

Current State and Demographic Trends of Medically Underserved Populations in Rare Disease Research in the United States

Lavanyaa Manjunatha¹; Saundarya MS¹; Deepika Dokuru¹; Nisha Venugopal¹; Jenifer Ngo
Waldrop⁵; Linda Goler Blount⁵; Reena V Kartha^{1,4}; Harsha K Rajasimha^{*1,2,3}

¹Indo US Organization for Rare Diseases (IndoUSrare), Herndon, VA, USA.

²Jeeva Clinical Trials, Inc., Manassas, VA, USA.

³School of Systems Biology, George Mason University, Fairfax, VA, USA.

⁴Center for Orphan Drug Research, Department of Experimental and Clinical
Pharmacology, University of Minnesota Twin Cities, Minneapolis, MN, USA.

⁵Rare Disease Diversity Coalition, Black Women's Health Imperative, Atlanta, GA, USA.

***Corresponding author:** Harsha K Rajasimha, MS, PhD, harshakarur@gmail.com

Abstract

Background

Addressing inequities and health disparities for medically underserved populations (MUPs) is critical, as they already face systemic bias and barriers, such as historical mistrust of healthcare systems. To achieve health equity, we need systematic approaches to measuring, quantifying, and reporting diversity, equity, inclusion, and accessibility (DEIA) metrics.

22 **Methods**

23 The objective of the study was to analyze literature and clinical trials to summarize the current
 24 state of demographics and socioeconomic factors (SF) reporting for MUPs in US-based RD
 25 clinical research. PubMed, Cochrane Library, and ClinicalTrials.gov were searched (1983–
 26 2023). A universal set of 30,303 unique RD terms from the Genetic and Rare Diseases
 27 Information Center, Orphanet, Rare-X, and ClinicalTrials.gov was used to filter publications
 28 and clinical trials. Publications that reported demographics or SFs, were US-based, and
 29 involved one or more RDs were included for analysis. Clinical trials that were US-based,
 30 involved an RD, and had study results posted were also included. Age, sex or gender, race,
 31 ethnicity, and SF data were extracted and analyzed using descriptive statistics. Race and
 32 ethnicity data were compared with the US census. The representation of MUPs in RD clinical
 33 research was assessed based on the frequency of publications and clinical trials reporting 13
 34 variables.

35 **Results**

36 We reviewed 234 publications and 8475 RD clinical trials. Age was the most reported
 37 demographic variable (publications: 94%; clinical trials: 100%), followed by sex or gender
 38 (86.3%; 100%). Race (50%; 45.7%) and ethnicity (29.9%; 38.5%) were less frequently
 39 reported and often in a variable format in publications compared with the ClinicalTrials.gov
 40 database. At least one SF was reported in 15.8% of the publications and 0.2% of the trials.
 41 American Indian or Native Alaskan, Asian, Hispanic, and Latino participants were
 42 significantly underrepresented compared with the US census averages. Data were largely

absent for other MUPs: lesbian, gay, bisexual, transgender, and queer or questioning individuals, rural residents, veterans, immigrants, and those affected by disability and poverty.

Conclusions

Significant gaps exist in demographics and SF reporting in RD clinical research, and several MUPs are underrepresented. Therefore, a framework to enhance DEIA in RD research is urgently needed.

Keywords

rare disease, clinical trials, medically underserved populations, demographics, socioeconomic factors, diversity, equity, inclusion, accessibility, MUPs

Background

Improving representation in clinical trials has been a major focus of United States (US) legislators and the Food and Drug Administration (FDA) in recent years (1–4). The FDA’s latest Diversity Action Plan (DAP) aims to enhance diversity in clinical trials, including in rare disease (RD) trials (4). Over 10,000 RDs collectively affect 30 million Americans (5,6), with a \$997 billion economic burden (7), making them a public health concern. Marking the 40th anniversary of the Orphan Drug Act in 2023, the approval of >880 treatments for RDs in this period indicate progress yet highlight unmet needs (8,9). The FDA has several programs to accelerate RD research, including the Rare Disease Innovation Hub and the Support for Clinical Trials Advancing Rare Disease Therapeutics Pilot Program (10–14). For these programs to be effective, accurate assessment and tracking of the current status of diversity, equity, inclusion, and accessibility (DEIA, Box 1) in US-based RD research is necessary.

65 The FDA’s Drug Trials Snapshot (DTS, 2015–2019) revealed underrepresentation in global
66 RD trials, with 11% Asian, 9% Black or African American, 1% American Indian or Alaska
67 Native, 6% Hispanic or Latino, and 34% missing ethnicity data (15). In cystic fibrosis (CF)
68 trials, Black or African American, Asian, and Hispanic or Latino participants were included at
69 lower rates than in the US CF Foundation registry (16). Similar underrepresentation has been
70 observed in US trials of muscular dystrophy, hemophilia, and acute leukemia (17–19),
71 indicating underrepresentation is prevalent in RD research.

72 This persistent underrepresentation could be explained by common RD features: low
73 prevalence, varying prevalence in medically underserved populations (MUPs), complex
74 natural history, dispersed geographic patient distribution, diagnostic difficulties, patient
75 registry reliance, and lack of awareness among care providers and patients (20). Given the
76 small trial sizes of RDs (21), defining what DEIA entails in these contexts is vital. Although
77 the US census serves as a representation benchmark (22,23), trial diversity should also be
78 assessed based on prevalence and epidemiology (24). Using estimates in the absence of
79 epidemiology data (25) adds another layer of complexity to ensure the inclusion of all
80 affected patients in RDs.

81 Demographic and socioeconomic factors (SFs) disparities are evident in RD research. Only
82 20% of RD trials had child participants (26), despite 50–70% of RDs affecting them (27,28).
83 Female patients with RDs, often dismissed by healthcare providers (29), face delayed
84 diagnosis and lower health-related quality of life than male patients (30,31). Disparities in
85 diagnostic delays and greater disease severity are apparent for Black or African American
86 patients with amyotrophic lateral sclerosis (32). Hispanic patients with CF had lower
87 pulmonary function than non-Hispanic patients with CF (33). Further, SFs impact the
88 mortality and survival of patients with pulmonary arterial hypertension (34). Delayed
89 diagnosis of RDs results in direct and indirect costs of \$517K per patient (35). These data

underscore the urgent need to enhance DEIA in RDs, emphasizing the moral, economic, and therapeutic imperatives for diverse representation. Understanding the current landscape of MUP participation is essential to advancing DEIA in RD research. Except for the FDA’s DTS report, which provides some data about representation, no other studies have reported such data (36). This study aimed to quantify the reporting of demographics and SFs in RD research to enhance DEIA for MUPs through evidence-based frameworks. We report the findings from a detailed literature assessment (2010–2023) and an analysis of US-based clinical trials (1983–2023) registered on ClinicalTrials.gov.

Box 1

Definitions

(a) **diversity^a** means the practice of including the many communities, identities, races, ethnicities, backgrounds, abilities, cultures, and beliefs of the American people, including underserved communities.

(b) **equity^{a,b}** means the consistent and systematic fair, just, and impartial treatment of all individuals, including individuals who belong to underserved communities that have been denied such treatment.

(c) **inclusion^a** means the recognition, appreciation, and use of the talents and skills of employees of all backgrounds.

(d) **accessibility^a** means the design, construction, development, and maintenance of facilities, information and communication technology, programs, and services so that all people, including people with disabilities, can fully and independently use them. Accessibility includes the provision of accommodations and modifications to ensure equal access to employment and participation in activities for people with disabilities, the reduction or elimination of physical and attitudinal barriers to equitable opportunities, a commitment to ensuring that people with disabilities can independently access every outward-facing and internal activity or electronic space, and the pursuit of best practices such as universal design.

(e) **medically underserved populations (MUPs)^c** are populations facing a shortage of primary healthcare services for a specific population subset within a geographic area. These groups may face economic, cultural, or language barriers to healthcare. MUPs could include people experiencing homelessness, low-income people, people eligible for Medicaid, Native Americans, and migrant farm workers.

(d) **medically underserved areas (MUAs)** have a shortage of primary health services within geographic areas such as a whole county, a group of neighboring counties, a group of urban census tracts, a group of county or civil divisions

(f) **underserved communities^a** refer to populations sharing a particular characteristic, as well as geographic communities, who have been systematically denied a full opportunity to participate in aspects of economic, social, and civic life. In the context of the Federal workforce, this term includes individuals who belong to communities of color, such as Black and African American, Hispanic and Latino, Native American, Alaska Native and Indigenous, Asian American, Native Hawaiian and Pacific Islander, Middle Eastern, and North African persons. It also includes individuals who belong to communities that face discrimination based on sex, sexual orientation, and gender identity (including lesbian, gay, bisexual, transgender, queer, gender non-conforming, and non-binary (LGBTQ+) persons); persons who face discrimination based on pregnancy or pregnancy-related conditions; parents; and caregivers. It also includes individuals who belong to communities that face discrimination based on their religion or disability; first-generation professionals or first-generation college students; individuals with limited English proficiency; immigrants; individuals who

belong to communities that may face employment barriers based on older age or former incarceration; persons who live in rural areas; veterans and military spouses; and persons otherwise adversely affected by persistent poverty, discrimination, or inequality. Individuals may belong to more than one underserved community and face intersecting barriers.

^a Executive order 14,035 on diversity, equity, inclusion, and accessibility in the federal workforce.

^b Executive order 13,985 on advancing racial equity and support for underserved communities through the federal government.

^c Health Resource and Services Administration

Methods

The study assessed 13 demographic and SF variables in RD research by reviewing the literature and analyzing the ClinicalTrials.gov database; methods are summarized below and detailed in the Supplementary Information. This study followed the Standards for Quality Improvement Reporting Excellence and Strengthening the Reporting of Observational Studies in Epidemiology guidelines (37,38).

Literature assessment

PubMed and the Cochrane Library were searched using specific terms (Additional file 1: Table S1). Screening and filtering details are in Additional file 1: Supplementary Methods 2, Figure S1A, and Table S2. Publications were included if they reported demographics or SFs, involved one or more RD conditions, and were conducted in the US. Qualitative and quantitative data were extracted from publications satisfying our defined inclusion criteria.

ClinicalTrials.gov database extraction and analysis

All clinical trials data (8 December 2023) and study information for US-based trials were downloaded from ClinicalTrials.gov (Filters: Location=USA, Study results=With Study results, Date range=01/01/1983–08/12/2023) (39). RD trials were identified by matching the “condition” section of clinical trials with a comprehensive list of RD terms generated from the Genetic and Rare Diseases Information Center (GARD), Orphanet, Rare-X, and the ClinicalTrials.gov classical website (find studies by topic/RDs) (40–43). See Supplementary

Information for details. Trials with SFs were identified using keywords (Additional file 1: Table S3).

Study Outcomes

The primary outcomes were representation of MUPs (historically underserved racial and ethnic groups, multiracial groups, lesbian, gay, bisexual, transgender, queer or questioning (LGBTQ+) individuals, caregivers, rural residents, veterans, religious minorities, immigrants, people living with disability, and those affected by poverty). We also gathered the frequency of reporting in publications and clinical trials of the variables: age, sex, gender, sexual orientation, race, ethnicity, and SFs (language, education, employment, income, residence location, health insurance, relationship status, social support, and disability allowance).

Statistical Analysis

Descriptive statistics were used to summarize the publication or trial characteristics, demographics, and SFs. Racial and ethnic proportions from trials reporting them in the National Institute of Health (NIH)/Office of Management and Budget (OMB) format ((44) were compared with the weighted average percentages of US census data (1990–2022). The weights were based on trial participant distribution in the census bins. These weighted average percentages are referred to as US census averages from this point forward. Chi-squared tests assessed significant differences in race and ethnicity distributions between the trials and the census. Bonferroni correction was used to adjust the significance threshold for multiple comparisons in the post hoc analyses. All statistical analyses were performed using Python 3.

Results

Literature assessment

Of the 16,187 selected publications, 1279 underwent full-text reviews to assess inclusion criteria. The analysis included 234 publications (56% US-only; 44% multinational; Additional file 1: Figure S1A). Table 1 presents the publication characteristics, participant type, demographics, and SFs reported in RD publications.

[Insert Table 1 here]

Demographic and SFs reporting in RD publications

The most reported demographic variable was age (94%), and the least reported was ethnicity (29.9%, Table 1). The most reported age category was “child and adult” (28.2%), followed by “adult and older adult” (25.5%), “child only” (14.2%), and prenatal (1.4%); “older adults” (0.9%) was the least reported category. Gender was reported by 86.3% of all publications; 66.8% of them reported it in the binary format. Only 5% of them reported a third option for gender as “other,” “non-binary,” “missing,” “not reported,” or “prefer not to say.” Race was reported in 50% of publications, and ethnicity in 29.9%; neither followed the standardized NIH/OMB format. Some publications used the term “Caucasian” for White participants, “non-White” for all other race participants, categorized race as ethnicity, reported data for some race categories, or combined data for two race categories (for example, Asian and Native Hawaiian or Other Pacific Islander were combined). In other publications, terms like “origin,” “ancestry,” or “country” were used to report race and ethnicity categories (Additional file 1: Table S4). Among the 117 publications reporting race, 96.6% included White participants, 65% Black or African American, 55.6% Asian, 21.4% American Indian or Alaska Native, 17.9% Native Hawaiian or Other Pacific Islander, and 20.5% more than one

race. Hispanic and Latino participants were reported in 80% of the publications, which included ethnicity data (Table 1). Only 37 of 234 (15.8%) publications reported one or more SFs; 75.7% of these were US-only publications. Education was the most reported SF (62.2%), and the least reported was social support (2.7%, Table 1). No standard reporting formats were applied; for example, education was reported in terms of the number of years or a categorical manner (“Bachelor,” “college,” “some college,” “high school”). Participants affected by RDs had varied educational backgrounds, insurance coverages, income levels, employment status, and relationship status, indicating potential inclusion and accessibility issues (Additional file 1: Table S5). English was the most reported language and was an inclusion/exclusion criterion in most publications reporting language data.

ClinicalTrials.gov data analysis

From 44,064 US-based trials with study results, we identified 8,475 RD trials (1,492,423 participants; Figure S1B); 69.9% were US-only, and 30.1% were multinational trials (Table 2). 74% of the trials had a 100% keyword–condition match, and <5% of the trials were non-specific matches (Additional file 1: Table S6). Table 2 and Additional file 1: Table S7 summarize the trial characteristics, demographics reporting, and participant distributions. [Insert Table 2 here]

Reporting of demographics and SFs in RD trials

Most US-based RD trials included “adult and older adult participants” (68.8%) and both male and female participants (85.6%). Trials with only children and older adults accounted for 5.8% and 0.5%, respectively. Only 0.3% of the trials included non-binary sex or genders and reported them as unknown. Race was reported in 45.7% of the trials, ethnicity in 38.5% in the NIH/OMB format, while 9.1% used a customized format for reporting race and ethnicity. SFs

were reported in 16 of 8475 (0.2%) trials; relationship status was most reported (68.8%), followed by education (50%), insurance status (37.5%), and the least reported was religion (6.3%, Additional file 1: Table S8 and S9).

Race and ethnicity in US-only RD trials

Participant-level analysis: US-only trials (1991–2022) reported race in 2,799 trials and ethnicity in 2,204 trials in the NIH/OMB format. Figure 1A summarizes racial and ethnic representation levels by presenting aggregate or overall proportions and mean proportions (the average of proportions from all trials) of participants. These aggregate and mean proportions significantly differed from the US census averages (race: $\chi^2=5477.9$ and 5500.9 ; ethnicity: $\chi^2=15491077.3$ and 13595347.8 , $p < 0.001$).

Both aggregate and mean proportions showed that the representation of American Indian or Alaska Native, Asian, and Hispanic or Latino participants was significantly below the US census averages; White representation was significantly above the US census average (Figure 1A, Additional file 1: Table S10). While aggregate proportions of Black or African American and Native Hawaiian or Pacific Islander participants were significantly above the US census averages, the mean proportion for Black or African American participants alone revealed the opposite trend. Across aggregate and mean proportions, 1–1.2% of the participants were categorized as more than one race, and 4.3–4.7% of the participants were “race unknown or not reported”; both were significantly below the US census averages. About 4.5–5.1% of the participants were “ethnicity unknown or not reported” with no corresponding category in the US census. (Figure 1A, Additional file 1: Table S10).

Trial-level analysis: We analyzed the trial-level diversity by assessing the percentage of trials having racial and ethnic representation above or below the US census average (Figure 1B). Most trials lacked representation for Native Hawaiian or Pacific Islander participants (92.4%),

followed by American Indian or Alaska Native (87.9%), Asian (55.1%), Black or African American (34.8%), and Hispanic or Latino (35%) participants. In contrast, only 3.9% of the trials did not include White participants.

Study phase and funder type analysis: Across study phases and funder types, aggregate and mean proportions for race and ethnicity significantly differed from the US census averages (Figure 2, Additional file 1: Table S11). Only Black or African American, and White representation significantly exceeded the US census averages in at least one study phase or funder type (Figure 2, Additional file 2: Table S12). White representation significantly exceeded the US census average levels across funder types and study Phases 1–4. Conversely, Black or African American representation significantly exceeded US census average levels only in NIH-funded and Phase 1, 3, and 4 trials. Hispanic or Latino representation was lower than the US census average across study phases and funder type (Figure 2). The highest representation of Hispanic or Latino participants was in industry-funded trials (10.6–11.4%) and Phase 2|Phase 3 trials (12.1–15.5%).

Temporal trends: The number of RD clinical trials reporting race and ethnicity began increasing in 1999 (race: $N_1=8$, ethnicity: $N_2=8$) and peaked in 2015 for race ($N_1=282$) and in 2016 for ethnicity ($N_2=230$, Additional file 1: Figure S2A). We analyzed participant distribution over three decades (1991–2020) to better understand representation trends of the different racial and ethnic groups (Figure 3A–F, Additional file 1: Figure S3A–D). American Indian or Alaska Native, Asian, and Hispanic or Latino participation consistently remained below the US census. In contrast, Black or African American, and Native Hawaiian participation consistently remained higher than the US census for those three decades. White participation exceeded the US census for the decades between 2001–2020 (Figure 3).

Race and ethnicity in multinational trials

Multinational trials (1997–2022) reported race in 1051 trials and ethnicity in 1040 trials.

Additional file 2: Table S13 presents the overall racial and ethnic distribution across funder types and study phases.

Discussion

Our results reveal key gaps in demographics and SF reporting practices in RD research, emphasizing the urgent need to include MUPs. Consistent with previous studies, age and sex or gender were well reported in our analysis (45,46). There was no reporting of sexual orientation and gender identity (SOGI) data. We note that the FDA has not included SOGI in the DAP 2024 (4), whereas a Health Resources and Services Administration mandate existed from 2016 (47).

Race and ethnicity were underreported despite the Final Rule for Clinical Trials 2017(48). A previous study also reported that only 43% of US-based trials had race/ethnicity data (23). Only 14.5% of the publications included child participants, compared to 5.8% of the trials (Table 1). The presence of genomics/genetics studies (12.4%) in our literature analysis, which often includes child participants, may explain this discrepancy. An earlier report showed that 20% of global RD trials had child participants (26), contrasting with our analysis of US-based trials (5.8%). Older adult participants were underrepresented in publications (0.9%) and clinical trials (0.5%). Although a direct comparison with the FDA’s DTS (2015–2019) RD clinical trial data (>65 years of age=24%) is not possible (15), the underrepresentation of older adults is evident, considering they comprise 17% of the US population (49). However, we note that older adult participants were included in other age categories: “adult and older adult” or “child, adult and older adult.”

Most publications did not report race and ethnicity in the NIH/OMB format, making it challenging to draw generalizable insights on the racial and ethnic representation in RD

research. Our analyses of aggregate and mean proportions of race and ethnicity highlight the persistent underrepresentation of Asian, Native Hawaiian or Other Pacific Islander, American Indian or Alaska Native, and Hispanic or Latino participants (Figure 2), consistent with other studies on US-based trials (22,23). Despite significantly higher aggregate proportions than the US census, the lower mean proportions (Figure 1A) and the percentage of trials lacking Black or African American (34.8%) and Native Hawaiian or Other Pacific Islander (92.4%) participants (Figure 1B) indicate a significant gap in participant recruitment. Hispanic or Latino participation was numerically higher in our study (9.1%) than in the FDA DTS report (6%). Conversely, Asian participation was numerically lower in our study (3.2%) than in the FDA DTS (9%), likely because we focused on US-only trials, whereas FDA DTS included global RD trials.¹⁵ In multinational trials, both Asian (10.4%) and Hispanic or Latino (12.1%) participation were numerically higher in our study (Additional file 2: Table S13) compared with the FDA DTS. Both the FDA DTS and our multinational studies included Asians at lower rates compared with their global representation of 59.1% (50). Although race and ethnicity reporting improved over time (Additional file 1: Figure S2, Figure S3B and S3D), the representation of American Indian, Asian, and Hispanic or Latino participants remained below the US census levels (Figure 3A, 3B, and 3F). The trend of overrepresentation of Black or African American, Native Hawaiian or Other Pacific Islanders could indicate uneven recruitment efforts. Although the percentage of people reporting multiple races increased to 10.2% in 2010, multiracial participation remained lower in RD trials in the period 2011–2020 and thereafter (Additional file 1: Figure S3A). The decline in trials reporting race and ethnicity data from 2018 onwards, consistent with a previous study (23), is likely due to our focus on trials with study results. However, whether the COVID-19 pandemic influenced the number of completed trials with study results is unclear.

The underrepresentation of racial and ethnic MUPs observed in this study could be due to barriers in accessing diagnostics and healthcare, limited awareness of trials, and resource constraints (51–53). Moreover, the historical mistrust in healthcare, stemming from events like the Tuskegee Syphilis Study (54), also contributes to lower participation of certain racial/ethnic groups (55–57). Implicit and explicit bias among healthcare providers results in certain racial and ethnic groups not being invited to participate in clinical research (58). Language barriers also hinder the participation of speakers with limited English proficiency (59,60). In our analysis, we found that English proficiency was used as an inclusion/exclusion criterion in studies reporting language data.

SFs were massively underreported, consistent with a previous study evaluating such reporting in high-impact factor medical journals (61). The discrepancy in SFs reporting between publications (15.8%) and clinical trials (0.2%, Table 1) is likely due to the focus on clinical and biomedical data rather than SFs in ClinicalTrials.gov. A recent report showed that SFs in US clinical research sites are the least collected data (62). The limited or lack of data on veterans, immigrants, religious minorities, rural residents, and those affected by poverty is likely because these are not federally mandated data elements. Additionally, privacy and discrimination concerns could also contribute to the barriers to collecting data on veterans and immigrants.

Limitations

This study has limitations. The literature search strategy used “RDs” or “rare disorders” in all fields and Medical Subject Headers (MeSH) terms. However, not all RD publications are annotated with these MeSH terms (63), causing dataset gaps. Most publications did not report race/ethnicity in a standardized format, and 30–50% omitted these data. Only publicly

available trials with study results in ClinicalTrials.gov were analyzed, and 1005 trials with customized race/ethnicity were excluded. Racial and ethnic representation in trials were compared with the US census, which is unsuitable for individual RDs with differing prevalence across races and ethnicities. Ideally, representation should also be evaluated based on the epidemiology of the disease (24), consistent with FDA guidance (4). However, as our analysis included all RD trials, the US census was a practical choice for comparisons.

Conclusions

This study is the first comprehensive report quantifying DEIA variables in US RD clinical research. We identified reporting gaps for race, ethnicity, and SFs in publications and clinical trials. There is a lack of data on several MUPs, coupled with the underrepresentation of racial and ethnic minority groups. Our observations support the urgent need for data-driven recommendations to improve measurement, reporting, and a national framework to enhance DEIA in RD research. Our study could serve as a template for future global studies evaluating DEIA variables in not just RD research but all clinical research.

Abbreviations

US: United States
 FDA: Food and Drug Administration
 DAP: Diversity Action Plan
 RD: Rare Disease
 DEIA: Diversity, Equity, Inclusion, Accessibility
 DTS: Drug Trial Snapshot Summary
 CF: Cystic Fibrosis

329 MUPs: Medically underserved populations
 330 SFs: Socioeconomic Factors
 331 MUAs: Medically underserved areas
 332 USA: United States of America
 333 GARD: Genetic and Rare Disease Information Centre
 334 LGBTQ+: lesbian, gay, bisexual, transgender, queer or questioning
 335 NIH: National Institute of Health
 336 OMB: Office of Management and Budget
 337 SOGI: Sexual Orientation and Gender Identity
 338 MeSH: Medical Subject Heading

339 **References**

340 1. U.S. Congress. Rep. H.R.7667 - Food and Drug Amendments of 2022. 117th
 341 Congress (2021-2022). 2022 [cited 2024 Dec 18]; Available from:
 342 <https://www.congress.gov/bill/117th-congress/house-bill/7667>
 343 2. U.S. Congress. Text - H.R.3503 - NIH Clinical Trial Diversity Act of 2023.
 344 118th Congress (2023-2024). 2023 [cited 2024 Dec 18]; Available from:
 345 <https://www.congress.gov/bill/118th-congress/house-bill/3503/text>
 346 3. U.S Congress H.R.6584 - DEPICT Act [Internet]. 117th Congress (2021-2022)
 347 2022. [cited 2024 Dec 16]. Available from: [https://www.congress.gov/bill/117th-](https://www.congress.gov/bill/117th-congress/house-bill/6584/text)
 348 [congress/house-bill/6584/text](https://www.congress.gov/bill/117th-congress/house-bill/6584/text)
 349 4. U.S. Food and Drug Administration. Diversity Action Plans to Improve
 350 Enrollment of Participants from Underrepresented Populations in Clinical Studies: Guidance
 351 for Industry. [cited 2024 Aug 19]; Available from:
 352 <https://www.fda.gov/media/179593/download>

353 5. Haendel M, Vasilevsky N, Unni D, et al. How many rare diseases are there?
354 *Nat Rev Drug Discov* 2019 19:2. 2019;19(2):77-78. doi:10.1038/d41573-019-00180-y

355 6. U.S Government Accountability Office (GAO). Rare Diseases: Although
356 Limited, Available Evidence Suggests Medical and Other Costs Can Be Substantial [Internet].
357 [cited 2024 Dec 30]. Available from: <https://www.gao.gov/products/gao-22-104235>

358 7. Yang G, Cintina I, Pariser A, Oehrlein E, Sullivan J, Kennedy A. The national
359 economic burden of rare disease in the United States in 2019. *Orphanet J Rare Dis*.
360 2022;17(1):163. doi:10.1186/S13023-022-02299-5

361 8. Orphan Drug Act, Pub L No. 97-414, 96 Stat 2049. [cited 2025 Jan 7];
362 Available from: <https://www.congress.gov/bill/97th-congress/house-bill/5238/text>

363 9. Fermaglich LJ, Miller KL. A comprehensive study of the rare diseases and
364 conditions targeted by orphan drug designations and approvals over the forty years of the
365 Orphan Drug Act. *Orphanet J Rare Dis*. 2023;18(1):163. doi:10.1186/S13023-023-02790-7

366 10. U.S. Food and Drug Administration. CDER Patient-Focused Drug
367 Development [Internet]. 2020 [cited 2024 May 16]. Available from:
368 [https://www.fda.gov/drugs/development-approval-process-drugs/cder-patient-focused-drug-](https://www.fda.gov/drugs/development-approval-process-drugs/cder-patient-focused-drug-development)
369 development

370 11. U.S. Food and Drug Administration. Breakthrough Devices Program
371 [Internet]. [cited 2024 May 16]. Available from: [https://www.fda.gov/medical-devices/how-](https://www.fda.gov/medical-devices/how-study-and-market-your-device/breakthrough-devices-program)
372 study-and-market-your-device/breakthrough-devices-program

373 12. U.S. Food and Drug Administration. Select CBER Projects and Activities
374 [Internet]. 2020 [cited 2024 May 16]. Available from: [https://www.fda.gov/vaccines-blood-](https://www.fda.gov/vaccines-blood-biologics/resources-you-biologics/select-cber-projects-and-activities)
375 biologics/resources-you-biologics/select-cber-projects-and-activities

376 13. U.S. Food and Drug Administration. Accelerating Rare disease Cures (ARC)
377 Program [Internet]. [cited 2024 May 16]. Available from: [https://www.fda.gov/about-](https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/accelerating-rare-disease-cures-arc-program)
378 [fda/center-drug-evaluation-and-research-cder/accelerating-rare-disease-cures-arc-program](https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/accelerating-rare-disease-cures-arc-program)

379 14. U.S. Food and Drug Administration. FDA Launches Pilot Program to Help
380 Further Accelerate Development of Rare Disease Therapies [Internet]. [cited 2024 May 16].
381 Available from: [https://www.fda.gov/news-events/press-announcements/fda-launches-pilot-](https://www.fda.gov/news-events/press-announcements/fda-launches-pilot-program-help-further-accelerate-development-rare-disease-therapies)
382 [program-help-further-accelerate-development-rare-disease-therapies](https://www.fda.gov/news-events/press-announcements/fda-launches-pilot-program-help-further-accelerate-development-rare-disease-therapies)

383 15. U.S Food and Drug Administration. Drug Trials Snapshots Summary Report:
384 Five-Year Summary and Analysis of Clinical Trial Participation and Demographics Contents.
385 2015 [cited 2024 Apr 27]; Available from:
386 <https://www.fda.gov/media/143592/download?attachment>

387 16. McGarry ME, Gibb ER, Oates GR, Schechter MS. Left behind: the potential
388 impact of CFTR modulators on racial and ethnic disparities in cystic fibrosis. *Paediatr Respir*
389 *Rev.* 2022;42:35. doi:10.1016/j.prrv.2021.1201

390 17. Mathews KD, Conway KM, Gedlinske AM, Johnson N, Street N, Butterfield
391 RJ, et al. Characteristics of clinical trial participants with Duchenne muscular dystrophy: data
392 from the Muscular Dystrophy Surveillance, Tracking, and Research Network (MD STARnet).
393 *Children.* 2021;8(10). doi:10.3390/children8100835

394 18. Fedewa SA, Valentino LA, Koo A, Cafuir L, Tran DQ, Antun A, et al. Race
395 and ethnicity reporting and representation in hemophilia clinical trials. *Blood Adv.*
396 2024;8(10):2351-2360. doi:10.1182/bloodadvances.2024012862

397 19. Hantel A, Luskin MR, Garcia JS, Stock W, DeAngelo DJ, Abel GA. Racial and
398 ethnic enrollment disparities and demographic reporting requirements in acute leukemia clinical
399 trials. *Blood Adv.* 2021;5(21):4352–60. doi:10.1182/bloodadvances.2021005148

400 20. Stoller JK. The challenge of rare diseases. *Chest*. 2018;153(6):1309-1314.
401 doi:10.1016/j.chest.2017.1218

402 21. Hee SW, Willis A, Tudur Smith C, Day S, Miller F, Madan J, et al. Does the
403 low prevalence affect the sample size of interventional clinical trials of rare diseases? An
404 analysis of data from the aggregate analysis of clinicaltrials.gov. *Orphanet J Rare Dis*.
405 2017;12(1):1-20. doi:10.1186/s13023-017-0597-1

406 22. Rottas M, Thadeio P, Simons R, Houck R, Gruben D, Keller D, et al.
407 Demographic diversity of participants in Pfizer sponsored clinical trials in the United States.
408 *Contemp Clin Trials*. 2021;106. doi:10.1016/j.cct.2021.106421

409 23. Turner BE, Steinberg JR, Weeks BT, Rodriguez F, Cullen MR. Race/ethnicity
410 reporting and representation in US clinical trials: A cohort study. *Lancet Reg Health Am*.
411 2022;11. doi:10.1016/j.lana.2022.100252

412 24. Agboola F, Wright AC. A framework for evaluating the diversity of clinical
413 trials. *J Clin Epidemiol*. 2024;169. doi:10.1016/j.jclinepi.2024.111299

414 25. Venugopal N, Naik G, Jayanna K, Mohapatra A, Sasinowski FJ, Kartha R V,
415 et al. Review of methods for estimating the prevalence of rare diseases. *Rare Dis Orphan*
416 *Drugs J* 2024;3:5. doi:10.20517/rdodj.2023.39

417 26. Sakate R, Fukagawa A, Takagaki Y, Okura H, Matsuyama A. Trends of
418 clinical trials for drug development in rare diseases. *Curr Clin Pharmacol*. 2018;13(3):199-
419 208. doi:10.2174/1574884713666180604081349

420 27. The Lancet Diabetes & Endocrinology. Spotlight on rare diseases. *Lancet*
421 *Diabetes Endocrinol*. 2019;7(2):75. doi:10.1016/s2213-8587(19)30006-3

422 28. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau
423 V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet
424 database. *Eur J Hum Genet*. 2020;28(2):165. doi:10.1038/s41431-019-0508-0

425 29. Merone L, Tsey K, Russell D, Nagle C. “I Just Want to Feel Safe Going to a
426 Doctor”: Experiences of Female Patients with Chronic Conditions in Australia. *Women’s*
427 *Health Rep.* 2022;3(1):1016. doi:10.1089/whr.2022052

428 30. Faye F, Crocione C, Anido de Peña R, Bellagambi S, Escati Peñaloza L,
429 Hunter A, et al. Time to diagnosis and determinants of diagnostic delays of people living with
430 a rare disease: results of a Rare Barometer retrospective patient survey. *Eur J Hum Genet.*
431 2024. Published online May 16, 2024;1-11. doi:10.1038/s41431-024-01604-z

432 31. Bogart KR, Irvin VL. Health-related quality of life among adults with diverse
433 rare disorders. *Orphanet J Rare Dis.* 2017;12(1):177. doi:10.1186/s13023-017-0730-1

434 32. Chen S, Carter D, Brockenbrough PB, Cox S, Gwathmey K. Racial disparities
435 in ALS diagnostic delay: a single center’s experience and review of potential contributing
436 factors. *Amyotroph Lateral Scler Frontotemporal Degener.* 2024;25(1-2):112-118.
437 doi:10.1080/21678421.2023.2273361

438 33. DiMango E, Simpson K, Menten E, Keating C, Fan W, Leu CS. Health
439 Disparities among adults cared for at an urban cystic fibrosis program. *Orphanet J Rare Dis.*
440 2021;16(1):1-5. doi:10.1186/s13023-021-01965-4

441 34. Talwar A, Morel KC, Perez VDJ, Talwar A. The impact of socioeconomic,
442 racial, and ethnic disparities on pulmonary hypertension diagnosis and treatment. *Adv Pulm*
443 *Hypertens.* 2022;21(2):30-34. doi:10.21693/1933-088X-21.2.30

444 35. The Lewin Group Inc. The Cost of Delayed Diagnosis in Rare Disease: A
445 Health Economic Study. Prepared for EveryLife Foundation for Rare Diseases [Internet].
446 2023 Sep [cited 2025 Jan 9]. Available from: [https://everylifefoundation.org/wp-](https://everylifefoundation.org/wp-content/uploads/2023/09/EveryLife-Cost-of-Delayed-Diagnosis-in-Rare-Disease_Final-Full-Study-Report_0914223.pdf)
447 [content/uploads/2023/09/EveryLife-Cost-of-Delayed-Diagnosis-in-Rare-Disease_Final-Full-](https://everylifefoundation.org/wp-content/uploads/2023/09/EveryLife-Cost-of-Delayed-Diagnosis-in-Rare-Disease_Final-Full-Study-Report_0914223.pdf)
448 [Study-Report_0914223.pdf](https://everylifefoundation.org/wp-content/uploads/2023/09/EveryLife-Cost-of-Delayed-Diagnosis-in-Rare-Disease_Final-Full-Study-Report_0914223.pdf)

- 449 36. Baynam G, Baker S, Steward C, Summar M, Halley M, Pariser A. Increasing
450 diversity, equity, inclusion, and accessibility in rare disease clinical trials. *Pharmaceut Med*.
451 Published online July 1, 2024. doi:10.1007/s40290-024-00529-8
- 452 37. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE
453 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication
454 guidelines from a detailed consensus process. *Am J Crit Care*. 2015;24(6):466-473.
455 doi:10.4037/ajcc2015455
- 456 38. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of
457 observational studies in epidemiology (STROBE) statement: guidelines for reporting
458 observational studies. *PLoS Med*. 2007;4(10):e296. doi:10.1371/journal.pmed.0040296
- 459 39. ClinicalTrials.gov [Internet]. [cited 2024 Jun 4]. Available from:
460 <https://www.clinicaltrials.gov/>
- 461 40. Genetic and Rare Diseases Information Center [Internet]. [cited 2024 Jun 2].
462 Available from: <https://rarediseases.info.nih.gov/>
- 463 41. Orphanet [Internet]. [cited 2024 Jun 2]. Available from:
464 <https://www.orpha.net/>
- 465 42. Rare Diseases – RARE-X [Internet]. [cited 2024 May 10]. Available from:
466 <https://rare-x.org/rare-diseases/>
- 467 43. ClinicalTrials.gov. Studies By Topic [Internet]. [cited 2024 Jun 2]. Available
468 from: https://classic.clinicaltrials.gov/ct2/search/browse?brwse=ord_alpha_all
- 469 44. Office of Management and Budget, *OMB Directive No. 15 Revisions to the*
470 *Standards for the Classification of Federal Data on Race and Ethnicity*. [Internet]. [cited
471 2025 Jan 9]. Available from:
472 https://obamawhitehouse.archives.gov/omb/fedreg_1997standards

473 45. Miller EG, Woodward AL, Flinchum G, Young JL, Tabor HK, Halley MC.
474 Opportunities and pitfalls of social media research in rare genetic diseases: a systematic
475 review. *Genet Med*. 2021;23(12):2250. doi:10.1038/S41436-021-01273-Z

476 46. Parsons M, Tong Y, Valenti SC, Gorelik V, Bhatnagar S, Boily M, et al.
477 Reporting of participant demographics in clinical trials published in general radiology
478 journals. *Curr Probl Diagn Radiol*. 2024;53(1):81-91. doi:10.1067/j.cpradiol.2023810

479 47. U.S. Department of Health and Human Services. Program Assistance Letter
480 (PAL 2016–02): Approved Uniform Data System Changes for Calendar Year 2016 [Internet].
481 2016 [cited 2025 Jan 7]. Available from: [https://bphc.hrsa.gov/sites/default/files/bphc/data-](https://bphc.hrsa.gov/sites/default/files/bphc/data-reporting/program-assistance-letter-2016-02.pdf)
482 reporting/program-assistance-letter-2016-02.pdf

483 48. U.S. Department of Health and Human Services. Clinical Trials Registration
484 and Results Information Submission. Final Rule. *Federal Register*. 2016; Vol 81(183):
485 64982-65157. [cited 2025 Jan 7]. Available from [https://www.govinfo.gov/content/pkg/FR-](https://www.govinfo.gov/content/pkg/FR-2016-09-21/pdf/2016-22129.pdf)
486 2016-09-21/pdf/2016-22129.pdf

487 49. Schwartz JB. Representative enrolment of older adults in clinical trials: the
488 time is now. *Lancet Healthy Longev*. 2023;4(7):e301-e303. doi:10.1016/S2666-
489 7568(23)00088-0

490 50. Population of Asia (2024) - Worldometer [Internet]. [cited 2024 Dec 22].
491 Available from: <https://www.worldometers.info/world-population/asia-population/>

492 51. Clark LT, Watkins L, Piña IL, et al. Increasing diversity in clinical trials:
493 overcoming critical barriers. *Curr Probl Cardiol*. 2019;44(5):148-172.
494 doi:10.1016/j.cpcardiol.2018.1102

495 52. Rodríguez-Torres E, González-Pérez MM, Díaz-Pérez C. Barriers and
496 facilitators to the participation of subjects in clinical trials: An overview of reviews. *Contemp*
497 *Clin Trials Commun*. 2021;23:100829. doi:10.1016/j.conctc.2021.100829

498 53. Office of Research on Women’s Health. Review of the Literature: Primary
499 Barriers and Facilitators to Participation in Clinical Research. [cited 2024 Jun 10]; Available
500 from: https://orwh.od.nih.gov/sites/orwh/files/docs/orwh_outreach_toolkit_litreview.pdf
501 54. Gamble VN. Under the shadow of Tuskegee: African Americans and health
502 care. *Am J Public Health*. 1997;87(11):1773-1778. doi:10.2105/ajph.87.11.1773
503 55. Amendola LM, Robinson JO, Hart R, et al. Why patients decline genomic
504 sequencing studies: experiences from the CSER consortium. *J Genet Couns*. 2018;27(5):1220.
505 doi:10.1007/s10897-018-0243-7
506 56. Claw KG, Anderson MZ, Begay RL, et al. A framework for enhancing ethical
507 genomic research with Indigenous communities. *Nat Commun*. 2018;9(1).
508 doi:10.1038/S41467-018-05188-3
509 57. Passmore SR, Jamison AM, Hancock GR, Abdelwadoud M, Mullins CD,
510 Rogers TB, et al. “I’m a little more trusting”: components of trustworthiness in the decision to
511 participate in genomics research for African Americans. *Public Health Genom*. 2019;22(5-
512 6):215. doi:10.1159/000505271
513 58. Kahn JM, Gray DM, Oliveri JM, Washington CM, DeGraffinreid CR, Paskett
514 ED. Strategies to improve diversity, equity, and inclusion in clinical trials. *Cancer*.
515 2022;128(2):216. doi:10.1002/cnecr.33905
516 59. Mugas J, Olayiwola W, Gardiner K. 119 Improving recruitment of children
517 from ethnic minorities with rare diseases in clinical research facility. *Arch Dis Child*.
518 2021;106(Suppl 3):A44. doi:10.1136/archdischild-2021-gosh.119
519 60. Serrano JG, O’Leary M, VanNoy GE, Mangilog BE, Holm IA, Fraiman YS, et
520 al. Advancing Understanding of Inequities in Rare Disease Genomics. *Clin Ther*.
521 2023;45(8):745-753. doi:10.1016/j.clinthera.2023610

- 522 61. Buttery SC, Philip KEJ, Alghamdi SM, Williams PJ, Quint JK, Hopkinson NS.
523 Reporting of data on participant ethnicity and socioeconomic status in high-impact medical
524 journals: a targeted literature review. *BMJ Open*. 2022;12(8):e064276. doi:10.1136/bmjopen-
525 2022-064276
- 526 62. Florez MI, Botto E, Kim JY. Mapping strategies for reaching
527 socioeconomically disadvantaged populations in clinical trials. *JAMA Netw Open*.
528 2024;7(6):e2413962-e2413962. doi:10.1001/jamanetworkopen.2024.13962
- 529 63. Walewski JL, Donovan D, Nori M. How many zebras are there, and where are
530 they hiding in medical literature? A literature review of publications on rare diseases. *Expert*
531 *Opin Orphan Drugs*. 2019;7(11):513-519. doi:10.1080/21678707.2019.1684260

532 **Declarations**

533 **Ethics approval and consent to participate**

534 Not applicable

535 **Consent for publication**

536 Not applicable

537 **Availability of data and material**

538 This study used publicly available data. The rare disease keywords, publications, and clinical
539 trial data included in the analysis are available from the corresponding author upon reasonable
540 request.

541 **Competing interests**

542 The Rare Disease Diversity Coalition (RDDC) funded this study. RDDC facilitated access to
543 members of the Diversity in Clinical Trials Working Group, some of whom participated as
544 non-author collaborators in this study. Additionally, RDDC contributed to and supported data
545 analysis, interpretation, journal selection, and manuscript review and approval. Dr Debra
546 Regier (Children’s National Hospital) provided editorial support. IndoUSrare independently
547 designed, executed, and reported the study, retaining full autonomy over all aspects of the
548 research process. The authors declare that they have no other competing interests to declare.

549 **Funding**

550 Indo US Organization for Rare Diseases (IndoUSrare) received funding from the RDDC.

551 **Author contributions**

552 HKR, LM, and SMS contributed to the conceptualization of the study. LM and SMS acquired
553 data, developed methods, conducted investigation, validated the data, drafted the manuscript,
554 and prepared figures; SMS created new software for data extraction. LM, DD, and SMS
555 analyzed the data, and DD performed statistical analysis. LM, SMS, DD NV, RVK, HKR,
556 JNW, and LGB interpreted the data, reviewed it for important intellectual content, and revised
557 and edited the manuscript. HKR, NV, and RVK obtained funding for the study. LM managed
558 the project administration with JNW and LGB, providing administrative, technical, or
559 material support. HKR supervised the study. All authors read and approved the manuscript.

560 **Acknowledgments**

561 IndoUSrare acknowledges the RDDC for providing funding. The authors thank the RDDC
562 Clinical Trials Working Group members for their invaluable contributions during manuscript
563 development. We acknowledge Ms Bethlehem Addisu Demissie, Ms Bethlehem Sisay Tefera,

564 Ms Cindy Diana Umanzoar Figueroa, Dr Dora Akinyi Mugambi, and Ms. Harriet Tunu
 565 Baraka for partial literature screening; Dr Dora Akinyi Mugambi, Ms Leanne Marie Woehlke,
 566 Ms. Kathleen Machuzak, Ms Veronica Mullins, and Dr Paul I Howard for analysis,
 567 interpretation, and content review; Dr Debra S Regier for editorial support; and Ms. Jocelyn
 568 Cooper and Ms. Lolita Smith-Moore for administrative support. We also acknowledge
 569 support received from Children’s National Hospital. We also thank Dr Pragya Chaube for her
 570 initial involvement in the literature search strategy and IndoUSrare volunteers for their
 571 support in partial literature screening and data extraction: Dr Dhanya K, Ms Ramya
 572 Muralikrishna, Mr Rohan Honganoor, and Mr Vikram Nadathur. We also thank Dr Kirkland
 573 Wilson for his input on rare disease filtering and statistical tests.

574 **Authors’ information**

575 Indo US Organization for Rare Diseases, Herndon, VA 20171, USA
 576 Lavanyaa Manjunatha, PhD; Saundarya MS, MSc; Deepika Dokuru, PhD; Nisha Venugopal,
 577 PhD; Reena V Kartha, PhD; Harsha K Rajasimha MS, PhD
 578 Center for Orphan Drug Research, Department of Experimental and Clinical Pharmacology,
 579 University of Minnesota Twin Cities, Minneapolis, MN, USA
 580 Reena V Kartha, PhD
 581 Jeeva Clinical Trials Inc., Manassas, VA 20109, USA; School of Systems Biology, George
 582 Mason University, Fairfax, VA 22030, USA
 583 Harsha K Rajasimha MS, PhD
 584 Rare Disease Diversity Coalition, Black Women’s Health Imperative, Atlanta, GA 3013, USA
 585 Jenifer Ngo Waldrop, MS; Linda Goler Blount, MPH

FIGURES

Figure 1: Distribution of participants by race and ethnicity: US-only RD trials (1991–2022)^a from ClinicalTrials.gov

Figure 1A Caption: Aggregate and Mean proportions of participants by race and ethnicity were compared with US census averages using a Chi² test. Error bars for mean proportions indicate standard deviation. The significance of the “Ethnicity unknown or not reported” subcategory could not be evaluated as the US census lacks a corresponding category.

Figure 1A Legend:

^aRefers to the trial start years.

^bWeighted average percentages of US census data (1990–2022) calculated using weights based on the distribution of trial participants over time.

^cAggregate proportions: calculated by dividing the number of participants of a racial or ethnic category by the total number of participants for whom race or ethnicity data are available.

^dMean proportions: calculated by averaging the racial or ethnic proportions in each trial.

*Indicates significance at Bonferroni-corrected level (race: 0071; ethnicity: 0167); *p-values* for significance in Table S10.

Abbreviations: N1, total number of participants analyzed for race; N2, total number of participants analyzed for ethnicity; RD, rare disease; US, United States.

Figure 1B Caption: The percentage of trials having representation for each race and ethnicity subcategory was evaluated by comparing aggregate proportions of participants with US census averages. “No representation” refers to trials with 0% aggregate proportions of participants for each race or ethnicity subcategory. The percentage of trials for representation of the “Ethnicity unknown or not Reported” subcategory could not be evaluated because the US census lacks a corresponding category. Data labels with category.

Figure 1B Legend:

^aRefers to trial start years.

^bWeighted average percentages of US census data (1990–2022) calculated using weights based on the distribution of trial participants over time.

Abbreviations: N1, number of trials with study results reporting race data in NIH/OMB format; N2, number of trials with study results reporting ethnicity data in NIH/OMB format; RD, rare disease; US, United States.

Figure 2: Participants distribution by funder type and study phase in US-only RD trials (1991–2022)^a

Figure 2 Caption: Aggregate and mean proportions by race and ethnicity were calculated and compared with the US census averages^b using Chi² tests. Post hoc analyses were conducted to determine the specific racial or ethnic groups contributing to the observed differences. Bonferroni correction was used to adjust the significance threshold. Proportions that significantly exceeded the US census averages^b are shaded in blue.

Figure 2 Legend:

^aRefers to trial start years.

^bWeighted average percentages of US census data (1990–2022) calculated using weights based on participant distribution in RD trials (1991–2022).

*Indicates significance at Bonferroni-corrected significance level (AP: 0071; MP: 0063; MP funder type: 0125, *p-values* for significance in Table S11.

Abbreviations: AP, aggregate proportion; MP, mean proportion; N, number of trials; NA, not available; RD, rare disease; US, United States.

Figure 3: Temporal participation trends by race and ethnicity: US-only RD trials (1991–2022)^a from ClinicalTrials.gov

Figure 3 Caption: Aggregate and mean proportions by race and ethnicity were calculated for three decades (1991–2000, 2001–2010, and 2011–2020), the years 2021–2022, and the entire 30-year period (1991–2022). US census data for the years 2000, 2010, 2020, 2022, and US census averages (1990–2020) are plotted as a red line graph for comparison. US census averages (1990–2020) refer to weighted average percentages of US census data calculated using weights based on participant distribution in RD trials (1991–2022).

Figure 3 Legend:

^aRefers to trial start years

Abbreviations: N, number of trials; RD, rare disease; US, United States.

625 TABLES

626 **Table 1: Characteristics of publications reporting RD research with US-based participants (2010–2023)**

Characteristics or Variables	Total publications, n (%)	US-only publications, n (%)	Multinational ^a publications, n (%)
Publications	234 (100.0)	131 (56.0)	103 (44.0)
Study type			
Clinical trial	79 (33.8)	34 (26.0)	45 (43.7)
Observational	47 (20.1)	29 (22.1)	18 (17.5)
Genetics/Genomics	29 (12.4)	17 (13.0)	12 (11.7)
Real-world data	18 (7.7)	18 (13.7)	0 (0.0)
Survey	21 (8.3)	13 (9.9)	8 (7.8)
Qualitative	18 (7.7)	10 (7.6)	8 (7.8)
Web-based study	4 (1.7)	4 (3.1)	0 (0.0)
Clinical report	7 (3.0)	3 (2.3)	4 (3.9)
Epidemiological	2 (0.9)	1 (0.8)	1 (1.0)
Secondary analysis	1 (0.4)	1 (0.8)	0 (0.0)
Systematic review/Literature review	4 (1.7)	1 (0.8)	3 (2.9)
Other	4 (1.7)	0 (0.0)	4 (3.9)
Participant type			
Patients	211 (90.2)	117 (89.3)	94 (91.3)
Caregivers	13 (5.6)	8 (6.1)	5 (4.9)
Patients and Caregivers	8 (3.4)	4 (3.1)	4 (3.9)

Characteristics or Variables	Total publications, n (%)	US-only publications, n (%)	Multinational ^a publications, n (%)
Care provider	2 (0.9)	2 (1.5)	0 (0.0)
Age^b	220 (94)	123 (93.9)	97 (94.2)
Prenatal ^c	3 (1.4)	1 (0.8)	2 (1.9)
Child	32 (14.5)	22 (16.8)	10 (9.7)
Adult	27 (12.3)	16 (12.2)	11 (10.7)
Older adult	2 (0.9)	2 (1.5)	0 (0.0)
Child and Adult	62 (28.2)	35 (26.7)	27 (26.2)
Adult and Older adult	56 (25.5)	27 (20.6)	29 (28.2)
Child, Adult, and Older adult	27 (12.3)	11 (8.4)	16 (15.5)
Not classified ^d	11 (5.0)	9 (6.9)	2 (1.9)
Sex or Gender^b	202 (86.3)	111 (84.7)	91 (88.3)
Male and Female or Men and Women	135 (66.8)	79 (71.2)	56 (61.5)
Male, Female, Other ^e	10 (5.0)	6 (5.5)	4 (3.9)
100% male	8 (4.0)	7 (5.3)	1 (1.0)
100% female	9 (4.5)	7 (5.3)	2 (1.9)
Male or Female	40 (19.8)	12 (10.8)	28 (27.2)
Not classified/other	0 (0.0)	0 (0.0)	0 (0.0)
Race^b	117 (50.0)	69 (52.7)	48 (46.0)
American Indian or Alaska Native	25 (21.4)	10(14.5)	11 (22.9)
Asian	65 (55.6)	29 (42.0)	36 (75.0)
Black or African American	76 (65.0)	41 (59.4)	32 (66.7)
Native Hawaiian or Other Pacific Islander	21 (17.9)	8 (11.6)	9 (18.8)

Characteristics or Variables	Total publications, n (%)	US-only publications, n (%)	Multinational ^a publications, n (%)
White	113 (96.6)	67 (97.1)	47 (97.9)
More than one race	24 (20.5)	11 (8.4)	12 (25.0)
Ethnicity^b	70 (29.9)	39 (29.7)	31 (30.1)
Hispanic or Latino	56 (80.0)	29 (74.4)	29 (93.5)
Not Hispanic or Latino	33 (47.1)	18 (46.2)	18 (58.1)
SFs^b	37 (15.8)	28 (21.3)	9 (8.7)
Language	16 (43.2)	9 (32.1)	7 (77.8)
Education	23 (62.2)	16 (57.1)	7 (77.8)
Employment	13 (35.1)	9 (32.1)	4 (44.4)
Income	14 (37.8)	12 (42.9)	2 (22.2)
Geographical location	7 (21.6)	6 (21.4)	1 (11.1)
Health insurance	14 (37.8)	12 (42.9)	2 (22.0)
Relationship status	11 (29.7)	5 (17.8)	4 (44.0)
Social support	1 (2.7)	1 (3.5)	0 (0.0)
Disability allowance	3 (8.1)	3 (10.7)	0 (0.0)

627

628 ^aPublications that reported studies conducted in the US and other countries.

629 ^bPercentages for subcategories of these characteristics were calculated based on the total number of publications reporting the characteristics.

630 ^cIncludes the categories “prenatal and child” and “prenatal, child, and adult.”

631 ^dIncludes publications that did not report age range and hence could not be classified into the existing age groups.

632 ^eIncludes the categories “Non-binary”/ “Prefer not to say”/ “Not reported.”

633 Abbreviations: n, number of publications; RD, rare diseases; US, United States.

634 **Table 2: Characteristics of RDs clinical trials^a from ClinicalTrials.gov with US-based participants (1986–2022)^b**

Characteristics	Total trials, n (%)	US-only trials, n (%)	Multinational ^c trials, n (%)
Clinical trials	8475 (100.0)	5927 (69.9)	2548 (30.1)
Study status^d			
Completed	6008 (70.9)	4127 (69.6)	1881 (73.8)
Terminated	2029 (23.9)	1553 (26.2)	476 (18.7)
Active not recruiting	408 (4.8)	221 (3.7)	187 (7.3)
Recruiting	1 (0.0)	1 (0.0)	0 (0.0)
Suspended	2 (0.0)	2 (0.0)	0 (0.0)
Unknown	27 (0.3)	23 (0.4)	4 (0.1)
Study type^d			
Interventional	8331 (98.3)	5806 (98)	2525 (99.1)
Observational	144 (1.7)	121 (2.0)	23 (0.9)
Study size (number of participants)			
Very small (1–10)	1326 (15.6)	1228 (20.7)	98 (3.8)
Small (1–50)	4972 (58.7)	4227 (71.3)	745 (29.2)
Medium (>50–500)	3078 (36.3)	1558 (26.3)	1520 (59.7)
Large (>500–1000)	245 (2.9)	49 (0.8)	196 (7.7)

Characteristics	Total trials, n (%)	US-only trials, n (%)	Multinational ^c trials, n (%)
Very Large (>1000)	134 (1.6)	53 (0.9)	81 (3.2)
Study phase^d			
Phase 1	442 (5.2)	360 (6.1)	82 (3.2)
Phase 1 Phase 2	1071 (12.6)	840 (14.2)	231 (9.1)
Phase 2	4097 (48.3)	3154 (53.2)	943 (37)
Phase 2 Phase3	166 (2.0)	76 (1.3)	90 (3.5)
Phase 3	1404 (16.6)	353 (6.0)	1051 (41.2)
Phase 4	341 (4.0)	261 (4.4)	80 (3.1)
Early Phase 1	33 (0.4)	29 (0.5)	4 (0.2)
Not applicable	921 (10.9)	854 (14.4)	67 (2.6)
Funder^d			
Federal	45 (0.5)	41 (0.7)	4 (0.2)
Industry	3206 (37.8)	1096 (18.5)	2110 (82.8)
NIH	853 (10.1)	710 (12.0)	143 (5.6)
Other	4371 (51.6)	4080 (68.8)	291 (11.4)
Age^d			
Child	488 (5.8)	271 (5.0)	217 (8.5)
Adult	271 (3.2)	234 (3.9)	37 (1.5)

Characteristics	Total trials, n (%)	US-only trials, n (%)	Multinational ^c trials, n (%)
Older adult	42 (0.5)	32 (0.5)	10 (0.4)
Child and Adult	644 (7.6)	430 (7.3)	214 (8.4)
Adult and Older adult	5832 (68.8)	4190 (70.7)	1642 (64.4)
Child, Adult, and Older adult	1198 (14.1)	770 (13.0)	428 (16.8)
Sex or Gender^d			
Male and Female	7252 (85.6)	4981 (84)	2277 (89.1)
Male, Female, Unknown or Not reported or Undifferentiated	27 (0.3)	21 (0.4)	6 (0.2)
Male only	435 (5.1)	295 (5.0)	140 (5.5)
Female only	630 (7.4)	528 (8.9)	102 (4.0)
Gender customized	78 (0.9)	57 (0.9)	21 (0.8)
No gender data	53 (0.6)	45 (0.8)	8 (0.3)
Race and ethnicity			
Race (NIH/OMB)	3869 (45.7)	2810 (47.4)	1059 (41.6)
Race only (NIH/OMB)	1088 (12.8)	795 (13.4)	293 (11.5)
Race customized	210 (2.5)	63 (1.1)	147 (5.8)
Ethnicity (NIH/OMB)	3259 (38.5)	2213 (37.4)	1046 (41.1)
Ethnicity only (NIH/OMB)	476 (5.6)	197 (3.3)	279 (11.0)
Ethnicity customized	20 (0.2)	9 (0.2)	11 (0.4)

Characteristics	Total trials, n (%)	US-only trials, n (%)	Multinational ^c trials, n (%)
Race and Ethnicity	2783 (32.8)	2016 (34.0)	767 (30.0)
Both race and ethnicity customised	775 (9.1)	383 (6.5)	392 (15.4)
No race or ethnicity	3353 (39.6)	2536 (42.8)	817 (32.1)
Race and Ethnicity subcategories^d			
American Indian or Alaska Native	659 (7.8)	339 (5.7)	320 (12.6)
Asian	2058 (24.3)	1257 (21.2)	801 (31.4)
Black or African American	2632 (31.1)	1824 (30.8)	808 (31.7)
Native Hawaiian or Other Pacific Islander	432 (5.0)	214 (2.5)	218 (8.6)
White	3732 (44.1)	2691 (45.4)	1041 (40.9)
Hispanic or Latino	2340 (27.6)	1433 (24.2)	908 (38.5)

^aIncludes only trials with ‘study results’ on ClinicalTrials.gov.

^bRefers to trial start years.

^cTrials conducted in the US and other countries.

^dSubcategories as reported by ClinicalTrials.gov.

Abbreviations: n, number of trials; NIH, National Institute of Health; OMB, Office of Management and Budget, US, United States.

Additional information

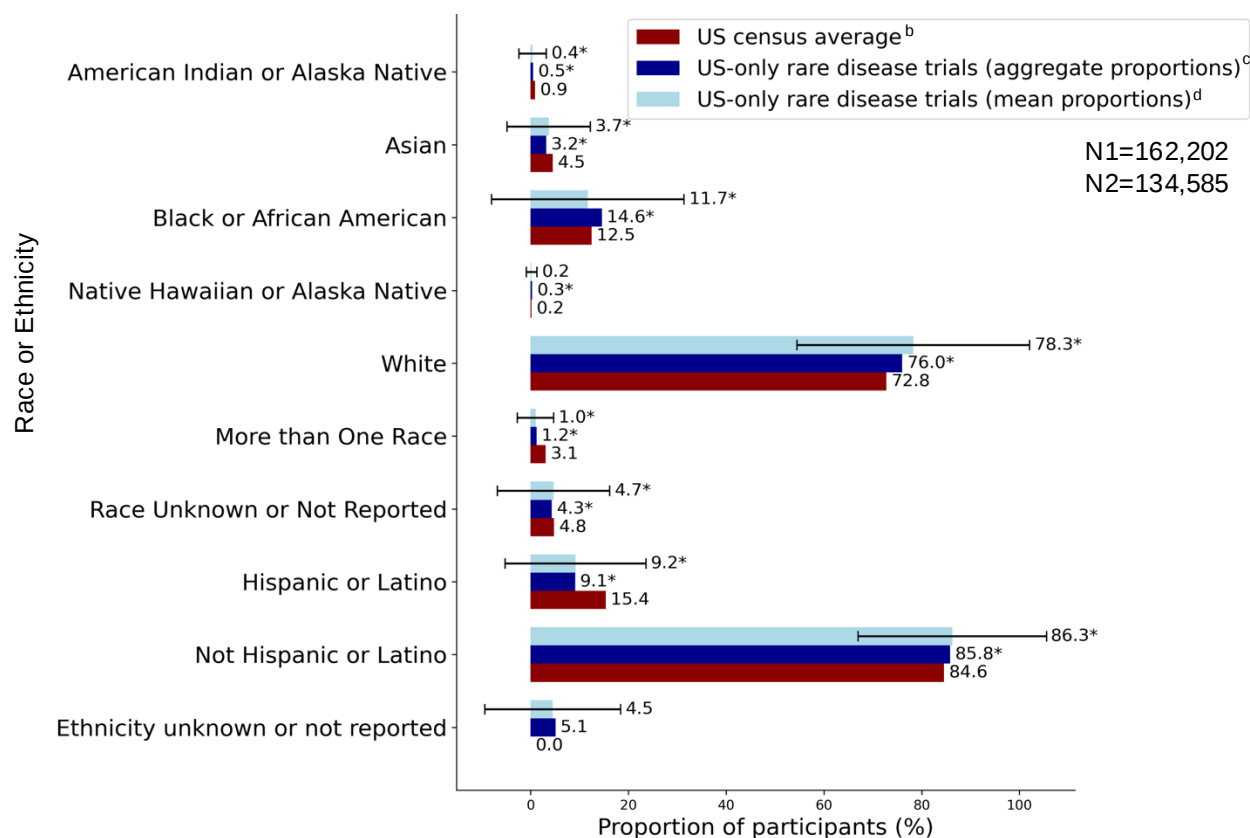
Additional File 1_Manjunatha et al_Current Trends_MUP State in RD research_16022025

Additional File 1: Supplementary Methods 1 and 2, Figures S1–S3, and Tables S1–S11. Supplementary Methods 1. Literature assessment screening and data extraction; Supplementary Methods 2. Rare disease clinical trials filtering and validation; Supplementary References; Figure S1: Screening and filtering of US-based RD publications and clinical trials from ClinicalTrials.gov; Figure S2: Temporal trends in US-only RD trials reporting race and ethnicity data over time; Figure S3: Temporal participation trends: More than one race, Not Hispanic or Latino, Race and Ethnicity unknown or not reported in US-only RD trials (1991–2022); Table S1: PubMed and Cochrane Library search queries; Table S2: Keywords used for title screening of publications; Table S3: Keywords for identifying RD clinical trials reporting socioeconomic factors data; Table S4: A snapshot of US-based RD publications reporting race and ethnicity in a non-standard format; Table S5: Socioeconomic factors reporting in US-based RD publications; Table S7: Participant distribution statistics from US-based RD clinical trials (1986–2022); Table S8: Socioeconomic factors reporting in US-based RD trials (1991–2022); Table S9: Participant distribution statistics in US-based RD trials reporting socioeconomic factors (1991–2022); Table S10: p-values for statistical significance of aggregate and mean race and ethnicity proportions, in US-only RD trials (1991–2022), compared with the US census averages; Table S11: p-values for statistical significance of aggregate and mean proportions of race and ethnicity in US-only RD trials (1991–2022), compared with US census averages: by study phase and funder type

654 **Additional File 2_Table S12 and S13_Manjunatha et al_Current State_MUP trends in RD research_1602025.xlsx**

655 Additional File 2: Table S12–S13. Table S12: p-values for significance of aggregate and MP of race and ethnicity in US-only RD trials (1991–
656 2022) compared with US census averages: by study phase and funder type; Table S13: Racial and ethnic distribution of participants in
657 multinational RD trials (1997–2022)

A Distribution of participants by race and ethnicity in US-only RD trials (1991–2022)^a compared with US census averages^b



B Representation of participants by race and ethnicity in US-only RD trials (1991–2022)^a

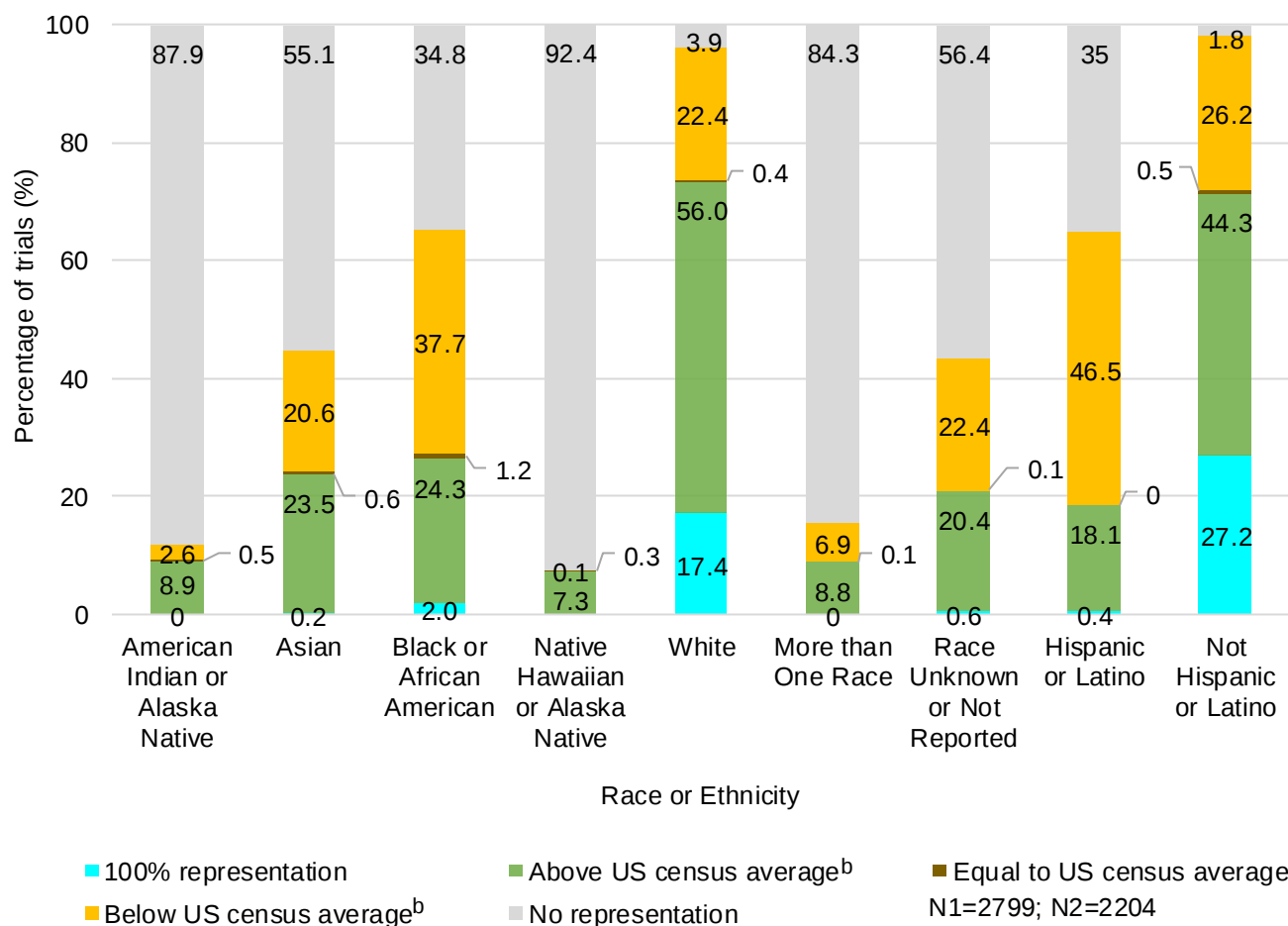
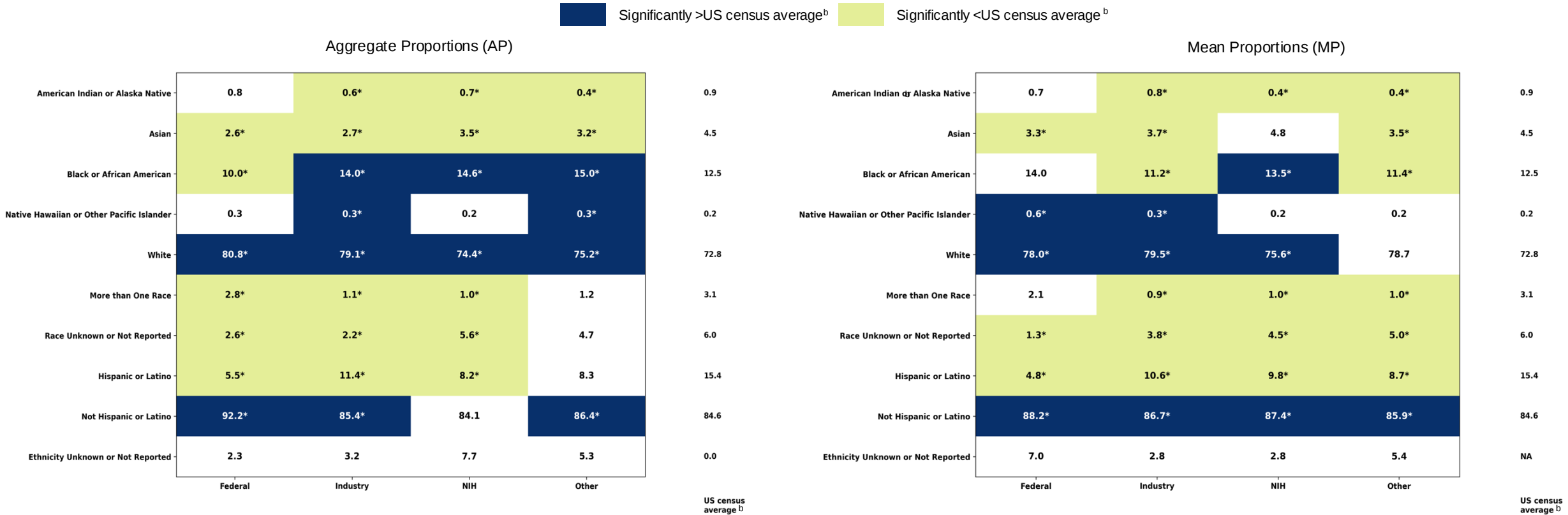


Figure 2: Participants distribution by funder type and study phase in US-only RD trials (1991–2022)^a

A Heat map of the distribution of participants by race and ethnicity across funder types



B Heat map of the distribution of participants by race and ethnicity across study phases



Figure 3: Temporal participation trends by race and ethnicity: US-only RD trials (1991–2022)^a from ClinicalTrials.gov

