailed regression tables) for violence victimisation (past 12 months) and fear

of stranger violence by standardised and alternative categorisations of ethnicity. We also ran supplementary analyses of migrant-only samples (Appendix E, Table A2) to check for consistency in our results based on like-for-like comparisons (although noting a smaller sample size), as well as uncontrolled analyses (Appendix F, Table A3), which supported our main results and conclusions.

Table 3. Cross-tabulations indicating the proportion of respondents who experienced past-year violence victimisation and fear of stranger violence by standardised and regrouped ethnicity, pooled data (waves 2004–2007 and 2013–2019).

	Violence Victimisation (Past Year), Sample 1 (N = 415,361)			Fear of Stranger Violence (Fairly/Very Worried), Sample 2 (N = 169,440)		
	N (Violence Victims)	*Un-Wgt %*	*Wgt %*	*N (Fearing Violence)*	*Un-Wgt %*	*Wgt %*
Standardised Ethnicity (5 groups)						
White	9852	2.52	2.93	67,576	29.69	30.81
Mixed	180	5.02	5.11	628	37.16	37.75
Asian	365	2.12	2.27	3877	47.54	48.53
Black	279	2.90	3.00	2194	46.69	46.29
Chinese or Other	99	2.20	2.44	1099	40.95	43.14
Total	10,775	2.53	2.91	75,374	30.79	32.41
Ethnic regrouping (7 groups)						
White	9838	2.52	2.93	67,495	29.68	30.78
Black/Black British	369	3.26	3.31	2514	45.66	45.64
Arab/MENA	21	4.46	4.38	214	45.53	48.70
South Asian	318	2.09	2.27	3438	47.26	48.47
ESEC Asian	116	2.33	2.46	1029	44.22	44.44
Latinx/Hispanic	27	3.45	3.80	148	44.85	45.35
Other	86	3.31	3.49	536	35.54	37.15
Total	10,775	2.53	2.91	75,374	30.79	32.41
Mixed Ethnicity						
Not Mixed (Single Ethnicity)	10,595	2.51	2.88	74,746	30.74	32.36
Mixed/Multiple Ethnicities	180	5.02	5.11	628	37.16	37.75
Total	10,775	2.53	2.91	75,374	30.79	32.41

Table 4. Odds ratios from logistic regressions of violence victimisation and fear of violence using standardised ethnicity and regrouped ethnicity as predictors, pooled data (waves 2004–2007, 2013–2019).

	Violence Victimisation	Fear of Violence
	OR	**OR**
Standardised Ethnicity (5 groups)		
1. White	*Ref*	*Ref*
2. Mixed	1.192 †	1.253 ***
3. Asian/Asian British	0.634 ***	1.967 ***
4. Black/Black British	0.925	1.635 ***
5. Chinese/Other	0.674 **	1.450 ***
Ethnic regrouping (7 groups)		
1. White	*Ref*	*Ref*
2. Black/Black British	0.884	1.670 ***
3. Arab/MENA	1.120	1.747 ***
4. South Asian	0.631 ***	1.974 ***
5. ESEC Asian	0.609 ***	1.736 ***
6. Latinx/Hispanic	1.200	1.508 **
7. Other	0.826	1.275 ***
Mixed Ethnicity (Ref = Not mixed)		
Mixed/Multiple Ethnicities	1.492 ***	0.783 ***

Notes: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

6.1. Physical Violence Victimisation

Standardised ethnic group comparisons (by cross-tabulated weighted percentages, Table 3) for violence victimisation showed that 'Mixed' ethnicity respondents were nearly twice as likely to have experienced violence in the past year (5.1%) than other ethnic groups. Black respondents had the second highest prevalence of violence victimisation (3.0%), followed by White (2.9%) and 'Other' (2.4%) respondents, with the lowest prevalence reported among Asian respondents (2.3%). Using our alternative categorisation of ethnicity, which introduced 'Latinx/Hispanic' as a distinctive ethnic group and which distinguished 'Arab/MENA' from 'Other', we found the highest victimisation rates among 'Arab/MENA' (4.4%) and 'Latinx/Hispanic' (3.8%) ethnic groups, followed by 'Other' (3.5%), 'Black' (3.3%), and 'White' (2.9%) groups (Table 3). 'ESEC Asian' and 'South Asian' groups reported the lowest prevalence of violence victimisation (2.5% and 2.3%, respectively).

Based on our logistic regression analyses (Table 4, see also Appendix D), we found that these results varied in significance, with our overall prevalence findings largely supported, although with some challenges noted regarding significance among smaller ethnic groups when using the alternative categorisation. Using the standardised categories, we found that 'Mixed' ethnicity significantly increased the odds of victimisation compared to all other ethnic groups, although only marginally ($p < 0.1$) compared to 'White' groups. The odds of victimisation among 'White' and 'Black' groups were significantly higher than 'Asian' and 'Other' groups, although we found no significant difference between 'White' and 'Black' ethnic groups (see Appendix D for full results).

Regression analyses using our alternative categorisation similarly found that 'Mixed' groups had significantly higher odds and prevalence of violence victimisation than non-mixed (OR 1.492 $p < 0.000$). 'ESEC Asian' and 'South Asian' ethnicities both had significantly lower odds of victimisation than all other ethnic groups.[6] Similarly to regression results using standardised measures, we found no significant difference between 'White' and 'Black' groups using our alternative categorisation. Furthermore, whilst we had previously found a higher prevalence (Table 3) among 'Arab/MENA' and 'Latinx/Hispanic' groups compared to 'White', these differences were not statistically significant, most likely due to the relatively small size of these groups. Victimisation odds were significantly higher for 'Arab/MENA' and 'Latinx/Hispanic' groups only when compared to 'ESEC Asian' and 'South Asian' respondents.

Whilst our regression results found few notable differences in violence victimisation by standardised and alternative categorisations, variations by prevalence provide several insights relating to the introduction, distinction, and recategorisation of ethnic groups. First, by distinguishing 'Arab/MENA' and 'Latinx/Hispanic' respondents from 'White' and 'Other' categories, we found that these groups experience a higher prevalence of violence victimisation than other ethnic groups; notably, there were higher prevalence rates among Latinx/Hispanic (3.80%) and Arab/MENA (4.38%) respondents than 'White' (2.93%) or 'Chinese/Other' (2.44%) respondents, which, based on the ONS standardised measure, they would otherwise be subsumed and hidden within. These prevalence rates (a measure commonly used in national statistical reports) indicate the importance of measuring and disaggregating these ethnic groups.

Second, we found that 'Mixed' ethnic groups experienced the highest prevalence and odds of violence victimisation in both standardised and alternative categorisations (although only marginally significantly ($p < 0.1$) higher than 'White' respondents using standardised categories). We posit that our alternative approach to the multiplicity of ethnicity enables researchers to better represent the different risk factors posed by (1) being 'Mixed' and (2) being part of particular ethnic groups, without obscuring the heighted risk of violence against 'Mixed' ethnic groups. Distinguishing between these different dimensions for ethnicity enables future research to account for their interaction and distinctiveness. Our finding that 'Mixed' ethnic groups are of particular risk of violence demonstrates the need for further research into the distinctive attribute of ethnic multiplicity in research on ethnic disparities in violence.

Third, distinguishing between 'ESEC' and 'South Asian' groups did not reveal any significant differences between these groups, with results equivalent to the original 'Asian' categorisation. Overall, we found that our recategorisation of ethnicity helped to reveal high levels of violence prevalence among otherwise under/mis-represented groups, but that this approach also faced challenges in terms of statistical power and significance necessary for reliable results.

6.2. Fear of Physical Stranger Violence

In analysing the fear of stranger violence by standardised ethnicity (Table 3), we found that nearly half of respondents from 'Black', 'Asian', or 'Other' ethnic groups reported being 'very' or 'fairly' worried about being physically attacked by a stranger, compared to under a third of White respondents. 'Asian' respondents reported the highest (weighted) prevalence of violence fears (48.53%), followed by 'Black' (46.29%), 'Other' (43.14%), and 'Mixed' (37.75%) groups, and 'White' respondents reported the lowest rate of violence fears (30.81%). Our regression models found that these results were all statistically significant ($p < 0.001$), although only marginally ($p < 0.1$) for 'Other' ethnic respondents when compared to 'Mixed' and 'Black' groups (OR 1.158, OR 0.887, respectively).

Using our alternative categorisation of ethnicity (proxy measure), we found that 'South Asian' (47.3%) respondents had the highest percentage of being 'very/fairly' worried about stranger violence, followed by 'Black' (45.8%), 'Latinx/Hispanic' (44.9%), 'ESEC Asian' (44.3%), 'Arab/MENA' (43.5%), and 'Other' (32.5%), with the lowest rate of violence fear among for white respondents (29.7%). The results of our logistic regression models estimated that 'South Asian' respondents had the highest odds of reporting violence fears (significantly higher than all ethnic groups except non-significantly higher than 'Latinx/Hispanic'), and that white respondents had the lowest (significance in all subgroup comparisons). 'Arab/MENA' respondents had the second highest odds of violence fear, although this was not significantly different from the 'ESEC Asian', 'Black', and 'Latinx/Hispanic' groups.

Notably, by distinguishing between 'South Asian' and 'ESEC Asian' groups, we found that 'ESEC Asians' experienced significantly lower odds of violence fear, which were more similar to Black respondents than South Asians. In recalling that MENA and Latin American-originating respondents would largely otherwise be categorised into either 'White' or 'Chinese/Other' ethnic groups (Table 2 and Figure 1), it is notable that when distinguishing these groups, both Arab/MENA and Latinx/Hispanic respondents had significantly higher odds of violence fear than both 'White' and 'Other' respondents.

7. Discussion

This study identified several methodological issues in the way the CSEW, and by extension most of the UK's national surveys and government statistics, measures and categorises ethnicity. These include excluding and conflating certain ethnic groups (e.g., 'Latinx/Hispanic', 'South Asian', and 'ESEC Asian'); conceptual incoherence of categories and conflation of multiple dimensions of ethnicity; accessibility and useability of the data by researchers; and longevity of ethnicity categories in the face of an evolving demographic landscape and research agenda (Sections 3 and 5). In particular, the categorisation of 'Mixed', 'Asian', and 'Other' ethnicities, the exclusion of 'Latinx/Hispanic' as an ethnic group, and issues surrounding the comparability of 'Chinese' and 'Arab/MENA' ethnicity data were all found to undermine respondent's self-recognition in categories, conceptual coherency, and practical usage of ethnicity data from the perspective of both respondents and researchers.

Our comparison of standardised ethnic measures and regional origins found high levels of Latin America and MENA-originating respondents being categorised as 'Other' which indicates a lack of appropriate categories which reflect respondents' self-identified ethnicity, particularly among 'Latinx/Hispanic groups'. It should be noted that although the latter is partially addressed by the 2012/13 introduction of 'Arab' as a response option,

these data are inaccessible to most researchers (current confined to the secure-access dataset) and the conflation of 'Arab' with 'Other' ethnicities means that Arab respondents remain marginalised within the data. The ethnic identity of Latinx/Hispanic respondents is entirely excluded from current measures of ethnicity, preventing analyses of the risks, adverse experiences, and inequalities faced by this particular group. Further, whilst South Asian-origin respondents were comparatively well represented by the group 'Asian', respondents from ESEC Asian countries had higher rates of 'Other' categorisation and there remains a lack of comparable data which distinguish 'Chinese' from the 'Chinese/Other' category which arbitrarily segregates this group from other ESEC Asian ethnicities.

This omission is not without consequence. In comparing the effects of standardised measures and our proposed alternative ethnicity categorisation as predictors of violence victimisation and fear of stranger violence, we found notable variations in ethnic patterns based on the distinction between South and ESEC Asians, the recategorisation of 'Mixed' ethnicities, and the extraction of 'Arab' and 'Latinx' ethnic groups from 'Other' and 'White' ethnicities.

Based on our regression analyses on ethnic disparities in the fear of violence, our finding that South Asian respondents experienced significantly higher levels of fear of violence than ESEC Asian respondents further supports the need for disaggregating Asian ethnicities. Our regression analyses on violence victimisation revealed that 'Mixed' ethnicities reported significantly higher odds of violence than all other groups, demonstrating the importance of research and policy which addresses ethnic multiplicity ('mixed-ness') as a distinctive and heightened risk factor, more so than specific ethnic group membership. It also suggests the intersecting risk-factors posed by both ethnic group membership and 'Mixed' ethnicities for violence victimisation, which would benefit from additional policy and research attention.

Further, our comparative analyses of prevalence rates (Table 3) found that Latinx/Hispanic and Arab/MENA respondents both had a higher prevalence of violence victimisation and fear than 'White' or 'Other' ethnicities, which they would otherwise be subsumed within. By comparing alternative approaches to ethnic categorisation to the study of ethnic disparities in violence victimisation and fear, this paper demonstrates the impact of measurement on our understanding of inequalities. Given that the standardised measurement of ethnicity is prevalent across most UK national surveys and government statistics, this paper provides a significant insight into the potential implications of ethnic categorisation across a much broader range of policy and research areas, which may be expanded through future research on this subject.

7.1. Limitations and Strengths

These findings are subject to some limitations relating to sample size restrictions, missing variables, and restrictions in ethnicity recoding based on available data. However, these limitations are also reflective of the limitations in the data and measures themselves which this study highlights.

Whilst the CSEW has a comparatively large sample size, with approximately 35,000 respondents per year, quantitative analyses of ethnic variation require higher numbers of minoritised ethnic participants in order to achieve sufficient statistical power when working with small sub-populations on the basis of minoritised ethnic groups, which in total only represent 8.5% of the sample population. This issue is exacerbated when analysing low-prevalence outcomes, such as violence victimisation (2.53%). Variation in statistical significance based on sub-population sizes (i.e., victims of violence from low-population ethnic groups) seemed to restrict several findings (e.g., violence against Arab and Latinx populations). The discontinuation of the CSEW's ethnic boost sample is therefore an important issue to address in order to alleviate the issue of insufficient ethnic minoritised sample sizes for analysis of ethnic variation, not least in relation to violence victimisation.

This study required access to the CSEW secure data files due to the removal of detailed ethnicity variables in 2012/13, as well as the lack of sufficient regional origin data in the main dataset. As previously noted, the removal of data on low-level ethnic groups restricts

researchers' ability to specify ethnic groups and instead relies on the high-level categories which this study has highlighted several issues with. Several years of detailed ethnicity data were also missing from both the non-secure and secure datasets, resulting in the combined data for this study being split across two time periods (2004–2007 and 2013–2019).

It should also be noted that whilst additional marginalised groups may have been considered, such as 'Eastern European', including these may further exacerbate issues of insufficient sample sizes, without improving representation in the high-level ethnic categorisation. By prioritising groups which do not align with any high-level categories and are over-represented as 'Other' (unlike 'Eastern European' which is predominantly white), we have sought to avoid excessively breaking down ethnic groups where misrepresentation is less prevalent based on our comparative analysis (Table 2).

Whilst considering the limitations of this study, several strengths and contributions should also be highlighted. Whilst there has been significant discussion and debate on the conceptualisation and categorisation of ethnicity, there are very few studies which practically engage with this subject through quantitative analysis, particularly regarding the implications of misrepresentation on specific outcomes. We have found no studies which have analysed the implications of ethnic measures and categorisation on study outcomes relating to experiences of violence, despite growing literature on the implications of how we measure violence itself (Pullerits and Phoenix 2023; Walby et al. 2016; Cooper and Obolenskaya 2021). This study contributes both towards existing critiques of standardised ethnicity measures and speculated potential implications, and, most notably, takes this further by practically applying these critiques to a specific dataset and implementing its recommendations through an alternative measure. This approach provides tangible evidence on how the way we measure ethnicity shapes our understanding of inequalities in the experience of violence in marginalised populations.

Further, whilst it may be argued that our results do not significantly change our understanding of binary ethnic inequalities (i.e., 'White' vs. minoritised ethnicities) and that it involves relatively small sub-groups (e.g., 'Arab/MENA' and 'Latinx/Hispanic'), we would argue that representativeness and visibility is crucial here. In prioritising broad categories, ethnicity data risk exacerbating how smaller ethnic groups are deprioritised and rendered invisible in policymaking and service provision. By making otherwise 'hidden' (e.g., 'Latinx/Hispanic'), obscured (e.g., 'ESEC Asian' or 'Arab/MENA'), or misrepresented (e.g., 'Mixed') groups visible in the data, we emphasise the importance of revealing inequalities and countering marginalisation through representation. For example, by introducing 'Latinx/Hispanic' as a distinctive ethnic group, we were able to reveal hidden inequalities and risks of violence among this group which would otherwise have been subsumed within 'White' or 'Other' experiences, which indicated lower levels of fear and victimisation of violence. In seeking to reduce the categorisation of individuals in ways that provide few meaningful insights into disparities across ethnic groups (e.g., 'Other' or 'Mixed'), we have sought to improve how ethnic categories provide representation and meaningful information on the experiences of marginalised groups.

7.2. Methodological Recommendations

In combining insights from the critical evaluation of the CSEW's approach to measuring ethnicity, and the results of our comparative analysis of ethnicity and regional origin, and further supported by variations in violence outcomes when utilising an alternative approach to ethnic categorisation, we propose several methodological adaptations to the CSEW, and more broadly to standardised ethnicity measures in the UK. These recommendations also draw upon insights and recommendations from existing critiques of standardised ethnic categorisations (Kautt 2011; Burton et al. 2010; Nerenz et al. 2009; Song 2010) and alternative approaches in other surveys and studies (Del Pinal et al. 2007; Morning 2015; Cormack and Robson 2010; Nerenz et al. 2009; Office for Management and Budget 2000; Kautt 2011; Oh 2021; Laws and Heckscher 2002). Based on our evaluation and analysis of the CSEW's standardised measure of ethnicity, we recommend the following:

A. **Introducing 'Latinx' as a response option for ethnicity**.

The introduction of 'Latinx/Hispanic' as a distinctive ethnic group is a much-needed addition to existing response options. Currently, Latinx respondents are not represented in ethnicity data, despite being one of the fastest growing ethnic groups in the UK. The absence of representation has been publicly noted (Coalition of Latin Americans in the UK 2023; UK Parliament Petition 2021). Our findings that Latinx respondents experience higher prevalence of violence fears and victimisation than 'White' and 'Other' ethnicities further support the need for their inclusion.

B. **Distinguishing between 'Mixed' as a dimension of inequality** rather than an ethnic group by enabling respondents to select multiple ethnicities if identifying as 'Mixed' and coding separate variables for 'singular/mixed' ethnicity and 'ethnic group'.

This would address the conflation of multiplicity as an ethnic dimension with ethnic groups, as well as improve representativeness for respondents and utility for researchers. Rather than lumping 'Mixed' respondents into a singular ethnic group which erases important information on group membership after the point of ethnic disclosure (Song 2012), this approach enables a more conceptually coherent and representative way of distinguishing the effect of 'Mixed' ethnicity and ethnic group. This is particularly important for violence research due to the estimated effect of 'Mixed' ethnicity in significantly increasing violence prevalence across groups.

C. **Distinguishing between 'South Asian' and 'ESEC Asian'.**

Greater distinction of 'Asian' groups would improve the representation of ESEC Asians who are currently only partially represented by a category which is designed to primarily capture 'South Asian' ethnic groups specifically, which insufficiently captures significant differences across 'Asian' ethnicities (Nerenz et al. 2009; Holup et al. 2007; Laws and Heckscher 2002). Similarly to the case of heightened violence against Arab and MENA-origin groups during the Global War on Terror period, the obscuring of 'ESEC Asian' respondents within the data meant that monitoring inequalities in violence during the COVID-19 period, which saw targeted violence against East Asian populations (Ha et al. 2020; Carr et al. 2022), was limited. Distinction between different Asian ethnic groups would enable better monitoring of discriminatory violence on the basis of ethnic stereotypes or group membership.

D. **Reviewing the conflation of 'Arab' with 'Other' ethnicities.**

Whilst the implications of small sub-sample sizes would need to be considered, the introduction of 'Arab' in 2012/13 is undermined by the inaccessibility of data which enables researchers to distinguish this group from others, and the inclusion of 'Arab' into the 'Other' category both in the data and the questionnaire risks repeating the issues posed by arbitrarily grouping 'Chinese/Other' together in previous waves. Further analysis of the implications of distinguishing 'Arab' from other ethnicities in terms of methodological practicalities is needed.

E. **Introducing 'write-in' options for 'Other' ethnic groups** (similarly to the UK Census).

This would improve the representativeness, utility, and longevity of ethnic categories by enabling respondents to specify their ethnic group where this information would otherwise be lost as 'Other', enabling data providers to recode the data for specific research purposes and access data on low-population groups (via secure-data if necessary for disclosure risks), and enabling the collection of groups which may become more demographically significant in the future without having to disrupt comparability over time in the ethnicity variables.

F. **Enabling greater flexibility, usability, and accessibility of ethnicity data.**

Finally, we propose that the low-level ethnicity data (16 groups) are re-introduced into the main dataset, as well as introducing measures of additional dimensions of ethnicity in the CSEW and any UK surveys lacking these, such as language and parental

heritage (as in other national surveys, see Morning 2015). This would enable researchers to capture the specificity of ethnic groups and the diverse dimensions which make up ethnicity. In cases of sufficient methodological concerns (e.g., disclosure risks, sample sizes), alternative categorisations of ethnicity (e.g., applying our alternative approach as a medium-level categorisation) should be added to existing variations in high-level ethnic variables currently available. We also recommend the re-introduction of the ethnic boost sample and introduction of other measures to improve the representation of minoritised ethnic populations in the CSEW, which would greatly alleviate the challenges relating to statistical power when analysing smaller marginalised groups.

In addition to these specific recommendations, recently developed resources for identifying and mitigating inequalities in the production of data and evidence on marginalised groups, such as Innes et al.'s (2023) 'Risk of Bias' assessment tool, may provide a productive framework for mainstreaming perspectives and practices which improve the representation and visibility of minoritised ethnic groups' experiences, particularly in relation to violence (see, in particular, p3 and 12).

These recommendations draw on key insights from other national measures (e.g., US and New Zealand) and studies of ethnicity, research on the need for balance and consideration of how ethnicity measures are constructed and implemented, and critiques of specific elements in standardised measures of ethnicity (Del Pinal et al. 2007; Ahlmark et al. 2015; Morning 2015; Greil et al. 2011; Song 2012; Coalition of Latin Americans in the UK 2023; Holup et al. 2007). The categorisation of 'Mixed' ethnicity, alternative response approaches (e.g., write-in boxes, list break-downs, hierarchal structures), 'Asian' group distinctions, and the exclusion of certain groups, such as Latinx/Hispanic, have both been highlighted in previous literature and identified as posing limitations and challenges for ethnic representation in the CSEW through our critical evaluation of its methodology. Whilst some challenges require changes to sampling (e.g., re-introducing the ethnic boost sample), the recommendations focus on recategorising ethnicity to address conceptual coherence, representation and inclusion, and revealing hidden violence and victims in the data which would otherwise be excluded or obscured based on current measures.

8. Conclusions

Despite the plethora of discussion and debate surrounding the conceptualisation and categorisation of ethnicity, there is a notable absence of studies which practically engage this subject through quantitative analysis, particularly those which explore the implications of misrepresentation on specific outcomes. No studies have thus far analysed the effect of ethnic misrepresentation on violence outcomes, despite a growing number of studies addressing the effect of flawed violence measures on inequality patterns in victimisation (Pullerits and Phoenix 2023; Walby et al. 2016; Cooper and Obolenskaya 2021). This study has not only contributed towards existing critiques of standardised ethnicity measures and speculated potential implications, but instead taken this further by utilising these critiques in application to a specific dataset and implementing its recommendations through a proxy alternative measure which has been used to more concretely analyse the implications of different measures for understanding inequalities in violence.

In practice, all standardised measures of ethnicity face challenges in fully capturing the heterogeneity, multi-dimensionality, complexity, subjectivity, and fluidity of the concept and its meaning for individuals and societies. Despite these challenges, however, it is essential that such measures do not further contribute towards the omission, exclusion, or marginalisation of certain ethnic groups and that the measurement of ethnicity reflects the conceptual and individual meaning of ethnic groups as accurately as possible. The measurement and categorisation of ethnicity plays a significant role our understanding of ethnic inequalities and the adverse experiences of marginalised ethnic groups.

They not only shape our understandings of systems of ethnic inequalities and how to tackle them through the production of research and evidence for policymakers and service providers, but often signify a source of alienation and misrepresentation among

minoritised ethnic groups. This paper has sought to bridge the gap between conceptual deliberations and practical applications by identifying limitations and misrepresentations in standardised measures of ethnicity, their implications for our understanding of violence and inequalities, and how they may be addressed through specific methodological amendments to categorisation.

We found evidence supporting the need for inclusion of 'Latinx/Hispanic', a group with is explicitly excluded from measurement. Whilst meaningful analysis of this group would require further methodological improvements to address insufficient sample sizes, our findings of disproportionately high levels of violence victimisation among Latinx/Hispanic populations demonstrate the importance of representation in uncovering hidden inequality patterns and victims. In seeking to address particular challenges and inadequacies of the standardised approach to 'Mixed' ethnicity, our alternative approach in turn faces a new set of challenges and limitations. We argue that enabling researchers to determine which categorisation of 'Mixed' best suits their research needs would be a significant improvement, providing an approach which better distinguishes between dimensions of ethnicity relating to heterogeneity and group membership. This paper has sought to account for the challenges in balancing conceptual complexity, individual meaning, and methodological practicalities, and proposed several key recommendations for improving the measurement and categorisation of ethnicity which both accounts for these competing factors whilst prioritising measures which avoid unnecessary exclusions and omissions, reveals hidden inequalities in the data, and better represents minoritised ethnic groups in UK national surveys and government statistics.

Whilst there are risks involved in measuring ethnicity and categorising people into distinctive groups, particularly in relation to race (Williams and Husk 2013), we cannot aim to reduce inequalities if we cannot see where they lie, what effect they have, and who they affect. Measuring ethnicity is therefore both equally challenging and necessary. By focusing on balancing the conceptual complexity and practical needs of measuring ethnicity, in particular by addressing some of the core issues relating to representativeness (respondent recognition of themselves within categories), conceptual coherence, (meaningful information conveyed through measures), and methodological practicality (useability of data for meaningful analyses), significant improvements to the ways we measure and categorise ethnicity may be made. Much work remains in revealing the intricate and expansive ways that systems of racial and ethnic inequalities operate and affect individuals and societies, not least in relation to violence. Improving the quality and representativeness of ethnicity data is a critical starting point to revealing hidden victims, characteristics, and outcomes at the intersection of violence and inequality.

Author Contributions: Conceptualisation, H.M.; methodology, H.M.; formal analysis, H.M. and N.B.; investigation, H.M.; data curation, H.M.; writing—original draft preparation, H.M.; writing—review and editing, H.M., N.B. and E.C.B.; supervision, E.C.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the UK Prevention Research Partnership (Violence, Health and Society; MR-VO49879/1), which is funded by the British Heart Foundation, Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Health and Social Care Research and Development Division (Welsh Government), Medical Research Council, National Institute for Health Research, Natural Environment Research Council, Public Health Agency (Northern Ireland), The Health Foundation, and Wellcome. The views expressed in this article are those of the authors and not necessarily those of the UK Prevention Research Partnership. The funder had no role in the study design, data collection, data analysis, data interpretation, or writing of this paper, or in the decision to submit the paper for publication. This research was additionally supported by the Violence and Society Centre Doctoral Studentship as part of the first author's Doctoral research, provided by City, University of London.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. The secondary data analyses were approved by the committee at City, University of London that considers medium-risk applications (ETH21220–299).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study via the CSEW.

Data Availability Statement: Data are available via the UK Data Service Secure Lab (Office for National Statistics 2023d).

Acknowledgments: The authors would like to thank Polina Obolenskaya, Ravi Thiara, Alexandria Innes, and Adrienne Yong for their comments and assistance on drafts of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A Timeline of Changes to the CSEW's Measurement and Categorisation of Ethnicity

1982	The CSEW (then the British Crime Survey) is launched (core sample size approx. 10,000). Interviewers record the observed 'race' of respondents as: (1) 'white', (2) 'Black (West Indian or African)', (3) 'Indian/Pakistani/Bangladeshi', (4) 'Other non-white', or (5) 'Mixed/uncertain'.
1984	Ethnic boost sample introduced (not present in 1998 and after 2007). 'Mixed/uncertain' category replaced with 'Other'.
1988	Interviewer assigned race is discontinued. Respondents are now asked "Which of the groups on this card best describe you?": (1) 'white', (2) 'black', (3) 'Indian', (4) 'Pakistani', (5) 'Bangladeshi', (6) 'Chinese', (7) 'Other Asian', or (8) 'Other', with write-in options for the latter two.
1992	Expanded ethnic categories: distinction between 'Black—Caribbean', 'Black—African' and 'Black—Other'. Removed 'Other Asian' category.
1994	Question now asks respondents "To which of these groups do you consider you *belong*?" [emphasis added]. The variable now references 'ethnicity' instead of 'race'.
2000	Mixed race' category added.
2001	Expanded ethnic categories and introduced high and low-level categorisation; amended question to align with 2000 Census; added a variable (Ethnic1) for 'Other' ethnic groups to specify 'English', 'Scottish', 'Welsh' or 'Other (specify)' background; amended category names to allow 'British' identifiers in other groups (not only 'White').
	Respondents asked to "choose one answer on this card to indicate your cultural background" [emphasis added] (low-level categories): (1) 'White—British', (2) 'White—Irish', (3) 'White—Other White Background', (4) 'Mixed—White and Black Caribbean', (5) 'Mixed—White and Black African', (6) 'Mixed—White and Asian', (7) 'Mixed—Any Other Mixed Background', (8) 'Asian or Asian British—Indian', (9) 'Asian or Asian British—Pakistani', (10) 'Asian or Asian British—Bangladeshi', (11) 'Asian or Asian British—Other Asian Background', (12) 'Black or Black British—Caribbean', (13) 'Black or Black British—African', (14) 'Black or Black British—Other Black Background', (15) 'Chinese', or (16) 'Other ethnic group'.
	Introduced a derived variable for high-level categories based on low-level responses: (1) 'white', (2) 'Mixed', (3) 'Asian', (4) Black', and (5) 'Other'.
2007	Ethnic boos sample discontinued.
2011	Relabelled categories to align with 2011 Census wording which allows for non-White ethnicities to also be 'British' (e.g., 'Black' changed to 'Black/Black British')
	Recategorised 'Chinese' as a high-level ethnic group (previously subsumed within 'Chinese/Other')
	Low-level ethnicity variable moved to the secure access dataset.
2012	Introduced low-level ethnic groups "White—Gypsy or Irish Traveller' (within the 'White' high-level group) and 'Arab' (within the 'Other' high-level group).

Appendix B Variable Description Table

Variable	Description	Values	Recoding
Independent Variables			
ONS Standardised Ethnicity (5 high-level groups)	High-level categorisation of ethnicity based on the ONS standardised measure of ethnicity. Responses derived from the survey question: "To which of these ethnic groups do you consider you belong?" (pre-2012/13) and "What is your ethnic group?" (post-2011)	1 = White; 2 = Mixed; 3 = Asian; 4 = Black; 5 = Other	Recoded post-2011 'Chinese' respondents into the high-level group 'Chinese/Other' for comparability with pre-2012/13 categorisation.
ONS Standardised Ethnicity (16 low-level groups)	Low-level categorisation of ethnicity based on the ONS standardised measure of ethnicity. Responses derived from the survey question: "To which of these ethnic groups do you consider you belong?" (pre-2012/13) and "What is your ethnic group?" (post-2011)	1 = 'White—British'; 2 = 'White—Irish'; 3 = 'White—Other White Background'; 4 = 'Mixed—White and Black Caribbean'; 5 = 'Mixed—White and Black African'; 6 = 'Mixed—White and Asian'; 7 = 'Mixed—Any Other Mixed Background'; 8 = 'Asian or Asian British—Indian'; 9 = 'Asian or Asian British—Pakistani'; 10 = 'Asian or Asian British—Bangladeshi'; 11 = 'Asian or Asian British—Other Asian Background'; 12 = 'Black or Black British—Caribbean'; 13 = 'Black or Black British—African'; 14 = 'Black or Black British—Other Black Background'; 15 = 'Chinese'; or 16 = 'Other ethnic group'	Recoded post-2011 'Arab' respondents into the low-level category 'Other', and 'White—Gypsy/Traveller' into the low-level category 'White—Other' for comparability with pre-2012/13 categorisation.
Alternative Recategorised Ethnicity (7 high-level groups)	Alternative categorisation of ethnicity proposed and tested in this paper, based on standardised low-level categories and regional origin (see Figure 1)	1 = White; 2 = Black/Black British; 3 = Arab/MENA; 4 = South Asian; 5 = ESEC Asian; 6 = Latinx/Hispanic; 7 = Other	See Figure 1 for recoding process
Dependent Variables			
Violence victimisation	Binary variable indicating whether the respondent had been a victim of a physical or sexual violence offence in the past 12 months. Based on whether the respondent was coded with a violent offence code by the ONS, based on various questions in the victim form.	0 = No, not a victim of violence in the past 12 months; 1 = Yes, a victim of violence in the past 12 months	Recoded respondents who had been victimised by at least one offence defined by the ONS as physical or sexual violence (serious wounding, other wounding, common assault, attempted assault, robbery, attempted robbery, snatch theft from the person, rape, serious wounding with sexual motive, other wounding with sexual motive, attempted rape, indecent assault) as violence victims.

Variable	Description	Values	Recoding
Dependent Variables			
Fear of Violence	Binary variable indicating whether the respondent reported fear of violence, based on the survey question "How worried are you about being physically attacked by strangers?"	0 = Not very/not at all worried; 1 = Fairly/very worried	Recoded respondents who reported being 'not very' or 'not at all' worried or reported themselves as 'not applicable' as 0 (no/low fear); and those who reported being 'fairly' or 'very' worried about stranger violence as 1 (fear)
Control Variables			
Mixed/Multiple Ethnicity	Additional variable indicating whether the respondent identified with one or multiple ethnic groups	0 = Single ethnicity; 1 = Mixed/Multiple ethnicities	Recoded respondents as having single or mixed/multiple ethnicity based on the high-level standardised ethnicity variable
Sex	Binary variable indicating whether the respondent is male or female, based on the interviewer's observation (asking "Is (name) male or female?" if needed).	0 = Male; 1 = Female	None
Migrant-status	Binary variable indicating whether the respondent was born in the UK (UK-born) or not (migrant), based on the questions "In which country were you born?"	0 = UK-born; 1 = Migrant	Recoded respondents as 'UK-born' if they were born in the UK and as 'migrant' if they answered 'Republic of Ireland' or 'Somewhere else'
Occupational Class	Categorical variable indicating the occupational class of the respondent derived by the ONS from multiple questions on occupation and employment.	1 = Higher managerial; 2 = Higher professional; 3 = Lower managerial and professional/higher technical occupations; 4 = Intermediate occupations; 5 = Small employers and own account workers; 6 = Lower supervisory and technical occupations; 7 = Semi-routine occupations; 8 = Routine occupations; 9 = Never worked; 10 = Not classified	None
Age	Continuous variable indicating the respondent's age (16+)	N/A Continuous variable	Removed responses 'don't know' and 'refused'
Other Variables			
Wave	Categorical variable indicating the year in which the survey wave was initiated (annually)	2004; 2005; 2006; 2007; 2013; 2014; 2015; 2016; 2017; 2018; 2019	Coded respondents by the year in which the survey wave was initiated. Coding based on individual datasets (per wave) which were then combined.
Weight	ONS weighting variable for population and non-response	N/A Weighting variable	Recoded for comparability over time (combining multiple comparable variables over time)

Appendix C Regional Origin Groupings (Adapted from ONS Regional Categorisation)

Region	Countries/Territories
UK-born	UK
Europe (excluding UK)	Albania; Andorra; Austria; Belarus; Belgium; Bosnia and Herzegovina; Bulgaria; Croatia; Cyprus; Czech Republic; Denmark; Estonia; Finland; France; Germany; Gibraltar; Greece; Hungary; Iceland; Ireland; Italy; Kosovo; Latvia; Liechtenstein; Lithuania; Luxembourg; Malta; Moldova; Monaco; Montenegro; Netherlands; North Macedonia; Norway; Poland; Portugal; Romania; Russia; San Marino; Serbia; Slovakia; Slovenia; Spain; Sweden; Switzerland; Ukraine
North America	Bermuda; Canada; Greenland; United States of America
Caribbean	Anguilla; Antigua and Barbud; Aruba; Bahamas; Barbados; British Virgin Islands; Cayman Islands; Cuba; Curaçao; Dominica; Dominican Republic; Grenada; Guatamala; Haiti; Jamaica; Martinique; Puerto Rico; Saint Kitts and Nevis; Saint Lucia; Saint Vincent and the Grenadines; Trinidad and Tobago; Turks and Caicos Islands; United States Virgin Islands
Latin America	Argentina; Belize; Bolivia; Brazil; Chile; Colombia; Costa Rica; Ecuador; El Salvado; Falkland Islands (Malvinas); French Guiana; Guatamala; Guyana; Honduras; Mexico; Nicaragua; Panama; Paraguay; Peru; Suriname; Uruguay; Venezuela
Africa (excluding North Africa)	Angola; British Indian Ocean Territory; Burundi; Cameroon; Central African Republic; Chad; Comoros; Congo; Djibouti; Eritrea; Ethiopia; French Southern Territories; Kenya; Madagascar; Malawi; Mauritius; Mayotte; Mozambique; Réunion; Rwanda; Seychelles; Somalia; South Sudan; Uganda; United Republic of Tanzania; Zambia; Zimbabwe
Middle East and North Africa (MENA)	Algeria; Armenia; Azerbaijan; Bahrain; Egypt; Georgia; Iran; Iraq; Israel; Jordan; Kurdistan; Kuwait; Lebanon; Libya; Morocco; Oman; Palestine; Qatar; Saudi Arabia; Sudan; Syria; Tunisia; Turkey; United Arab Emirates; Western Sahara; Yemen
South Asia	Afghanistan; Bangladesh; Bhutan; India; Iran; Maldives; Nepal; Pakistan; Sri Lanka; Kashmir
East, South East and Central (ESEC) Asia	Brunei; Cambodia; China; East Timor; Hong Kong; Indonesia; Japan; Kazakhstan; Kyrgyzstan; Laos; Macao; Malaysia; Mongolia; Myanmar; North Korea; Philippines; Singapore; South Korea; Taiwan; Tajikistan; Turkmenistan; Thailand; Uzbekistan; Vietnam
Oceana/Other	American Samoa; Australia; Cook Islands; Fiji; French Polynesia; Guam; Kiribati; Marshall Islands; Micronesia; Nauru; New Caledonia; New Zealand; Palau; Papua New Guinea; Samoa; Solomon Islands; Tonga; Tuvalu; Vanuatu

Appendix D Detailed Table A1: Logistic Regression Results by Rotating Referent Groups

Table A1. (a) Odds ratios from logistic regressions of violence victimisation using standardised ethnicity and regrouped ethnicity as predictors, pooled data (waves 2004–2007, 2013–2019), sample 1 (N = 415,361) *. (b) Odds ratios from logistic regressions of fear of violence using standardised ethnicity and regrouped ethnicity as predictors, pooled data (waves 2004–2007, 2013–2019), sample 2 (N = 169,440) *.

	(a)				
	Set of Models 1: Violence by standardised ethnicity (5 grps)				
	Changing reference category for ethnicity (5 groups):				
	1	2	3	4	5
Standardised Ethnicity (5 groups)					
1. White	*Ref*	0.839 †	1.577 ***	1.081	1.483 **
2. Mixed	1.192 †	*Ref*	1.880 ***	1.289 *	1.768 ***
3. Asian/Asian British	0.634 ***	0.532 ***	*Ref*	0.686 ***	0.940
4. Black/Black British	0.925	0.776 *	1.459 ***	*Ref*	1.372 *
5. Chinese/Other	0.674 **	0.566 ***	1.063	0.729 *	*Ref*

Table A1. *Cont.*

	Set of Models 2: Violence by regrouped ethnicity (7 grps), controlling for 'Mixed' ethnicity						
	Changing reference category for ethnicity (7 groups):						
	1	2	3	4	5	6	7
Ethnic regrouping (7 groups)							
1. White	*Ref*	1.132	0.893	1.584 ***	1.641 ***	0.833	1.211
2. Black/Black British	0.884	*Ref*	0.789	1.400 ***	1.450 **	0.736	1.070
3. Arab/MENA	1.120	1.267	*Ref*	1.774 *	1.838 *	0.933	1.356
4. South Asian	0.631 ***	0.714 ***	0.564 *	*Ref*	1.036	0.526 **	0.765 †
5. ESEC Asian	0.609 ***	0.690 **	0.544 *	0.965	*Ref*	0.508 **	0.738 †
6. Latinx/Hispanic	1.200	1.358	1.072	1.902 **	1.970 **	*Ref*	1.454
7. Other	0.826	0.934	0.737	1.308 †	1.355 †	0.688	*Ref*
Mixed Ethnicity (Ref = Not mixed)							
Mixed/Multiple Ethnicities	1.492 ***						
	(b)						
	Set of Models 3: Fear by standardised ethnicity (5 grps)						
	Changing reference category for ethnicity (5 groups)						
	1	2	3	4	5		
Standardised Ethnicity (5 groups)							
1. White	*Ref*	0.798 ***	0.508 ***	0.612 ***	0.690 ***		
2. Mixed	1.253 ***	*Ref*	0.637 ***	0.766 ***	0.864 †		
3. Asian/Asian British	1.967 ***	1.570 ***	*Ref*	1.203 ***	1.357 ***		
4. Black/Black British	1.635 ***	1.305 ***	0.831 ***	*Ref*	1.128 †		
5. Chinese/Other	1.450 ***	1.158 †	0.737 ***	0.887 †	*Ref*		
	Set of Models 4: Fear by regrouped ethnicity (7 grps), controlling for 'Mixed'						
	Changing reference category for ethnicity (7 groups)						
	1	2	3	4	5	6	7
Ethnic regrouping (7 groups)							
1. White	*Ref*	0.599 ***	0.572 ***	0.507 ***	0.576 ***	0.663 **	0.784 ***
2. Black/Black British	1.670 ***	*Ref*	0.956	0.846 ***	0.962	1.107	1.310 ***
3. Arab/MENA	1.747 ***	1.046	*Ref*	0.885	1.007	1.159	1.371 *
4. South Asian	1.974 ***	1.182 ***	1.13	*Ref*	1.137 *	1.309 †	1.548 ***
5. ESEC Asian	1.736 ***	1.040	0.993	0.880 *	*Ref*	1.151	1.362 ***
6. Latinx/Hispanic	1.508 **	0.903	0.863	0.764 †	0.869	*Ref*	1.183
7. Other	1.275 ***	0.763 ***	0.729 *	0.646 ***	0.734 ***	0.845	*Ref*
Mixed Ethnicity (Ref = Not mixed)							
Mixed/Multiple Ethnicities	0.783 ***						

Notes: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix E Supplementary Analyses of Migrant-Only Sample

Table A2. (a) Odds ratios from logistic regressions of violence victimisation using standardised ethnicity and regrouped ethnicity as predictors for Migrant-Only Sample, pooled data (waves 2004–2007, 2013–2019), sample 3 (N = 49,283) *. (b) Odds ratios from logistic regressions of fear of violence using standardised ethnicity and regrouped ethnicity as predictors for Migrant-Only Sample, pooled data (waves 2004–2007, 2013–2019), sample 4 (N = 24,310) *.

	(a)						
	Set of Models 5: Violence by standardised ethnicity (5 grps)						
	Changing reference category for ethnicity (5 groups):						
	1	2	3	4	5		
Standardised Ethnicity (5 groups)							
1. White	*Ref*	0.56 **	1.413 ***	0.916	1.734 ***		
2. Mixed	1.783 **	*Ref*	2.519 ***	1.633 *	3.092 ***		
3. Asian/Asian British	0.708 ***	0.397 ***	*Ref*	0.648 ***	1.228		
4. Black/Black British	1.092	0.613 *	1.543 ***	*Ref*	1.894 ***		
5. Chinese/Other	0.577 ***	0.323 ***	0.815	0.528 ***	*Ref*		
	Set of Models 6: Violence by regrouped ethnicity (7 grps), controlling for 'Mixed' ethnicity						
	Changing reference category for ethnicity (7 groups):						
	1	2	3	4	5	6	7
Ethnic regrouping (7 groups)							
1. White	*Ref*	0.949	0.838	1.453 ***	1.716 ***	0.859	1.608 *
2. Black/Black British	1.054	*Ref*	0.884	1.532 ***	1.809 ***	0.905	1.695 *
3. Arab/MENA	1.193	1.132	*Ref*	1.733 *	2.047 *	1.025	1.918 *
4. South Asian	0.688 ***	0.653 ***	0.577 *	*Ref*	*Ref*	0.591 *	1.107
5. ESEC Asian	0.583 ***	0.553 **	0.488 *	0.847	1.181	0.500 **	0.937
6. Latinx/Hispanic	1.164	1.105	0.976	1.692 *	1.998 *	*Ref*	1.872 *
7. Other	0.622 *	0.590 *	0.521 *	0.904	1.067	0.534 *	*Ref*
Mixed Ethnicity (Ref = Not mixed)							
Mixed/Multiple Ethnicities	2.102 ***						
	(b)						
	Set of Models 3: Fear by standardised ethnicity (5 grps)						
	Changing reference category for ethnicity (5 groups)						
	1	2	3	4	5		
Standardised Ethnicity (5 groups)							
1. White	*Ref*	0.637 ***	0.467 ***	0.504 ***	0.613 ***		
2. Mixed	1.570 ***	*Ref*	0.732 **	0.791 *	0.962		
3. Asian/Asian British	2.143 ***	1.365 **	*Ref*	1.080	1.314 ***		
4. Black/Black British	1.983 ***	1.264 *	0.926	*Ref*	1.216 **		
5. Chinese/Other	1.631 ***	1.039	0.761 ***	0.823 **	*Ref*		
	Set of Models 4: Fear by regrouped ethnicity (7 grps), controlling for 'Mixed'						
	Changing reference category for ethnicity (7 groups)						
	1	2	3	4	5	6	7
Ethnic regrouping (7 groups)							
1. White	*Ref*	0.496 ***	0.518 ***	0.468 ***	0.492 ***	0.622 ***	0.727 ***
2. Black/Black British	2.016 ***	*Ref*	1.044	0.943	0.992	1.254	1.467 ***
3. Arab/MENA	1.931 ***	0.957	*Ref*	0.903	0.950	1.200	1.404 *

Table A2. *Cont.*

	(b)						
	Set of Models 4: Fear by regrouped ethnicity (7 grps), controlling for 'Mixed'						
	Changing reference category for ethnicity (7 groups)						
	1	2	3	4	5	6	7
Ethnic regrouping (7 groups)							
4. South Asian	2.137 ***	1.060	1.107	*Ref*	1.051	1.329 *	1.554 ***
5. ESEC Asian	2.033 ***	1.008	1.053	0.951	*Ref*	1.264	1.478 ***
6. Latinx/Hispanic	1.608 ***	0.798	0.833	0.753 *	0.791	*Ref*	1.170
7. Other	1.375 ***	0.682 ***	0.712 *	0.643 ***	0.676 ***	0.855	*Ref*
Mixed Ethnicity (Ref = Not mixed)							
Mixed/Multiple Ethnicities	0.857						

Notes: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix F Supplementary Analyses of Main Regression Models Without Controls

Table A3. Odds ratios from uncontrolled logistic regressions of violence victimisation and fear of violence using standardised ethnicity and regrouped ethnicity as predictors, pooled data (waves 2004–2007, 2013–2019).

	Violence Victimisation	**Fear of Violence**
	OR	**OR**
Standardised Ethnicity (5 groups)		
1. White	*Ref*	*Ref*
2. Mixed	1.485 ***	1.362 ***
3. Asian/Asian British	0.769 ***	2.118 ***
4. Black/Black British	1.027	1.936 ***
5. Chinese/Other	0.828	1.705 ***
Ethnic regrouping (7 groups)		
1. White	*Ref*	*Ref*
2. Black/Black British	0.997	1.967 ***
3. Arab/MENA	1.429	2.178 ***
4. South Asian	0.767 ***	2.118 ***
5. ESEC Asian	0.720 **	1.891 ***
6. Latinx/Hispanic	1.165	1.925 ***
7. Other	0.940	1.423 ***
Mixed Ethnicity (Ref = Not mixed)		
Mixed/Multiple Ethnicities	1.950 ***	0.743 ***

Notes: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes

1 Using shorthand for waves, e.g. 2004/05 = 2004

2 The secure datasets are subject to additional security requirements by the ONS as they contain sensitive and potentially disclosive data. We used the secure dataset to include the low-level ethnicity variable (which is missing from the non-secure dataset for several years) in our merged dataset.

3 Whilst the CSEW has since recategorised 'Chinese' from 'Chinese/Other' into 'Asian', a lack of comparable variables for the study period required us to use the 'Chinese/Other' version (see Appendices A and B).

4 In focusing our analysis of violence/fear of violence on migrants only, the only ones who could potentially be recategorised based on country of origin, led to the same conclusions (Appendix E).

5 Further analyses found that 'White' and 'Asian' African-origin respondents corresponded with countries subject to British colonial rule and policies which provided historical explanations to the ethnic heterogeneity of African-origin migrants to the UK (e.g., the expulsion of Asians from Uganda).

[6] Note: although with only marginally significant differences when comparing the former to 'Arab/MENA' and 'Other' groups, respectively.

References

Ahlmark, Nanna, Maria Holst Algren, Teresa Holmberg, Marie Louise Norredam, Signe Smith Nielsen, Astrid Benedikte Blom, Anne Bo, and Knud Juel. 2015. Survey nonresponse among ethnic minorities in a national health survey–a mixed-method study of participation, barriers, and potentials. *Ethnicity & Health* 20: 611–32.

Alcoff, Linda Martín. 1999. Towards a phenomenology of racial embodiment. *Radical Philosophy* 95: 15–26.

Anthias, Floya. 1992. *Ethnicity, Class, Gender and Migration: Greek Cypriots in Britain*. Aldershot: Avebury.

Anthias, Floya. 2007. Ethnic Ties: Social Capital and the Question of Mobilisability. *The Sociological Review* 55: 788–805. [CrossRef]

Anthias, Floya, and Nira Yuval-Davis. 2005. *Racialized Boundaries: Race, Nation, Gender, Colour and Class and the Anti-Racist Struggle*. London: Routledge.

Aspinall, Peter. 2000. The challenges of measuring the ethno-cultural diversity of Britain in the new millennium. *Policy and Politics* 28: 109–18. [CrossRef]

Aspinall, Peter. 2009. The future of ethnicity classifications. *Journal of Ethnic and Migration Studies* 35: 1417–35. [CrossRef]

Ballard, Roger. 1997. 'The construction of a conceptual vision: "ethnic groups" and the 1991 UK Census'. *Ethnic and Racial Studies* 20: 182–94. [CrossRef]

Barot, Rohit, and John Bird. 2001. Racialization: The genealogy and critique of a concept. *Ethnic & Racial Studies* 24: 601–18.

Bjørnholt, Margunn, and Ole Kristian Hjemdal. 2018. Measuring violence, mainstreaming gender: Does adding harm make a difference? *Journal of Gender-Based Violence* 2: 465–79. [CrossRef]

Blair, Laurence. 2019. Latin Americans are one of the UK's Fastest-Growing Groups. So Why Aren't They Recognised? *The New Statesman*. December 11. Available online: https://www.newstatesman.com/politics/2019/12/latin-americans-are-one-of-the-uks-fastest-growing-groups-so-why-arent-they-recognised (accessed on 1 January 2024).

Blom, Niels. 2023. *Code for Merging Waves of the Crime Survey of England and Wales and the British Crime Survey, 1982–2020*. [Data Collection]. Colchester: UK Data Service. [CrossRef]

Brooks, Rachel. 2019. Representations of East Asian students in the UK media. In *Asian Migration and Education Cultures in the Anglosphere*. London: Routledge, pp. 81–95.

Bruce, Marino A., and Vincent J. Roscigno. 2003. "Race effects" and conceptual ambiguity in violence research: Bringing inequality back in. In *Interpersonal Violence: The Ethnicity, Race, and Class Nexus*. Cambridge: Cambridge University Press, pp. 238–53.

Burton, Jonathan, Alita Nandi, and Lucinda Platt. 2010. Measuring ethnicity: Challenges and opportunities for survey research. *Ethnic and Racial Studies* 33: 1332–49. [CrossRef]

Bush, Stephen. 2022. What Is an Ethnic Minority Anyway? *The Financial Times*. November 8. Available online: https://www.ft.com/content/62ff7211-312d-4cf0-b919-aaf9fe20fde9 (accessed on 1 January 2024).

Callister, Paul, Robert Didham, and Anna Kivi. 2009. *Who Are We? The Conceptualisation and Expression of Ethnicity*. Wellington: Statistics New Zealand.

Carr, J., J. James, J. Clifton-Sprigg, and S. Vujic. 2022. *Hate in the time of COVID-19: Racial Crimes against East Asians*. Bonn: IZA—Institute of Labor Economics.

Coalition of Latin Americans in the UK. 2023. Coalition of Latin Americans in the UK. Available online: https://clauk.org.uk/ (accessed on 1 January 2024).

Cooper, Kerris, and Polina Obolenskaya. 2021. Hidden victims: The gendered data gap of violent crime. *The British Journal of Criminology* 61: 905–25. [CrossRef]

Cormack, D., and C. Robson. 2010. *Classification and Output of Multiple Ethnicities: Issues for Monitoring Māori Health*. Wellington: Te Rōpū Rangahau Hauora a Eru Pōmare.

Del Pinal, J., E. Martin, C. Bennett, and A. Cresce. 2007. Overview of Results of New Race and Hispanic Origin Questions in Census 2000. In *Research Report Series: Survey Methodology #2007-5*; Washington: U.S. Census Bureau. Available online: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=93faff9e7a7a4b7044a3b3c64ca90719e0b77ad5 (accessed on 1 January 2024).

Department for Work and Pensions. 2022. State Support. In *Ethnicity Facts and Figures*. Available online: https://www.ethnicity-facts-figures.service.gov.uk/work-pay-and-benefits/benefits/state-support/latest/ (accessed on 1 January 2024).

Didham, Robert, and Paul Callister. 2012. The effect of ethnic prioritisation on ethnic health analysis: A research note. *The New Zealand Medical Journal* 125: 58–66. [PubMed]

European Social Survey. 2020. *ESS Round 10 Source Questionnaire*. London: ESS ERIC Headquarters c/o City, University of London.

Femi-Ajao, Omolade, Sarah Kendal, and Karina Lovell. 2020. A qualitative systematic review of published work on disclosure and help-seeking for domestic violence and abuse among women from ethnic minority populations in the UK. *Ethnicity & Health* 25: 732–46.

Fenton, S., and H. Bradley. 2002. *Ethnicity and Economy*. Basingstoke: Palgrave Macmillan.

Finney, Nissa, James Nazroo, Laia Bécares, Dharmi Kapadia, Natalie Shlomo, Nico Ochmann, Joseph Harrison, Michaela Kyclova, Harry Taylor, Emma Hill, and et al. 2023. *Racism and Ethnic Inequality in a Time of Crisis*. Policy Press: Bristol.

Garrett, Helen, Justine Piddington, and Simon Nicol. 2014. The housing conditions of minority ethnic households in England. *Better Housing Briefing Paper* 24: 1–12. Available online: https://raceequalityfoundation.org.uk/wp-content/uploads/2022/09/housing_briefing_24-final.pdf (accessed on 1 January 2024).

Goldstein, Joshua R., and Ann J. Morning. 2000. The multiple-race population of the United States: Issues and estimates. *Proceedings of the National Academy of Sciences* 97: 6230–35. [CrossRef] [PubMed]

Greil, Arthur L., Julia McQuillan, Karina M. Shreffler, Katherine M. Johnson, and Kathleen S. Slauson-Blevins. 2011. Race-ethnicity and medical services for infertility: Stratified reproduction in a population-based sample of US women. *Journal of health and social behavior* 52: 493–509. [CrossRef] [PubMed]

Ha, Sierra K., Ann T. Nguyen, Chloe Sales, Rachel S. Chang, Hillary Ta, Malathi Srinivasan, Sukyung Chung, Latha Palaniappan, and Bryant Lin. 2020. Increased self-reported discrimination and concern for physical assault due to the COVID-19 pandemic in Chinese, Vietnamese, Korean, Japanese, and Filipino Americans. *medRxiv*, 2020–9. [CrossRef]

Holup, Joan L., Nancy Press, William M. Vollmer, Emily L. Harris, Thomas M. Vogt, and Chuhe Chen. 2007. Performance of the U.S. Office of Management and Budget's revised race and ethnicity categories in Asian populations. *International Journal of Intercultural Relations* 31: 561–73. [CrossRef] [PubMed]

Hoover, Eddie L. 2007. There is no scientific rationale for race-based research. *Journal of the National Medical Association* 99: 690. [PubMed]

Innes, A., Niels Blom, A. Bunce, Anastasiia Fadeeva, H. Manzur, R. Thiara, and Estela Capelas Barbosa. 2023. *Tool to Mitigate Risk of Bias when working with data on Ethnicity and Migration*. VISION. [Online tool and publication]. Available online: https://openaccess.city.ac.uk/id/eprint/31787/ (accessed on 1 January 2024).

Kautt, Paula M. 2011. Ethnicity and the British crime survey: Issues, obstacles and solutions. *The Howard Journal of Criminal Justice* 50: 275–88. [CrossRef]

Khanna, Nikki. 2010. "If You're Half Black, You're Just Black": Reflected Appraisals and the Persistence of the One-Drop Rule. *The Sociological Quarterly* 51: 96–121. [CrossRef]

Kumar, Sumit, Lawrence W. Sherman, and Heather Strang. 2020. Racial disparities in homicide victimisation rates: How to improve transparency by the Office of National Statistics in England and Wales. *Cambridge Journal of Evidence-Based Policing* 4: 178–86. [CrossRef]

Laux, Richard. 2019. 50 years of collecting ethnicity data. History of Government. Cabinet Office. [Online blog]. Available online: https://history.blog.gov.uk/2019/03/07/50-years-of-collecting-ethnicity-data/ (accessed on 1 January 2024).

Laws, Barton M., and Rachel A. Heckscher. 2002. Racial and ethnic identification practices in public health data systems in New England. *Public Health Reports* 117: 50–61. [CrossRef] [PubMed]

Modood, Tariq, Richard Berthoud, and James Nazroo. 2002. "'Race', racism and ethnicity": A response to Ken Smith. *Sociology* 36: 419–27. [CrossRef]

Morning, Ann. 2014. And you thought we had moved beyond all that: Biological race returns to the social sciences. *Ethnic and Racial Studies* 37: 1676–85. [CrossRef]

Morning, Ann. 2015. Ethnic classification in global perspective: A cross-national survey of the 2000 census round. In *Social Statistics and Ethnic Diversity: Cross-National Perspectives in Classifications and Identity Politics*. Cham: Springer, pp. 17–37.

Nerenz, David R., Bernadette McFadden, and Cheryl Ulmer, eds. 2009. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington: The National Academic Press.

NHS Digital. 2022. Health Survey England Additional Analyses, Ethnicity and Health, 2011–2019. *Health Survey England Additional Analyses*. Available online: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-england-additional-analyses/ethnicity-and-health-2011-2019-experimental-statistics# (accessed on 9 April 2024).

Office for Management and Budget. 2000. OMB bulletin No. 00-02. Available online: http://www.whitehouse.gov/omb/bulletins/b00-02.html (accessed on 1 January 2024).

Office for National Statistics. 2009. Information Paper: Deciding Which Tick-Boxes to Add to the Ethnic Group Question in the 2011 England and Wales Census. Available online: https://www.ons.gov.uk/file?uri=/census/2011census/howourcensusworks/howweplannedthe2011census/questionnairedevelopment/finalisingthe2011questionnaire/ethnicgroupprioritisationtool_tcm77-183986.pdf (accessed on 1 January 2024).

Office for National Statistics. 2021. Victims of Crime 2020 Archived. In *Ethnicity Facts and Figures*. Available online: https://www.ethnicity-facts-figures.service.gov.uk/crime-justice-and-the-law/crime-and-reoffending/victims-of-crime/latest/ (accessed on 1 January 2024).

Office for National Statistics. 2022. Ethnic Group, England and Wales: Census 2021. Available online: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021 (accessed on 1 January 2024).

Office for National Statistics. 2023a. Crime in England and Wales QMI. Available online: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/crimeinenglandandwalesqmi (accessed on 1 January 2024).

Office for National Statistics. 2023b. Understanding Consistency of Ethnicity Data Recorded in Health-Related Administrative Datasets in England: 2011 to 2021. Available online: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthinequalities/articles/understandingconsistencyofethnicitydatarecordedinhealthrelatedadministrativedatasetsinengland2011to2021/2023-01-16 (accessed on 1 January 2024).

Office for National Statistics. 2023c. Employment. In *Ethnicity: Facts and Figures*. Available online: https://www.ethnicity-facts-figures.service.gov.uk/work-pay-and-benefits/employment/employment/latest/ (accessed on 1 January 2024).
Office for National Statistics. 2023d. *Crime Survey for England and Wales, 1996–2020: Secure Access*. [Data Collection]. *12th Edition*. UK Data Service. SN: 7280. Available online: https://doi.org/10.5255/UKDA-SN-7280-12 (accessed on 1 January 2024).
Office for Statistics Regulation. 2021. Improving Health and Social Care Statistics: Lessons Learned from the COVID-19 Pandemic. Available online: https://osr.statisticsauthority.gov.uk/publication/improving-health-and-social-care-statistics-lessons-learned-from-the-covid-19-pandemic/ (accessed on 1 January 2024).
Oh, Hyunsu. 2021. The Association Between Discriminatory Experiences and Self-Reported Health Status among Asian Americans and Its Subethnic Group Variations. *Journal of Racial and Ethnic Health Disparities* 9: 1689–96. [CrossRef] [PubMed]
Payson, Kenneth E. 1996. Check One Box: Reconsidering Directive No. 15 and the Classification of Mixed-Race People. *California Law Review* 84: 1233–91. [CrossRef]
Powell, Andrew, and Brigid Francis-Devine. 2023. Research Briefing Number 6385: Unemployment by Ethnic Background. House of Commons Library. Available online: https://researchbriefings.files.parliament.uk/documents/SN06385/SN06385.pdf (accessed on 1 January 2024).
Prewitt, Kenneth. 2013. Opinion: Fix the Census' Archaic Racial Categories. *New York Times*. August 21. Available online: https://www.nytimes.com/2013/08/22/opinion/fix-the-census-archaic-racial-categories.html (accessed on 1 January 2024).
Pullerits, Merili, and Jessica Phoenix. 2023. How priority ordering of offence codes undercounts gendered violence: An analysis of the Crime Survey for England and Wales. *The British Journal of Criminology* 64: 381–99. [CrossRef]
Rhead, Rebecca D., Charlotte Woodhead, Gargie Ahmad, Jayati Das-Munshi, Sally McManus, and Stephani L Hatch. 2022. A comparison of single and intersectional social identities associated with discrimination and mental health service use: Data from the 2014 Adult Psychiatric Morbidity Survey in England. *Social Psychiatry and Psychiatric Epidemiology* 57: 2049–63. [CrossRef]
Rose, Hilary, and Steven Rose. 2005. Why we should give up on race. *The Guardian*. April 9. Available online: https://www.theguardian.com/world/2005/apr/09/race.science (accessed on 1 January 2024).
Saini, Angela. 2019. *Superior: The Return of Race Science*. Boston: Beacon Press.
Schneider, Mike. 2023. US Government Ponders the Meaning of Race and Ethnicity. *The Independent*. March 13. Available online: https://www.independent.co.uk/news/ap-black-office-of-management-and-budget-biden-japan-b2299576.html (accessed on 1 January 2024).
Song, Miri. 2010. Does 'race'matter? A study of 'mixed race'siblings' identifications. *The Sociological Review* 58: 265–85. [CrossRef]
Song, Miri. 2012. Making sense of 'mixture': States and the classification of 'mixed'people. *Ethnic and Racial Studies* 35: 565–73. [CrossRef]
UK Data Service. 2023. Ethnicity. Available online: https://ukdataservice.ac.uk/find-data/browse/ethnicity/ (accessed on 1 January 2024).
UK Government Equality Hub. 2023. Guidance: Standards for Ethnicity Data. Available online: https://www.gov.uk/government/publications/standards-for-ethnicity-data/standards-for-ethnicity-data (accessed on 1 January 2024).
UK Parliament Petition. 2021. Include Latinx/Hispanic as an ethnic group in the 2021 census". UK Government and Parliament Petition 557528. [Online Petition]. Available online: https://petition.parliament.uk/petitions/557528 (accessed on 1 January 2024).
Valles, S. A., Raj Singh Bhopal, and P. J. Aspinall. 2015. Census categories for mixed race and mixed ethnicity: Impacts on data collection and analysis in the US, UK and NZ. *Public health* 129: 266–70. [CrossRef] [PubMed]
Wagner, Jennifer K., Joon-Ho Yu, Jayne O. Ifekwunigwe, Tanya M. Harrell, Michael J. Bamshad, and Charmaine D. Royal. 2017. Anthropologists' views on race, ancestry, and genetics. *American Journal of Physical Anthropology* 162: 318–27. [CrossRef] [PubMed]
Walby, Sylvia, Jude Towers, and Brian Francis. 2016. Is violent crime increasing or decreasing? A new methodology to measure repeat attacks making visible the significance of gender and domestic relations. *British Journal of Criminology* 56: 1203–34. [CrossRef]
Walby, Sylvia. 2009. Globalization and inequalities: Complexity and contested modernities. In *Globalization and Inequalities*. Thousand Oaks: SAGE Publications, pp. 1–520.
Williams, David R. 1996. Race/ethnicity and socioeconomic status: Measurement and methodological issues. *International Journal of Health Services* 26: 483–505. [CrossRef]
Williams, Malcolm, and Kerryn Husk. 2013. Can we, should we, measure ethnicity? *International Journal of Social Research Methodology* 16: 285–300. [CrossRef]
Wood, Douglas. 1984. *British Crime Survey, 1982: Technical Report*. P.670/675. Mexico City: Centro de Excelencia UNODC-INEGI.
Woozeer, Laila. 2022. Opinion: As a 'Mixed' Person, the Language to Describe Me Isn't Fit for Purpose. *The Guardian*. August 30. Available online: https://www.theguardian.com/commentisfree/2022/aug/30/mixed-person-language-vocabulary (accessed on 1 January 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.