


BMJ Open Quality Racial and ethnic diversity in clinical studies reported to ClinicalTrials.gov, 2009–2024

Maryam Aziz,¹ Emily C O'Brien,^{1,2} Jay B Lusk,³ Sudarshan Krishnamurthy,⁴ Vani Garcha,¹ M Alan Brookhart,¹ Robert M Califf,^{2,5} Michael D Green ¹

To cite: Aziz M, O'Brien EC, Lusk JB, *et al*. Racial and ethnic diversity in clinical studies reported to ClinicalTrials.gov, 2009–2024. *BMJ Open Quality* 2026;**15**:e003801. doi:10.1136/bmjopen-2025-003801

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjopen-2025-003801>).

Received 25 August 2025
Accepted 10 May 2026



© Author(s) (or their employer(s)) 2026. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

¹Department of Population Health Sciences, Duke University School of Medicine, Durham, North Carolina, USA

²Duke University Duke Clinical Research Institute, Durham, North Carolina, USA

³Department of Family Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

⁴Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA

⁵Division of Cardiology, Duke University School of Medicine, Durham, North Carolina, USA

Correspondence to

Dr Michael D Green;
michael.d.green@duke.edu

ABSTRACT

Objective A lack of transparent reporting of race and ethnicity in clinical research limits the ability to identify health inequities and evaluate to what extent clinical research includes diverse populations. Our objectives are: (1) to identify study characteristics associated with reporting race and ethnicity of clinical study participants and (2) to document temporal trends in race and ethnicity reporting on ClinicalTrials.gov.

Design Cross-sectional analysis of interventional trials and observational studies from 2009 to 2024; multivariable logistic regression assessed study-level factors associated with reporting race and ethnicity.

Setting Global registry of clinical studies (ClinicalTrials.gov).

Participants 58 163 studies with posted results and without early termination.

Exposures Study characteristics: sponsor trial phase, study type and country.

Main outcomes and measures Reporting of race, reporting of ethnicity, reporting of both.

Results Among 58163 studies (mean enrolment=1215 participants), 44.8% did not report race or ethnicity to the repository (mean enrolment=1481 participants). The proportion of studies reporting both race and ethnicity rose from 7.4% in 2013 to 54.6% in 2024. In multivariable models, observational studies had lower odds of reporting race and ethnicity (OR 0.55, 95% CI 0.49 to 0.61) compared with interventional trials. Phase 4 trials were least likely phase to report race and ethnicity (OR=0.32; 95% CI 0.29 to 0.35), and studies with only National Institute of Health funding were more likely to report race and ethnicity compared with studies with any industry funding or sponsorship (OR=1.70, 95% CI 1.61 to 1.79). For studies that reported race, White participants comprised ≥50% each year based on study-level percentages; proportions of Asian participants declined, and Black participants fluctuated. 'Not Hispanic or Latino' remained ≥80% of reported ethnicity annually.

Conclusions Race and ethnicity reporting on ClinicalTrials.gov has improved markedly yet remains incomplete, with shortfalls in late-phase and observational studies.

BACKGROUND

Racial and ethnic disparities in access to care and health outcomes are well documented across a range of medical conditions.^{1–3}

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Racial and ethnic disparities in health outcomes are well documented, yet clinical studies often lack diverse participation. Prior assessments of diversity tend to focus on specific diseases or therapeutic areas, limiting our understanding of reporting practices across the broader clinical research landscape.

WHAT THIS STUDY ADDS

⇒ Among over 58000 clinical studies, reporting of both race and ethnicity rose from 7.4% in 2013 to 54.6% in 2024. National Institute of Health-sponsored studies, interventional trials and US-based studies were more likely to report race and ethnicity, while observational and phase 4 studies were less likely.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These findings suggest that while reporting has improved, gaps persist in certain study types. Efforts to address reporting in observational and late-phase studies alongside initiatives to recruit and retain diverse participants may help ensure clinical research better reflects the populations it aims to serve.

Race is a socially constructed trait which can sort humans into groups based on observed characteristics,^{4 5} exists as a byproduct of the process of racialisation, and can be used as a proxy for exposure to parts of systemic racism.⁶ Ethnicity is a socially constructed trait used or assigned to sort humans into groups based on characteristics like shared culture and national origin. Despite recognised disparities identified using race and ethnicity, an extensive body of literature suggests that clinical studies—the foundation of evidence-based medicine—often lack diverse participation and fail to adequately represent populations disproportionately affected by diseases.^{7–9} These representation gaps are concerning given that racial inequities in highly prevalent conditions like cardiovascular disease persist. Black patients, for instance, experience elevated

heart failure mortality compared with White patients.¹⁰ Similarly, among older adults in the USA, racial disparities in mortality become more pronounced as cardiovascular disease burden increases, even when Black adults initially show comparable or lower risks of disease progression.¹¹ Many researchers and institutions have emphasised the rationale for greater diversity in clinical research.^{8,9,12,13} Without appropriate representation, study findings may lack generalisability and risk exacerbating existing disparities among underrepresented racial and ethnic groups.^{14,15} In response, several US federal agencies launched initiatives prior to 2025 to encourage greater inclusion in clinical studies.^{16,17} Notable efforts include the National Institute of Health (NIH) Revitalisation Act of 1993¹⁸ and the Food and Drug Administration 2020 guidance on Enhancing the Diversity of Clinical Trial Populations.¹⁹

Despite these policy efforts, comprehensive and systematic evaluations of racial and ethnic representation in clinical studies remain limited.²⁰ Existing assessments tend to focus on specific diseases, time-periods or therapeutic areas, resulting in a lack of systematic evaluation of diversity in clinical trials.^{7,21–24} ClinicalTrials.gov, the largest publicly available repository of clinical study results, includes participant characteristics such as race and ethnicity.²⁵ Since 2017, the NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information requires clinical studies to report race and ethnicity information to ClinicalTrials.gov.⁷

To address existing gaps, we analysed more than 58000 interventional and observational clinical studies registered on ClinicalTrials.gov from 2000 to 2024. Our objectives were twofold:

1. To assess the state of race and ethnicity reporting to ClinicalTrials.gov and examine key study characteristics associated with reporting race and ethnicity.
2. To present the distribution of participant racial and ethnic identities across key study characteristics.

METHODS

Data extraction

We queried the ClinicalTrials.gov registry using version 2 of the application programming interface (API) from the registry's inception in February 2000 to 19 December 2024. This period was selected to align with the establishment of the NIH Office of Management and Budget (OMB) Race (and) Ethnicity Standard in 1997, ensuring that studies in the USA reported since the inception of ClinicalTrials.gov in 2000 would be consistent with reporting guidelines.

We only included studies with results using the Study Results binary indicator retrieved using the API version 2. On filtering, we retrieved 58163 studies. We obtained data for the following columns for each of the studies: NCT ID (the unique study identifier), Overall Status (status of study recruitment), Start Date (date first participant enrolled), Sponsor (the organisation or person

who initiates the study and has authority and control over the study), Funder Type (category of organisation that provides funding for the study), Primary Completion Date, Last Update Post Date, Study Type (interventional or observational), Phases, Enrolment, Characteristics (demographic information collected at beginning of study). We selected this subset of columns because we hypothesised they may impact race and ethnicity reporting practices.

We restricted our analysis to exclude terminated trials, which are 'stud(ies) (which) halted prematurely and will not resume; participants are no longer being examined or receiving intervention'.²⁶ Under the Public Health Service Act, only trials with approved or licensed products that reach their primary completion date are subject to reporting results to ClinicalTrials.gov. Terminated trials that are stopped before their primary completion date may have incomplete or inconsistently reported demographic results (including race and ethnicity), which could introduce misclassification bias into our analysis.²⁷

Additionally, we used publicly available data from the Aggregate Analysis of ClinicalTrials.gov (AACT) database to obtain information on the country where the trials were registered.²⁸ We used the 1 January 2025, pipe-delimited files to obtain the 'country' data which has the registered facilities/sites of the individual trial. AACT files are available at: aact.ctti-clinicaltrials.org/downloads.

Race and ethnicity data processing

The ClinicalTrials.gov data was retrieved from the API version 2 in a JSON format and Python (V.3.8.16) was used to load, parse and convert the data into a data frame. We then used Python to iteratively extract race and ethnicity data from the 'Characteristics' column, which revealed significant variability in the methods used to report race/ethnicity. In contrast, the other variables we pulled were provided in a standardised format. For race and ethnicity, some data were reported using the NIH/OMB standards, while others used customised categories and table formats. This variability resulted in the need for refinement of our data extraction script to uniformly capture race and ethnicity data. Our script returned an error for 29 entries (0.04% of studies) which contained empty tables or 'NA' placeholders—these entries had to be manually judged by a member of the study team.

Following the initial extraction process, we manually reviewed a random sample of 1000 rows for accuracy and completeness. The manual review uncovered 9 (0.9% of the 1000 checked) errors due to nested tables, inconsistencies in table naming, incorrect data types, and studies with more than 2 race/ethnicity tables. Examples of these cases are provided as an online supplemental file 2. We corrected seven of these errors by updating the logic in our Python script; however, we were unable to implement robust logic to address studies with greater than two race and ethnicity tables and with inconsistencies in table naming due to wide variation in the data structures. As a result, 115 studies (0.20% of all studies) with more

than two race/ethnicity tables and 27 studies (0.05% of all studies) with inconsistencies in table naming had to be manually adjudicated by a member of the study team. Following the corrections, we performed a manual validation of 250 rows where we identified no errors.

Race and ethnicity categorisation

We mapped all customised race and ethnicity categories to the standardised NIH/OMB categories as defined in the online supplemental table 1) (accessed in 2024).²⁹ When the race or ethnicity was listed as ‘unknown or not reported’, those participants were included in our participant level tabulations. However, studies that did not report any race or ethnicity data were excluded from analyses focused on participant diversity. We used keyword search to make initial categorisations. A member of the study team (VG) reviewed all categorisations and manually judged any records that did not align with the predefined keywords. The NIH/OMB definitions used for categorisations are presented in online supplemental table 1.

While all studies included in the ethnicity analysis reported some form of ethnicity data, many tables also included an ‘unknown or not reported’ category, which reflects the portion of participants for whom ethnicity is not provided. Additionally, a subset of studies (n=1560, 2.7% of all studies) reported race and ethnicity in a single combined table. In these cases, it was unclear whether ‘unknown or not reported’ referred to race, ethnicity or both. To avoid potential misclassification, these ambiguous entries were excluded from analyses.

Determining funding source for studies

We derived the funding source of a study based on the listed sponsors and collaborators using an approach documented elsewhere.^{30–32} The schema for determining the funding source of a study is presented in online supplemental table 2.

Country classification

Using the ‘countries’ pipe delimited file from the AACT database, we matched the listed country to the NCTID number for the clinical studies included in our analysis. All countries that were listed ‘United States’, ‘United States of America’, ‘USA’ or ‘US’ were considered United States-based studies, and all others were considered non-United States-studies. For studies with multiple countries listed, we also had a category for ‘at least-one US site’.

Analysis plan

We evaluated factors associated with the reporting of race and/or ethnicity information on ClinicalTrials.gov using multivariable logistic regression. Three binary outcomes were created: (1) reporting race, (2) reporting ethnicity and (3) reporting both. For each outcome, we fit a generalised linear model with a binomial error distribution and logit link using the RStudio glm function (RStudio 2024.04). Covariates included in the models were funder type (Industry (reference), NIH, Other, Other US

Federal), trial phase (phase 1 (reference), early phase 1, phase 1/2, phase 2, phase 2/3, phase 3, phase 4, other or unspecified), study type (Interventional (reference) vs Observational), and study location (no US sites (reference), at least 1 US-based site, only US based sites). We also performed a secondary analysis limited to studies conducted only in the USA to assess race and ethnicity reporting within a single regulatory environment.

Next, we examined temporal trends in race and ethnicity reporting. We presented trends in reporting both categories together and separately, as well as changes in participant diversity over time by study characteristics. To reduce the influence of larger studies on annual trends, we first calculated the within-study percentages for each race and ethnicity category, then averaged these percentages across studies for each year.

For these trends, we used the year of the study’s *Last Update Post Date*,³³ which reflects the most recent date on which changes to the study record (including race/ethnicity) were made. We selected this date because it best reflects the point at which results are most likely to be complete. In addition, other time fields such as study registration and start date often precede the availability of results. The study’s *Last Update Post Date* values range from 2009 to 2024; thus, 2009 was the first year that race and ethnicity data were reported to the repository.

Patient and public involvement

As an analysis of trial level data, this section is not applicable.

RESULTS

Our results consider a total of 58 163 clinical studies, with 72.2% having at least one site in the USA. **Table 1** presents key characteristics of the included studies, separated by reporting type.

Likelihood to report race and/or ethnicity

Among the studies registered worldwide between 2009 and 2024, several design and operational characteristics were independently associated with whether investigators reported both race and ethnicity to ClinicalTrials.gov (**figure 1**). Compared with interventional trials, observational studies had lower odds of reporting race and ethnicity (OR=0.55; 95% CI 0.49 to 0.61). Phase 4 trials were the least likely of all phases to report race and ethnicity (OR=0.32, 95% CI 0.29 to 0.35), and studies with other funding sources (OR=0.78, 95% CI 0.74 to 0.83) were also less likely to report race and ethnicity data. In contrast, studies with only US sites (OR=5.14, 95% CI 4.84 to 5.47), or with at least one site in the USA (OR=5.93, 95% CI 5.53 to 6.36) were more likely to report race and ethnicity compared with those outside the USA.

Temporal trends in reporting race and ethnicity

Reporting of both race and ethnicity has improved over time (**figure 2**). In 2024, 54.6% of all studies reported

Table 1 Characteristics of clinical trials stratified by race and ethnicity reporting structure

		Overall	Customised	NIH/OMB	Not collected
n		58163	10438 (17.95)	21663 (37.25)	26062 (44.80)
Phase, n (%)	Early phase 1	310 (0.5)	35 (0.3)	175 (0.8)	100 (0.4)
	Phase 1	4274 (7.3)	908 (8.7)	1883 (8.7)	1483 (5.7)
	Phase 2	16006 (27.5)	2767 (26.5)	7040 (32.5)	6199 (23.8)
	Phase 3	11706 (20.1)	2952 (28.3)	3512 (16.2)	5242 (20.1)
	Phase 4	6187 (10.6)	995 (9.5)	1696 (7.8)	3496 (13.4)
	Other phase	19680 (33.8)	2781 (26.6)	7357 (34.0)	9542 (36.6)
Sponsor, n (%)	Industry	32362 (55.6)	7133 (68.3)	10251 (47.3)	14978 (57.5)
	NIH	8624 (14.8)	1232 (11.8)	4846 (22.4)	2546 (9.8)
	Other	14358 (24.7)	1596 (15.3)	5449 (25.2)	7313 (28.1)
	Other US Federal	2819 (4.8)	477 (4.6)	1117 (5.2)	1225 (4.7)
Country, n (%)	At least one US site	9732 (16.7)	3025 (29.7)	3471 (16.3)	3236 (13.2)
	No US sites	14085 (24.2)	2360 (23.1)	2954 (13.9)	8771 (35.7)
	Only US sites	32258 (55.5)	4816 (47.2)	14 870 (69.8)	12572 (51.1)
	Missing	2088 (3.58)	–	–	–
Study type, n (%)	Interventional	54481 (93.7)	10026 (96.1)	20920 (96.6)	23535 (90.3)
	Observational	3682 (6.3)	412 (3.9)	743 (3.4)	2527 (9.7)
Enrolment, mean (SD)		1214.6 (36339.7)	2212.4 (55628.1)	412.9 (6711.2)	1481.4 (40859.3)

NIH, National Institute of Health; OMB, Office of Management and Budget.

both race and ethnicity, a gradual increase from 7.4% in 2013.

In studies reporting ethnicity, ‘Not Hispanic or Latino’ was the most common category, making up 83.2% of participants in 2024 (figure 3). While certain subgroups, including early phase 1 and industry-funded studies, showed higher proportions of participants identifying as ‘Hispanic or Latino’ in certain years, these differences were not sustained (online supplemental figures 1–12).

In studies that reported race, ‘White’ was the most common category, making up more than 50% of participants each year (figure 4). The racial distribution remained relatively stable over time, and there was a gradual decline in the proportion of ‘Asian’ participants, while the proportion of ‘Black or African American’ participants showed variability from year to year. When stratified by trial phase, funder type, and study design, the overall patterns persisted: ‘White’ participants remained the majority, and changes in representation of other racial and ethnic groups were inconsistent across subgroups. Stratified results are shown in online supplemental figures 1–12.

DISCUSSION

To our knowledge, this is the largest study to systematically document long-term trends in the reporting of race and ethnicity to ClinicalTrials.gov and to quantify the characteristics of studies that are more likely to report race and ethnicity. Sufficient representation in clinical studies is essential in evaluating interventions

and products across diverse populations, ensuring that they do not reinforce existing biases. In 2009, over 50% of registered studies failed to report race or ethnicity, or both; by 2024, this proportion had fallen to 17%. Early-phase trials, interventional trials, and studies that are conducted in the USA were most likely to report race and ethnicity. Among US studies, the frequency of race and ethnicity reporting differs between federally sponsored and industry-sponsored studies. However, NIH-funded studies were more likely to report both race and ethnicity.

Race is more commonly reported than ethnicity; however, in recent years, over half of registered studies report both race and ethnicity, and more than three-quarters report race alone. In 2024, most adult participants were classified as either White or Black. Given that ClinicalTrials.gov also includes international studies, the low and declining percentage of Asian and other racial groups is concerning, especially considering the large prevalence of Asian individuals globally. Racial categories like ‘Asian’ are descriptive of a diaspora of diverse people. For example, Asian is a racial group which can represent the populations of the two largest countries in the world (India and China). In the USA, Asian individuals represent approximately 7% of the country’s population.³⁴ While the geographic context and racial dynamics of the countries where the research is being conducted must be considered, the availability of multilingual recruitment and reporting resources is critical to diversity initiatives.³⁵ Across all study years, observational studies were consistently less likely than interventional trials to

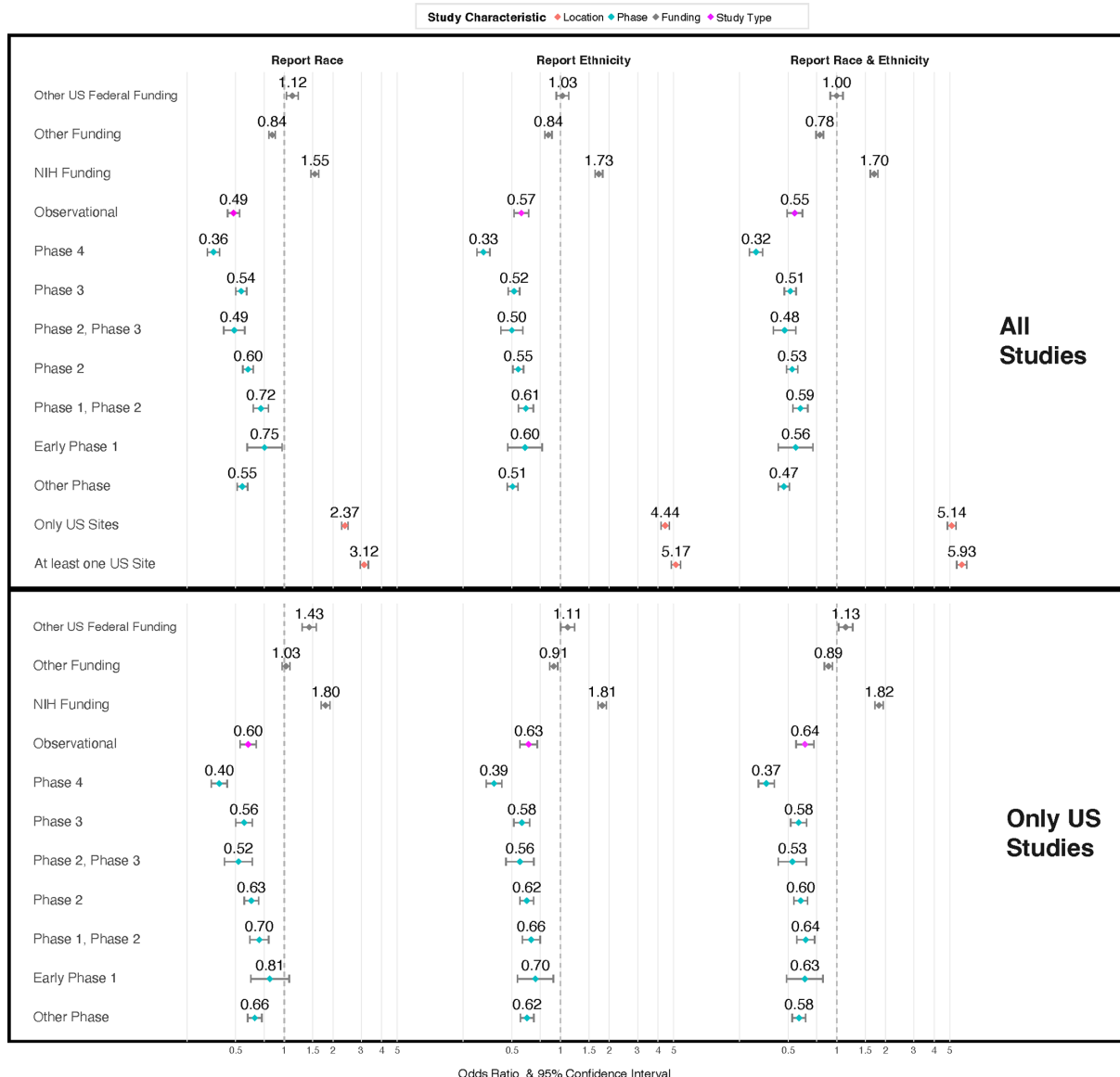


Figure 1 Characteristics of studies that report race and ethnicity to ClinicalTrials.gov, 2009–2024. ORs from multivariable logistic regression models. NIH, National Institute of Health.

report race, ethnicity or both. Additional work is needed to better understand why these studies are not reporting race and ethnicity data, as this information is available to study sponsors. Previous analyses have shown that trials delay reporting results to ClinicalTrials.gov, despite ethical and regulatory mandates.³⁶ In examples where both peer-reviewed literature and the registry are used, there is a large proportion of studies which have no available public results, despite concluding more than 7 years prior to evaluating reporting.³⁷ The largest study examining race and ethnicity of participants of studies registered to ClinicalTrials.gov examined only interventional studies.³⁸ Their findings on trial characteristics which are associated with reporting race and ethnicity largely matched ours, and they also found that the proportions of the race and ethnicity of the participants varied greatly when compared with the US Census data. Progress in

reporting and evaluating the representativeness of trials is still lacking collectively.

Although identifying the race and ethnicity of individuals who are participating in clinical research is important, identifying the societal mechanisms which cause health disparities is also important. Racism is the act (or presence) of structured inequality based on racial categories operating through interpersonal, institutional, and/or structural mechanisms.³⁹ Racism is one of the societal mechanisms which can affect health.⁴⁰ For example, examining racism from an ecosocial framework illustrates how discrimination can exert harmful effects across multiple sources and across lifecourse.⁴¹ It has been posited that the unequal treatment of individuals through racism is one of the primary drivers of health differences for individuals.^{42 43} Using research to improve clinical care requires including individuals from diverse

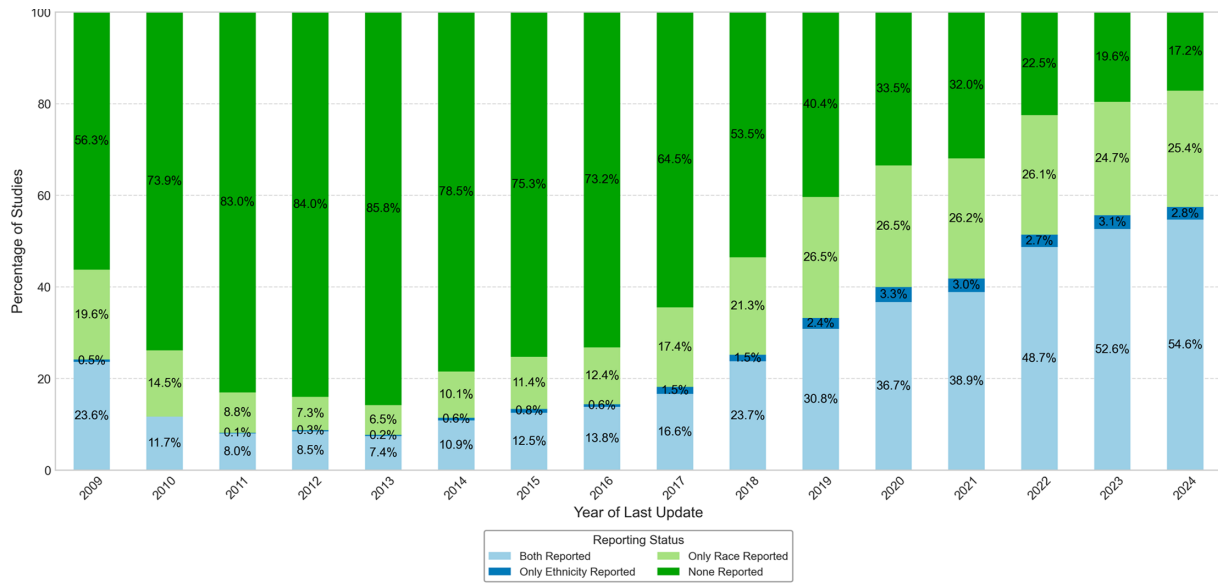


Figure 2 Proportion of studies reporting race/ethnicity over time.

social circumstances, but it also demands that interventional research extends beyond just inclusion to examine equity in clinical outcomes.

We did not stratify by intervention type or disease area, which limits our ability to evaluate specialty-specific reporting patterns. This represents an important area for future work, as prior studies using both ClinicalTrials.gov and peer-reviewed manuscripts have demonstrated variation in representation across specialties.^{44 45} Additionally, clinical trials should consider disease-specific disparities, ensuring that the composition of participants reflects the populations most affected by the condition. ClinicalTrials.gov includes more than just interventional trials, and the repository is based on the most recent updates provided by the study sponsors which can lead to some changes over time from various incentives and monitoring.⁴⁶ The

dynamic nature of the database raises the possibility of some variation in how data are recorded over time.

Limitations

Race and ethnicity definitions differ across countries. In this study, we used NIH/OMB categories, which are standard for reporting in the USA where ClinicalTrials.gov is hosted and these classifications were developed. However, we acknowledge that race and ethnicity are social constructs that depend on context, which may affect interpretation of our findings. We used a combination of keyword search and manual review to verify that the hundreds of customised race options matched the NIH/OMB options that were presented. While both keyword searches and human review are subject to error, we are encouraged by the

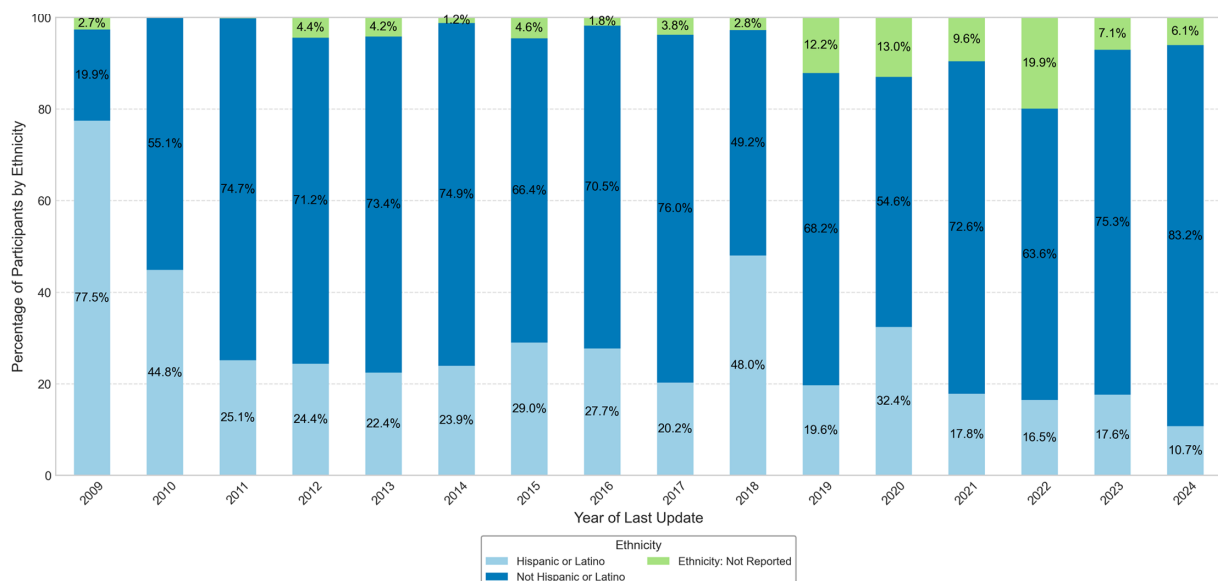


Figure 3 Distribution of ethnicity over time.

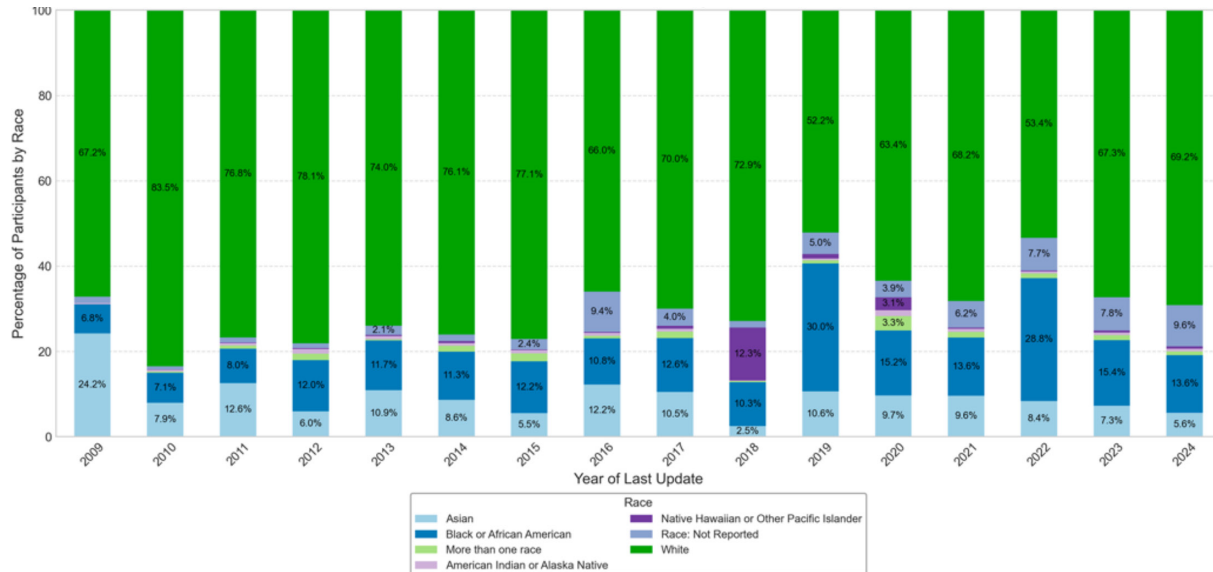


Figure 4 Distribution of race over time.

diminishing proportion of trials using customised methods for race and ethnicity categorisation over time. Terminated trials were excluded from our study, but there is relevant concern about the representation of the individuals within these trials as well.⁴⁷ There is increasing pharmaceutical research and development activity in low- and middle-income countries,⁴⁸ which have differing racial and ethnic populations alongside interpretations of race. Our process provided an overview from our US-centric perspective. We encourage other investigators to evaluate race, ethnicity, and other systems of social hierarchy resulting in inequities in different settings. Finally, while ClinicalTrials.gov is the most comprehensive public registry of trial data, it does not capture all relevant information. A previous analysis found approximately 19% of drug trials published in peer-reviewed journals were not reported to ClinicalTrials.gov.³⁷ There are valid concerns of missingness of race and ethnicity data and should not be considered the entire picture of clinical study representation.

Future directions

There are many initiatives and strategies aimed at increasing diversity in clinical trials.^{9 15 49-51} Advocacy groups, like the American Heart Association, have initiated a research network focused on the science of promoting diversity in clinical trials, with ongoing projects related to community engagement to address barriers to clinical trial participation in minority populations, including mistrust, lack of comfort, and lack of information.⁵²⁻⁵⁴ Decentralised clinical trials have also emerged as a mechanism for providing more opportunities for underrepresented racial and ethnic populations to participate in clinical research.⁵⁵ With the proliferation of technology and the internet, digital tools are also another promising venue for reaching more diverse populations,^{56 57} although

consistent internet access remains a barrier. Collectively, these approaches for randomised controlled trials, along with other study designs,^{58 59} should be implemented in parallel to increase representation in clinical research. This includes having clear targets and priorities for improving diversity and reporting their progress towards this goal.

Since they are socially constructed categories, the meanings of race and ethnicity as well as targets for diversity in trials will change over time. We conducted this analysis to assess the current state of racial and ethnic diversity in clinical research. Transparency in reporting clinical study findings remains essential. While progress in reporting race and ethnicity data to ClinicalTrials.gov may be a sign of understanding the value of this information, it is important that investigators also prioritise initiatives to consistently recruit and retain socio-demographically diverse participants.¹⁷ Further investigating the reporting of gender after the introduction of a new gender eligibility description registry in 2017⁶⁰ could provide another avenue for investigating sociodemographic diversity.

CONCLUSIONS

We found that reporting of both race and ethnicity has improved over time, with several study characteristics associated with whether sponsors reported both participant race and ethnicity to ClinicalTrials.gov. While there has been an improvement in reporting practices, more work is still needed. In 2024, 17.2% of studies still did not report any data on race and ethnicity, and 28.2% only reported either race or ethnicity. Additionally, ‘Not Hispanic or Latino’ and ‘White’ were the most common categories across time, with the breakdown of race and ethnicity categories not aligning with US Census data across time.

Each study team has the resources, guidance and ability to retain race and ethnicity data in some form. Emphasising diversity in clinical trials, reporting participant sociodemographic information and consistently implementing guidelines and policies around reporting are crucial for continued progress and improvement.

Acknowledgements We are appreciative of Annika Kruse for providing an early-stage evaluation of our idea and analysis.

Contributors MA: methodology, software, validation, investigation, data curation, writing—original draft, visualisation. ECO'B: methodology, supervision, writing—original draft, writing—review and editing. JBL: writing—review and editing. SK: writing—review and editing. VG: data curation, writing—review and editing. MAB: conceptualisation, writing—review and editing, supervision. RMC: conceptualisation, writing—review and editing, supervision. MDG: conceptualisation, methodology, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, supervision, project administration, funding acquisition. Guarantor of overall content is MDG.

Funding Michael D. Green was supported by the National Institute on Aging of the National Institutes of Health under Award Number F99AG088695. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests RMC served as the Commissioner for Food and Drug at FDA until January 2025. Prior to his appointment as FDA Commissioner, RMC was an employee of and held equity in Verily Life Sciences and Google Health (Alphabet). He also served on boards of directors for Cytokinetics, Centessa, Clinetic, Keystone Symposia, the Critical Path Institute (C-Path), the Clinical Research Forum and One Fifteen. All authors have no competing interests to report.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval As an analysis of secondary data at the trial level without reporting individual participant characteristics, institutional review board approval is not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. The datasets used are referenced in the manuscript alongside the API used.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Michael D Green <https://orcid.org/0000-0002-4982-8154>

REFERENCES

- Nair L, Asuzu P, Dagogo-Jack S. Ethnic Disparities in the Risk Factors, Morbidity, and Mortality of Cardiovascular Disease in People With Diabetes. *J Endocr Soc* 2024;8:bvae116.
- Osborn CY, de Groot M, Wagner JA. Racial and ethnic disparities in diabetes complications in the northeastern United States: the role of socioeconomic status. *J Natl Med Assoc* 2013;105:51–8.
- Islami F, Guerra CE, Minihan A, et al. American Cancer Society's report on the status of cancer disparities in the United States, 2021. *CA Cancer J Clin* 2022;72:112–43.
- Bonilla-Silva E. The Essential Social Fact of Race. *Am Sociol Rev* 1999;64:899–906.
- Graves J. *The Race Myth: Why We Pretend Race Exists in America*. Penguin, 2005.
- Lett E, Asabor E, Beltrán S, et al. Conceptualizing, Contextualizing, and Operationalizing Race in Quantitative Health Sciences Research. *Ann Fam Med* 2022;20:157–63.
- Fain KM, Nelson JT, Tse T, et al. Race and ethnicity reporting for clinical trials in ClinicalTrials.gov and publications. *Contemp Clin Trials* 2021;101:106237.
- Gedela K, Wong R, Balendra S, et al. Embedding equity, diversity and inclusion processes within clinical trials and health and social care research. *BMJ Open* 2025;15:e091807.
- Byiringiro S, Garcia JK, Farrell N, et al. Advocating for collaboration among key partners to promote diversity in clinical studies amid policy challenges in the United States of America. *Trials* 2025;26:117.
- Lewsey SC, Breathett K. Racial and ethnic disparities in heart failure: current state and future directions. *Curr Opin Cardiol* 2021;36:320–8.
- Green MD, Mentz RJ, Greene SJ, et al. Cardiovascular Multimorbidity in Older Adults in the United States by Race and Sex. *JACC Adv* 2025;4:102103.
- Varma T, Gross CP, Miller JE. Clinical Trial Diversity—Will We Know It When We See It? *JAMA Oncol* 2023;9:765–7.
- Knepper TC, McLeod HL. *When Will Clinical Trials Finally Reflect Diversity*. Nature Publishing Group UK London, 2018.
- Schwartz AL, Alsan M, Morris AA, et al. Why Diverse Clinical Trial Participation Matters. *N Engl J Med* 2023;388:1252–4.
- Brangman SA. Achieving diversity in study populations: The importance of community engagement. *J Am Geriatr Soc* 2022;70:3080–6.
- Hwang TJ, Brawley OW. New Federal Incentives for Diversity in Clinical Trials. *N Engl J Med* 2022;387:1347–9.
- VersavelS, Subasinghe A, JohnsonK, et al. Diversity, equity, and inclusion in clinical trials: A practical guide from the perspective of a trial sponsor. *Contemp Clin Trials* 2025.
- Chen MS Jr, Lara PN, Dang JHT, et al. Twenty years post-NIH Revitalization Act: Enhancing minority participation in clinical trials (EMPaCT): Laying the groundwork for improving minority clinical trial accrual. *Cancer* 2014;120:1091–6.
- ib HS, United States F, et al. Guidance for Industry Center for Drug Evaluation and Research. Enhancing the diversity of clinical trial populations: eligibility criteria, enrollment practices, and trial designs. 2020.
- Sanyi A, Byiringiro S, Dabiri S, et al. Measuring Representativeness in Clinical Trials. *Circulation* 2025;151:318–30.
- Buffenstein I, Kaneakua B, Taylor E, et al. Demographic recruitment bias of adults in United States randomized clinical trials by disease categories between 2008 to 2019: a systematic review and meta-analysis. *Sci Rep* 2023;13:42.
- Loree JM, Anand S, Dasari A, et al. Disparity of Race Reporting and Representation in Clinical Trials Leading to Cancer Drug Approvals From 2008 to 2018. *JAMA Oncol* 2019;5:e191870.
- Allison K, Patel DG, Greene L. Racial and Ethnic Disparities in Primary Open-Angle Glaucoma Clinical Trials: A Systematic Review and Meta-analysis. *JAMA Netw Open* 2021;4:e218348.
- Tahhan AS, Vaduganathan M, Greene SJ, et al. Enrollment of Older Patients, Women, and Racial/Ethnic Minority Groups in Contemporary Acute Coronary Syndrome Clinical Trials: A Systematic Review. *JAMA Cardiol* 2020;5:714–22.
- Califf RM, Cutler TL, Marston HD, et al. The importance of ClinicalTrials.gov in informing trial design, conduct, and results. *J Clin Transl Sci* 2025;9:e42.
- National Library of Medicine. Protocol registration data element definitions for interventional and observational studies. Available: <https://clinicaltrials.gov/policy/protocol-definitions#study-status> [Accessed 7 Jun 2025].
- National Library of Medicine. FDAAA 801 and the final rule. 2025 Aggregate Analysis of ClinicalTrials.gov (AACT) Database. *Clinical Trials Transformation Initiative (CTTI)* 2025. Available: <https://aact.ctti-clinicaltrials.org/>
- National Institute of Health. Racial and ethnic categories and definitions for nih diversity programs and for other reporting purposes. Available: <https://grants.nih.gov/grants/guide/notice-files/not-od-15-089.html> [Accessed 6 Jun 2025].
- Hirsch BR, Califf RM, Cheng SK, et al. Characteristics of oncology clinical trials: insights from a systematic analysis of ClinicalTrials.gov. *JAMA Intern Med* 2013;173:972–9.
- Goswami ND, Pfeiffer CD, Horton JR, et al. The state of infectious diseases clinical trials: a systematic review of ClinicalTrials.gov. *PLoS One* 2013;8:e77086.

- 32 Califf RM, Zarin DA, Kramer JM, *et al.* Characteristics of clinical trials registered in ClinicalTrials.gov, 2007-2010. *JAMA* 2012;307:1838-47.
- 33 National Library of Medicine. Study data structure. 2025 Available: <https://clinicaltrials.gov/data-api/about-api/study-data-structure> Jun7
- 34 Krogstad JM, Im C. Key Facts about Asians in the US Pew Research Center. *May* 2025;1. Available: www.pewresearch.org/short-reads/2025/05/01/key-facts-about-asians-in-the-us/
- 35 Willis A, Isaacs T, Khunti K. Improving diversity in research and trial participation: the challenges of language. *Lancet Public Health* 2021;6:e445-6.
- 36 Anderson ML, Chiswell K, Peterson ED, *et al.* Compliance with results reporting at ClinicalTrials.gov. *N Engl J Med* 2015;372:1031-9.
- 37 Fain KM, Rajakannan T, Tse T, *et al.* Results Reporting for Trials With the Same Sponsor, Drug, and Condition in ClinicalTrials.gov and Peer-Reviewed Publications. *JAMA Intern Med* 2018;178:990-2.
- 38 Turner BE, Steinberg JR, Weeks BT, *et al.* Race/ethnicity reporting and representation in US clinical trials: a cohort study. *Lancet Reg Health Am* 2022;11:100252.
- 39 Bonilla-Silva E. Rethinking Racism: Toward a Structural Interpretation. *Am Sociol Rev* 1997;62:465.
- 40 Bailey ZD, Krieger N, Agénor M, *et al.* Structural racism and health inequities in the USA: evidence and interventions. *The Lancet* 2017;389:1453-63.
- 41 Krieger N. Methods for the scientific study of discrimination and health: an ecosocial approach. *Am J Public Health* 2012;102:936-44.
- 42 Graves Jr JL, Goodman AH. *Racism, Not Race: Answers to Frequently Asked Questions*. Columbia University Press, 2021.
- 43 Williams DR, Lawrence JA, Davis BA. Racism and Health: Evidence and Needed Research. *Annu Rev Public Health* 2019;40:105-25.
- 44 Rees CA, Stewart AM, Mehta S, *et al.* Reporting of Participant Race and Ethnicity in Published US Pediatric Clinical Trials From 2011 to 2020. *JAMA Pediatr* 2022;176:e220142.
- 45 Lee LK, Narang C, Rees CA, *et al.* Reporting and Representation of Participant Race and Ethnicity in National Institutes of Health-Funded Pediatric Clinical Trials. *JAMA Netw Open* 2023;6:e2331316.
- 46 Tse T, Fain KM, Zarin DA. How to avoid common problems when using ClinicalTrials.gov in research: 10 issues to consider. *BMJ* 2018;361:k1452.
- 47 Pak TR, Rodriguez MD, Roth FP. 2015 Why clinical trials are terminated. *bioRxiv*.
- 48 Vieira M, Andia T, Karim O, *et al.* Rising pharmaceutical innovation in the Global South: a landscape study. *J Pharm Policy Pract* 2023;16:155.
- 49 Igwe J-K, Wangdak Yuthok TY, Cruz E, *et al.* Opportunities to Increase Science of Diversity and Inclusion in Clinical Trials: Equity and a Lack of a Control. *J Am Heart Assoc* 2023;12:e030042.
- 50 Woodcock J, Araujo R, Thompson T, *et al.* Integrating Research into Community Practice - Toward Increased Diversity in Clinical Trials. *N Engl J Med* 2021;385:1351-3.
- 51 Bibbins-Domingo K, Helman A. Improving representation in clinical trials and research. 2025.
- 52 Patel N, Whitman C, Lieberman A, *et al.* A series of randomized trials of behavioral economic interventions to increase racial and ethnic diversity of research participants: Rationale and design of ITERATE. *Am Heart J* 2025;286:80-7.
- 53 Dabiri S, Raman R, Grooms J, *et al.* Examining the Role of Community Engagement in Enhancing the Participation of Racial and Ethnic Minoritized Communities in Alzheimer's Disease Clinical Trials: A Rapid Review. *J Prev Alzheimers Dis* 2024;11:1647-72.
- 54 Reopell L, Nolan TS, Gray DM, *et al.* Community engagement and clinical trial diversity: Navigating barriers and co-designing solutions—A report from the “Health Equity through Diversity” seminar series. *PLoS One* 2023;18:e0281940.
- 55 van Rijssel TI, van Delden JJM, Lagerwaard B, *et al.* Diversity in decentralized clinical trials: prioritizing inclusion of underrepresented groups. *BMC Med Ethics* 2025;26:51.
- 56 Tomiwa T, Wong E, Miller HN, *et al.* Leveraging digital tools to enhance diversity and inclusion in clinical trial recruitment. *Front Public Health* 2024;12:1483367.
- 57 Tan RKJ, Wu D, Day S, *et al.* Digital approaches to enhancing community engagement in clinical trials. *NPJ Digit Med* 2022;5:37.
- 58 Evans SR, Paraoan D, Perlmutter J, *et al.* Real-World Data for Planning Eligibility Criteria and Enhancing Recruitment: Recommendations from the Clinical Trials Transformation Initiative. *Ther Innov Regul Sci* 2021;55:545-52.
- 59 Propes C, Mehl K, Morain SR. Untapped Potential? Representativeness in Pragmatic Clinical Trials. *JAMA* 2025;333:375-6.
- 60 Burton A, Mohan S, Puplampu-Dove Y, *et al.* Characterizing Gender Eligibility Descriptions for Clinical Trials Registered on ClinicalTrials.gov. *JAMA* 2023;330:2019-21.