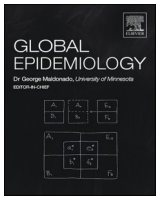
Global Epidemiology 7 (2024) 100134

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology



Improved clinical trial race/ethnicity reporting and updated inclusion profile, 2017–2022: A New Jersey snapshot

Elli Gourna Paleoudis [a,b], Zhiyong Han [a], Simon Gelman [b], Hernan Arias-Ruiz [a], Destiney Carter [a], Jovan Bertrand [a], Nicole Mastrogiovanni [a], Stanley R. Terlecky [a,*]

[a] Department of Medical Sciences, Hackensack Meridian School of Medicine, Nutley, NJ, USA
[b] Office of Research Administration, Hackensack Meridian Health Research Institute, Nutley, NJ, USA

ARTICLE INFO

Keywords:
Clinical trials
Race
Ethnicity
Evidence-based medicine
Clinicaltrials.gov

ABSTRACT

Background: Diverse representation in clinical trials is an important goal in the testing of a medical, diagnostic, or therapeutic intervention. To date, the desired level of trial equity and inclusivity has been unevenly achieved.
Methods: Employing the US National Library of Medicine's Clinicaltrials.gov registry, we examined 481 clinical trials conducted - at least in part - in the state of New Jersey. These trials were initiated after the FDA-mandated Common Rule changes, i.e., between January 2017 and October 2022, were enacted, and had their results posted. We analyzed sex/race/ethnicity reporting as well as applicable enrollment. Using meta-analysis, we estimated group participation proportions of a subset of the 481 identified trials; specifically, the 229 studies that were conducted solely within the US (i.e., without international sites) and compared them to US census data.
Findings: Within the 481 clinical trials analyzed, over 97% reported on the race and/or ethnicity of their enrollees; all included information on sex. Reporting was not affected by funding source or therapeutic area. Based on the 229 solely US-based studies, the participants overall were 76.7% White; 14.1% Black; 2.7% Asian; and 15% Hispanic. Inclusion of Black participants did not differ from the 2020 US census data; in contrast, the levels of Asian and Hispanic participation were below the corresponding census percentages.
Interpretation: The past five years have seen an overall uptick in the equity of race/ethnicity reporting and inclusivity of clinical trials, as compared to previously reported data, presaging the potential acquisition of ever more powerful and meaningful results of such interventional studies going forward.
Funding: Support for this study comes from the Hackensack Meridian *Health* Research Institute and the Hackensack Meridian School of Medicine.
Research in context: Evidence before this study
Clinical trials are a critical part of determining whether or not a medical (drug/device/biologic) or socio-behavioral intervention is safe and truly effective. Through their use, scientific understanding is advanced and, ideally, human health is improved. To gain the most impactful information from a clinical trial, it should be sufficiently representative, that is, should enroll an adequate number of participants, and include a diverse population. Without such inclusion, the study is of only limited generalizability. Efforts are underway by funders, sites, and other stakeholders, to enhance reporting and promote inclusive enrollment. The extent to which such attempts are yielding results - at least for clinical trials in the state of New Jersey - is the focus of this data-driven analysis. The ClinicalTrials.gov registry database was carefully mined for the information contained in this report.
Added value of this study
Our analysis of clinical trials initiated in the state of New Jersey and conducted there or elsewhere in the US reveals several positive trends. Our 5-year snapshot reveals that a very large percentage of trials report on race/ethnicity - and inclusivity is improving. While there is still some way to go to have the demographic numbers in these trials match US census values, our results suggest that recent efforts are having an effect.
Implications of all the available evidence
For myriad reasons, clinical trials have not enjoyed the public's universal trust over the years. In many ways, medicine moves at the speed of trust - without it, the promise of modern healthcare is brought into question.

* Corresponding author at: Hackensack Meridian School of Medicine, 123 Metro Blvd., Suite M-4113, Nutley, NJ, USA.
E-mail address: stanley.terlecky@hmhn.org (S.R. Terlecky).

https://doi.org/10.1016/j.gloepi.2023.100134
Received 22 June 2023; Received in revised form 30 November 2023; Accepted 20 December 2023
Available online 23 December 2023

Clinical trials must include a commitment to diverse enrollment pools and equitable reporting under the law. Creating a legacy of trust - through greater inclusivity in clinical trials and more transparent reporting of results - will begin to heal the divide and engender faith in modern medicine and today's healthcare system. It would also allow for the desired far-reaching generalizability of results across patient populations. To better appreciate what needs to be done going forward, we must truly understand the state of clinical trials reporting and demographic inclusion. This report initiates such an analysis, by carefully documenting how New Jersey's clinical trials are performing. By virtue of its location (e.g., proximity to the cities of New York and Philadelphia) the state is part of a large biopharma cluster and healthcare nexus; it is critical that it performs well with respect to adopting/adhering to updated clinical trial guideline mandates. This report provides a glimpse - an important first look - into the state of clinical trials in New Jersey - from 2017 through 2022.

Introduction

Desired health outcomes are best achieved when evidence-based medicine, translational science, and clinical research are brought together at the level of patient care. Information and data are key; nowhere is this more true than in the conduct of clinical trials associated with the development of new drugs, devices, or treatments. Ideally, randomized trials are completed with appropriately powered (patient) sample sizes, and with all necessary steps taken to assure safety and efficacy. To permit such new interventions to be approved and widely employed, they must be tested in diverse populations - including across sex, racial, and ethnic groups. Recent evidence suggests that clinical trials are moving towards this goal of inclusive design and execution [1].

In 2007, the Food and Drug Administration (FDA) issued an amendment [2] requiring all applicable clinical trials (ACT) using FDA regulated drug/devices and pediatric postmarket surveillance to be publicly registered in order to ensure transparency and inclusion. In 2016, the Department of Health and Human Services (DHH), expanding on that law, published additional requirements on the clinical trial registration and result posting [3] that went into effect in January 2017. These changes in the common law after 2017 have begun influencing result posting within ClinicalTrials.gov [1], which is the largest publicly available data bank for clinical trials in the US. Additional efforts made by the FDA and other federal bodies, including funders (e.g., NIH grant funded opportunities), have mandated and provided incentives [4] for result reporting, as well as the inclusion of racial and ethnic minorities that have traditionally been underrepresented within clinical trials [5,6]. These include, among others, monetary fines and other penalties, as well as targeted funding opportunities supporting diverse trial enrollment.

In April of 2021, the FDA issued the first "Notice of Noncompliance" to the sponsor of a clinical trial for failing to submit clinical trial results to ClinicalTrials.gov as required by the Common Rule [7]. Subsequently, three letters were sent to other sponsors (through March of 2023), along with the establishment of significant civil and monetary penalties (over $10,000) [8] - suggest the systematic enforcement of the law. This enforcement has however received harsh criticism as to its effectiveness from both public [9] and academic groups, suggesting the inadequacy of these approaches.

While funders, federal bodies [10], and other organizations [11,12] continue working to ensure transparency within clinical trials and accurate representation of all populations, newer investigations focusing on the results posted in the ClinicalTrials.gov data bank are warranted.

The goal of this study was to document the race/ethnicity reporting and inclusivity profiles of recently completed clinical trials in the state of New Jersey and beyond. Against the backdrop of the new common law posting requirements; the recent pandemic; and the continued emergence of new molecular therapeutics, behavioral interventions, diagnostic treatments, and medical devices; such an up-to-date analysis was very much in order. Going forward, rigorous inventorying of reported ClinicalTrials.gov data may help to identify areas in which improvement strategies could be implemented to ensure more inclusive and equitable clinical testing.

Methods

Data sources

On October 11, 2022 we conducted an Advanced Search (Fig. 1a), using the publicly available ClinicalTrials.gov data bank (https://clinicaltrials.gov), for all studies initiated after January 2017 (i.e., post common law changes), with results posted, that were conducted in at least one site in the state of New Jersey. We excluded all studies that were not completed at the time to ensure accurately posted results; we included studies using participant populations of all ages.

For the comparative analysis, we examined only the subset of studies taking place solely in the US, referencing our data against the 2020 Census Bureau database [13,14]. See Fig. 1b for a graphic depiction state-by-state of the number of studies that meet these criteria.

Definitions and variables used

To ensure consistency with reporting requirements and the way results are reported within ClinicalTrials.gov (CT.gov), we adopted the sex, race, and ethnicity definitions (and categories) as employed by NIH/Office of Management and Budget (OMB) Standards [15]; we added an Unknown or Not Reported category to capture inconsistencies in data entry and to limit missing data points.

More specifically, we included Male, Female, Other for sex; White, Black or African American, Asian, Native Hawaiian or other Pacific Islander, American Indian or Alaska Native, Two or more races, Other, and Unknown or Not Reported for race; and Hispanic, Latino, and Unknown/Not Reported for ethnicity.

Information on sex/race/ethnicity was manually extracted from the data bank and entered into a REDCap database (Research Electronic Data Capture - which is a secure, web-based software platform designed to support data acquisition for research studies) developed specifically for this project [16,17]. The following data points were automatically extracted from CT.gov and imported directly into the (same) REDCap database: NCT number, Study Title, Intervention, Age of Participants, Phase, Start Date, Primary Completion Date, and Study Location.

In addition to sex/race/ethnicity, we also manually extracted information about the therapeutic area and whether or not the study was related to COVID-19.

Statistical analysis

We first explored reporting of sex, race, and ethnicity information in clinical trials conducted from 2017 to 2022 and assessed how reporting may be affected by study-level characteristics (*Reporting of Information*). Second, we analyzed racial, ethnic, and sex proportions, their heterogeneity, and associations with the study-level characteristics in trials conducted in the US (*Analysis of Racial, Ethnic, and Sex Proportions*). Study-level characteristics our team considered factors that could influence both reporting of race/ethnicity information and/or the racial/ethnic composition of clinical trials were: i. study location, ii. therapeutic area, iii. COVID-19 relatedness, iv. trial phases, and v. source of funding. Study location was based on the location of participants or sites

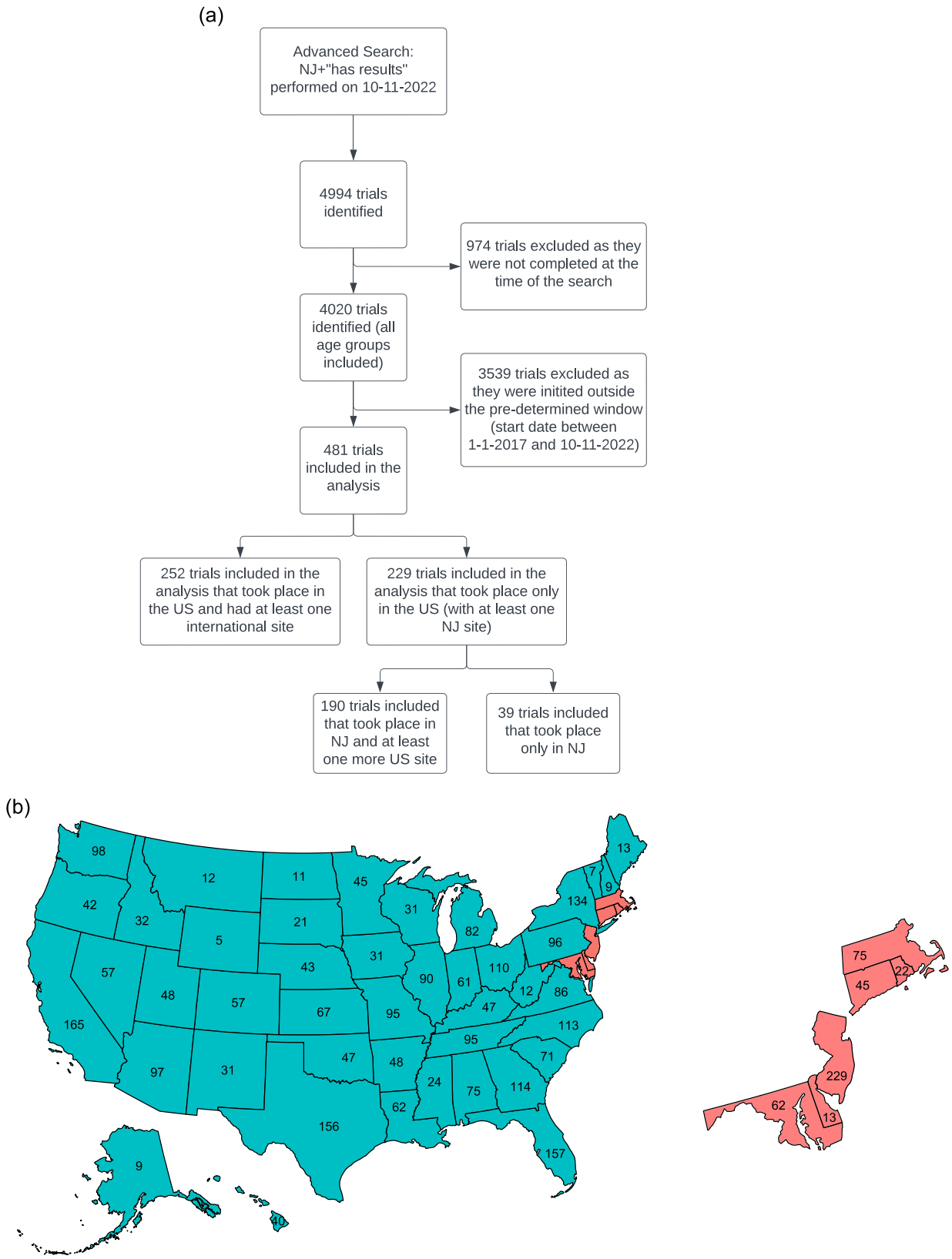


Fig. 1. a. CONSORT flow diagram of study search paring parameters and inclusion of trials.
b. Clinical trial distribution.
Map of the US with distribution of clinical trials that were conducted – at least in part – in New Jersey and solely within the US (i.e., without international sites. 229 such studies meet this criteria in total – hence the number 229 appears inside New Jersey.

involved in a study and contained three categories - US/International, US, and New Jersey. Therapeutic areas comprised 15 fields: anesthesia and analgesia, cardiovascular diseases, dermatology, endocrinology and metabolism, gastroenterology, immunology, infectious diseases, neurology, obstetrics and gynecology, oncology and hematology, ophthalmology, orthopedics, psychiatry, pulmonology and rheumatology, and urology. COVID-19 was an indicator of whether or not the study was related to aspects of COVID-19. Phase included phase 1, 2, 3, 4, other, and not applicable categories. The funding sources identified were Government, Industry, and Other.

Reporting of information

For the first objective, we used three binary (Yes/No) outcome variables, each indicating whether information was reported (1 - Was information on race and/or ethnicity reported?; 2 - Was information on race reported?; and 3 - Was information on ethnicity reported?) in each examined clinical trial. The data was summarized in contingency tables as frequencies and percentages. Univariate associations between reporting and study-level characteristics were tested with Fisher's exact tests.

Additionally, we tested association between reporting and study-level characteristics (study location, COVID-19, and source of funding) using logistic regression. We fit three generalized linear models of the form logit(*Outcome*) $\sim$ *location* + *COVID* + *funding* (see Supplementary Table 1 for details of the three models). Using the models, we predicted probabilities of reporting race and/or ethnicity, race, and ethnicity information (Yes) for various combinations of study-level characteristics (Supplementary Table 2).

Analysis of racial, ethnic, and sex proportions: meta-analysis

For the second objective, our outcome variables were proportions of Whites, Blacks, Asians, Hispanics, and Females. To estimate the racial/ethnic/sex proportions in clinical trials, we conducted a meta-analysis. We assumed that racial/ethnic/sex proportions were part of the trials' outcomes. Additionally, racial proportions are usually not uniform across a large geographical area; therefore, we assumed that there is a distribution of racial proportions and hence, used a random-effects model. An inverse variance method under a random-effects model was used to estimate the mean proportions. To describe heterogeneity in the estimated effects, we report the I^2 statistic, which is the ratio of excess dispersion to total dispersion [18]. It was of interest to explore the sources of variability in the racial proportions. Therefore, we conducted separate subgroup meta-analyses for each of the study-level characteristics.

Analysis was performed using R software, version 4.1.1 [19]. Meta-analysis was performed using *meta* package [20]. Figures were built with the *ggplot2* package [21].

Results

Reporting of information

Reporting of race and/or ethnicity information over a span of 5 years (2017–2022) is summarized in Table 1 and Supplementary Fig. 1. Overall, 97.1% (467/481) of all trials provided race and/or ethnicity data, 95.6% (460/481) of trials provided race data, and 76.5% (368/481) of trials reported ethnicity information; 100% of studies reported on sex.

We examined the relationship between reporting information and study-level characteristics. Univariate analysis revealed that the location of the trial was associated with reporting of race and/or ethnicity, race, and ethnicity. Trials taking place within the state of New Jersey reported this information less frequently than trials conducted at other US sites (Table 1 and Supplementary Fig. 1). Therapeutic area, COVID-19, and funding source did not affect whether a trial reported information or not; it is unclear if the study phase affected reporting since the observed differences are associated with the "Not applicable" category.

To simultaneously test whether study-characteristics affect reporting of race and/or ethnicity, race, and ethnicity information, logistic models that included study location, COVID-19, and source of funding as predictors, were used (see Methods). The models indicate that only study location had an effect on whether information was reported or not (see Supplementary Table 1 for model details). The models were used to predict probability of reporting for combinations of study-level characteristics. Supplementary Table 2 lists predicted probabilities of reporting information (Yes) and 95% prediction intervals for industry funded studies, adjusting for COVID-19 and location. New Jersey consistently showed lower probability of reporting on all three outcomes compared to trials with US/International or US sites (Supplementary Table 2). For example, 85% of New Jersey trials associated with COVID and funded by industry reported race information compared to 97% in US-conducted trials and only 82% of studies not related to COVID reported race compared to 96% in other US-conducted trials.

Reporting race, ethnicity, and sex proportions in clinical trials

Boxplots in Fig. 2 provide a basic summary of the representation of races, ethnicities, and sex that were observed in trials conducted within the US. Subsequently, we focused our analysis on three racial categories (Whites, Blacks, and Asians), one ethnic category (Hispanic), and one sex category (Females).

Variability in racial, ethnic, and sex proportions

Analysis of race and ethnicity variability for clinical trials conducted from 2017 to 2022 in the US revealed a high degree of variability without a noticeable trend, except that a reduction in the number of trials could be observed in and around the year 2020 (Fig. 3).

Overall meta-analysis

We estimated the mean proportion of Whites, Blacks, Asians, Hispanics, and Females in clinical trials using a random-effects meta-analysis. Summaries of the overall meta-analysis for all outcomes are shown in Table 2. Percentage of racial, ethnic, or sex categories from meta-analysis were tested against the 2020 census values. The 2020 Census values were taken from https://www.census.gov/quickfacts/fact/table/US/RHI125221. It should be noted that Race and Ethnic origin are two different categories. People of Hispanic origin could be of any race. The census splits racial categories by ethnic origin. Here for hypothesis testing, we used racial categories (White, Black, and Asian) defined by the census as "Race, alone", not conditioning on ethnicity.

The estimated combined proportion of Whites was 74.9% (95% CI: 72.0%–77.7%). This proportion did not differ from the 2020 US census value of 75.8%. Proportion of Blacks 15.1% (95% CI: 13.0%–17.4%) also did not differ from the US census (13.6%). However, Asians and Hispanics were underrepresented by 3.2% and 3.9%, respectively. Females were overrepresented by 8.4% compared to the 2020 US population. Graphical summary of the overall meta-analysis is shown in Fig. 4.

We also compared New Jersey's estimated racial proportions to the state's 2020 census data. Proportions of Whites and Blacks in these trials were not different from the 2020 census data for the state of New Jersey because the census values (70.7% and 15.4%, respectively) fell within 95% confidence intervals of our estimates (see Supplementary Table 3). Hispanics and Asians were slightly underrepresented in these trials because the 2020 New Jersey census values of 21.9% and 10.5% were higher than the upper limits of respective 95% confidence intervals of our estimates (see Supplementary Table 3).

Table 1
Reporting of race and/or ethnicity, race, and ethnicity information.

Characteristic	Reported Race and/or Ethnicity?		Reported Race?		Reported Ethnicity?	
	No, $N = 14$ (2.91%)	Yes, $N = 467$ (97.1%)	No, $N = 21$ (4.37%)	Yes, $N = 460$ (95.6%)	No, $N = 135$ (28.1%)	Yes, $N = 346$ (71.9%)
Therapeutic Area, n (%)						
Anesthesia and Analgesia	1 (14.3)	6 (85.7)	1 (14.3)	6 (85.7)	2 (28.6)	5 (71.4)
Cardiovascular diseases	4 (10.3)	35 (89.7)	4 (10.3)	35 (89.7)	13 (33.3)	26 (66.7)
Dermatology	4 (6.1)	62 (93.9)	5 (7.6)	61 (92.4)	24 (36.4)	42 (63.6)
Endocrinology / Metabolism	0 (0.0)	43 (100.0)	1 (2.3)	42 (97.7)	7 (16.3)	36 (83.7)
Gastroenterology	1 (5.9)	16 (94.1)	1 (5.9)	16 (94.1)	3 (17.6)	14 (82.4)
Immunology	0 (0.0)	8 (100.0)	0 (0.0)	8 (100.0)	1 (12.5)	7 (87.5)
Infectious Diseases	0 (0.0)	50 (100.0)	1 (2.0)	49 (98.0)	14 (28.0)	36 (72.0)
Neurology	3 (5.1)	56 (94.9)	3 (5.1)	56 (94.9)	14 (23.7)	45 (76.3)
OBGYN	0 (0.0)	10 (100.0)	0 (0.0)	10 (100.0)	4 (40.0)	6 (60.0)
Oncology and Hematology	0 (0.0)	40 (100.0)	0 (0.0)	40 (100.0)	9 (22.5)	31 (77.5)
Ophthalmology	0 (0.0)	17 (100.0)	2 (11.8)	15 (88.2)	7 (41.2)	10 (58.8)
Orthopedics	0 (0.0)	4 (100.0)	1 (25.0)	3 (75.0)	0 (0.0)	4 (100.0)
Other / Combination of two categories	0 (0.0)	32 (100.0)	0 (0.0)	32 (100.0)	8 (25.0)	24 (75.0)
Psychiatry	0 (0.0)	34 (100.0)	0 (0.0)	34 (100.0)	12 (35.3)	22 (64.7)
Pulmonology / Rheumatology	0 (0.0)	43 (100.0)	1 (2.3)	42 (97.7)	13 (30.2)	30 (69.8)
Urology	1 (8.3)	11 (91.7)	1 (8.3)	11 (91.7)	4 (33.3)	8 (66.7)
COVID study? (Yes) n (%)	1 (3.1)	31 (96.9)	1 (3.1)	31 (96.9)	5 (15.6)	27 (84.4)
Phases, n (%)						
Phase 1	1 (5.0)	19 (95.0)	1 (5.0)	19 (95.0)	5 (25.0)	15 (75.0)
Phase 2	0 (0.0)	157 (100.0)	1 (0.6)	156 (99.4)	47 (29.9)	110 (70.1)
Phase 3	0 (0.0)	204 (100.0)	4 (2.0)	200 (98.0)	47 (23.0)	157 (77.0)
Phase 4	1 (3.8)	25 (96.2)	1 (3.8)	25 (96.2)	7 (26.9)	19 (73.1)
Other	0 (0.0)	7 (100.0)	1 (14.3)	6 (85.7)	3 (42.9)	4 (57.1)
Not applicable	12 (17.9)	55 (82.1)	13 (19.4)	54 (80.6)	26 (38.8)	41 (61.2)
Funding, n (%)						
Government	0 (0.0)	11 (100.0)	0 (0.0)	11 (100.0)	2 (18.2)	9 (81.8)
Industry	12 (2.7)	435 (97.3)	19 (4.3)	428 (95.7)	126 (28.2)	321 (71.8)
Other	2 (8.7)	21 (91.3)	2 (8.7)	21 (91.3)	7 (30.4)	16 (69.6)
Study Location, n (%)						
US/International	5 (2.0)	247 (98.0)	7 (2.8)	245 (97.2)	69 (27.4)	183 (72.6)
US	4 (2.1)	186 (97.9)	7 (3.7)	183 (96.3)	52 (27.4)	138 (72.6)
NJ	5 (12.8)	34 (87.2)	7 (17.9)	32 (82.1)	14 (35.9)	25 (64.1)

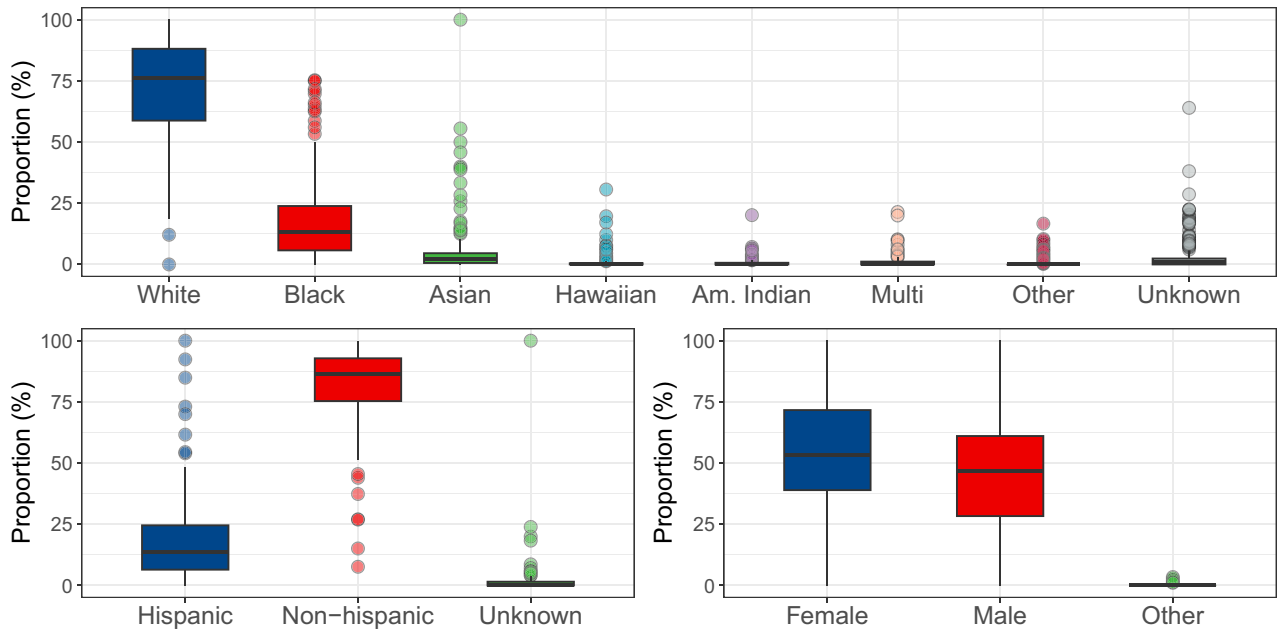


Fig. 2. Summary of the racial, ethnic, and sex proportions in trials conducted in the US. Standard box and whisker plots depict medians, 1st and 3rd quartiles (horizontal lines within colored boxes), data points located 1.5 times interquartile range from 1st and 3rd quartile (whisker ends), and data beyond that distance (circles; possible outliers).

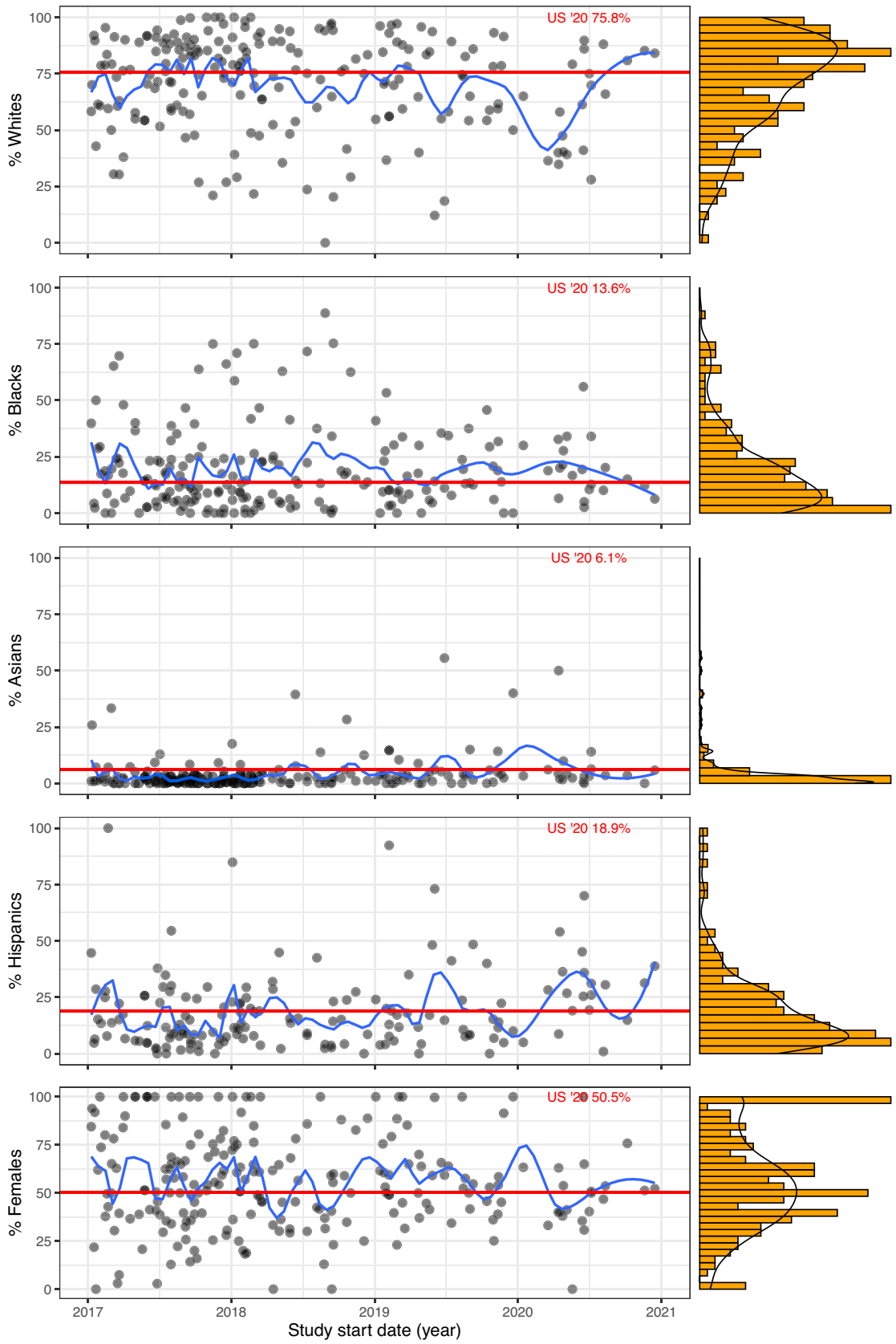


Fig. 3. Variability in racial, ethnic, and sex proportions in the clinical trials (US locations only).
The left panel shows the variability of racial, ethnic, and sex proportions in examined clinical trials that started in the period between 2017 and 2021 in the US. The right panel shows histograms of distributions of proportions in clinical trials. Each Black symbol represents a clinical trial conducted in the US. Red line indicates the percentage of interest from the 2020 US census. Blue line is a LOESS curve, which is a smoothed curve used here to qualitatively assess patterns in the data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Summary of meta-analysis (US locations only).

Overall meta-analysis					
	N studies	N in group / N recruited	% (95% CI)	2020 US Census	I^2 (95% CI)
White	215	39,420/ 54,492	74.9 (72.0–77.7)	75.8	97.1% (96.8–97.3)
Black	213	10,199/ 54,492	15.1 (13.0–17.4)	13.6	96.6% (96.4–96.9)
Asian	203	1,767/ 54,492	2.9 (2.4–3.4)	6.1	90.5% (89.4–91.4)
Hispanic	162	7,859/ 44,682	15.0 (12.9–17.3)	18.9	95.9% (95.5–96.2)
Female	229	34,286/ 56,849	58.9 (54.5–63.1)	50.5	97.0% (96.8–97.2)

For each population category, the table lists the number of clinical trials, the number of participants in each category (N in group), the total number of recruited participants (N recruited), the estimated percentage with 95% confidence interval, the 2020 US census percentage, and the 2020 US census value, and a heterogeneity statistic I^2.

Comparisons of group proportions to US census

Location

Whites, in trials conducted in New Jersey, constituted 61.9% (95% CI: 48.9%–73.5%) of participants, which was below the 2020 US census population, 75.8%. However, in studies including other US states the proportion was 76.7% (95% CI: 73.9%–79.3%) which was not different from the census. Blacks were overrepresented in New Jersey trials, 22.7% (95% CI: 14.4%–33.9%) compared to 13.6% from the census, but not in other US states, 13.8% (95% CI: 12.1%–16.3%). Asians were underrepresented in other US states 2.7% (95% CI: 2.3%–3.2%) compared to national 6.1%, but not in New Jersey 3.9% (95% CI: 2.3%–6.7%). Hispanics were slightly underrepresented in other US states 15% (95% CI: 12.7%–17.6%) compared to the census 18.9%, but not in New Jersey 15.1% (95% CI: 10.4%–21.3%).

Therapeutic area

As could be seen in Fig. 5 and Supplementary Table 3, Whites were underrepresented in Psychiatry and Infectious disease (55.7% and 54.5%, respectively) compared to the 2020 US census (75.8%) and overrepresented in Dermatology and Cardiovascular disease (90.0% and 88.6%, respectively). Blacks were underrepresented in Dermatology (5.8%) and Cardiovascular disease (7.7%) and overrepresented in Psychiatry (31.6%), Orthopedics (22.5%) and Infectious disease (25.1) compared to the 2020 US census (13.6%). Asians were underrepresented in Pulmonology and Rheumatology (1.5%), Psychiatry (3.2%), Neurology (2.3%), Immunology (1.9%), Dermatology (3.1%), Cardiovascular disease (0.9%), and Anesthesia/Analgesia (2.5%) compared to the 2020 US census (6.1%). Hispanics were underrepresented in Psychiatry (12.9%), Ophthalmology (11.8%), Oncology/Hematology (9.4%), Immunology (6.7%), and Cardiovascular disease (3.6%) compared to 18.9% in the US population. Interestingly, Asians and Hispanics were *not* overrepresented in any therapeutic area. Females were underrepresented in Cardiovascular disease (36.6%) and overrepresented in Ophthalmology (65.7%), OBGYN (98.1%, naturally), Gastroenterology (70.2%), and Dermatology (63.5%) relative to the 2020 US census (50.5%).

COVID-19

Compared to the 2020 US census proportions, Whites (60.6%) and Females (44.6%) were underrepresented and Hispanics (32.0%) were overrepresented in COVID studies (Fig. 4 and Supplementary Table 3).

Phase

White, Black, and Female representation in different trial phases did not differ from the 2020 US census. Asians were underrepresented in Phase 1, 2, and 3 clinical trials (2.0%, 3.1%, and 2.4%, respectively). Hispanics were underrepresented in Phase 3 (14.3%) and 4 (11.5%) clinical trials (Supplementary Table 3).

Funding

Whites were underrepresented in clinical trials in which sponsorship did not originate from Industry or Government (52.6%). Representation of Blacks and Females in trials sponsored by Government, Industry, or Other did not statistically differ from the 2020 US census. Asians and Hispanics were underrepresented in Industry sponsored trials (Supplementary Table 3).

Sources of heterogeneity: subgroup meta-analysis

As we expected, there was a high degree of heterogeneity in proportions of all three major racial groups, Hispanics, and Females (I^2 values are larger than 92%; Table 2). In order to understand the sources of heterogeneity, we explored the relationship between study-level

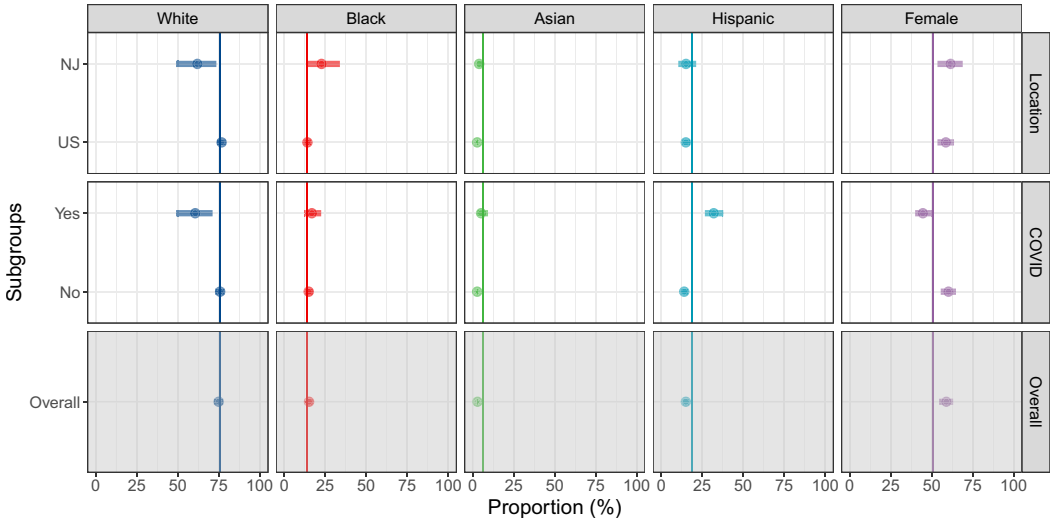


Fig. 4. Modified forest plots for overall and subgroup meta-analysis, Location and COVID-19.
Vertical lines indicate the 2020 US census percentages. Circles and bars indicate overall or subgroup (Location and COVID) estimates of population group proportions and 95% confidence intervals, respectively, from the random-effects model.

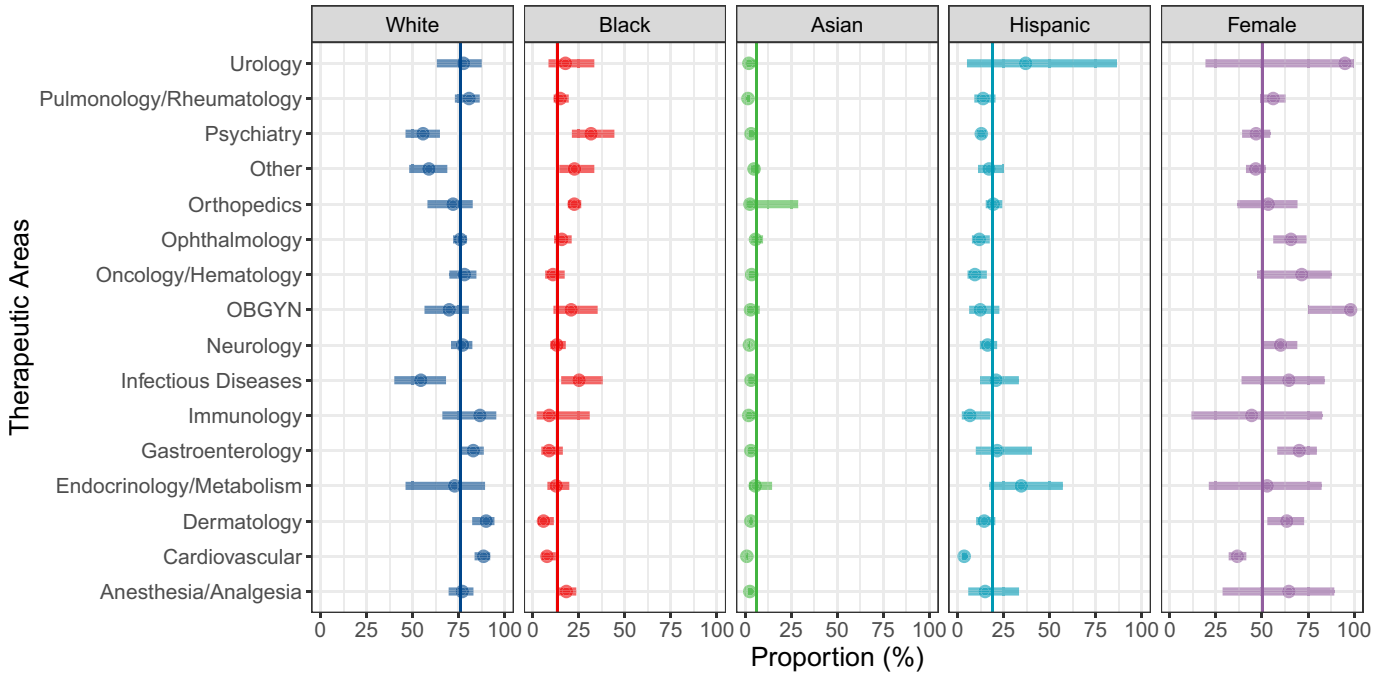


Fig. 5. Proportions by Therapeutic area.
Modified forest plots for subgroup meta-analysis, Location and COVID-19. Vertical lines indicate the 2020 US census percentages. Circles and bars indicate overall or subgroup (Therapeutic area)) estimates of population group proportions and 95% confidence intervals, respectively, from the random-effects model.

characteristics and the outcome proportions by conducting a separate subgroup meta-analysis for each study-level characteristic: location, therapeutic area, COVID-19, phase, and source of funding. Supplementary Table 3 and Figs. 4 and 5 show results of these subgroup analyses.

Examination of clinical trials conducted solely in New Jersey, indicates that the proportion of Whites (61.9%) was lower compared to trials conducted in other US states (76.7%). Proportion of Whites also varied by therapeutic area. For example, Psychiatry trials, on average, had the smallest proportion of Whites, 55.7%, compared to the largest proportion of Whites, in Dermatology trials, 90%. Proportion of Whites was lower in COVID-19 trials compared to non-COVID-19 trials (60.6% vs. 75.8%, respectively). Proportions of Whites were similar in trials of different phases. Trials funded by 'Other' sources exhibited a lower proportion of Whites compared to Industry funded trials (52.6% vs. 77.3%).

Similarly to Whites, the proportion of Blacks also differed between New Jersey and other US states (22.7% vs. 14.1%). The lowest proportion of Blacks was observed in Dermatology trials, 5.8% and the highest in Psychiatry, 31.6%. Overall, the proportion was different between therapeutic areas and was not different between COVID-19 and non-COVID-19 studies, study phases, and funding sources.

Contrary to Whites and Blacks, proportions of Asians, Hispanics, and Females did not differ between New Jersey and other US states (Supplementary Table 3 and Fig. 4). The lowest proportion of Asians was noted in cardiovascular studies, 0.9% and the highest in Endocrinology/Metabolism and Ophthalmology, 5.5% and 5.6%, respectively. Again, the proportion of Asians was different between therapeutic areas and was not different between COVID-19 and non-COVID-19 studies, study phases, and funding sources.

The lowest proportion of Hispanics was observed in cardiovascular studies, 3.6% and the highest in Endocrinology/Metabolism and Urology, 24.6% and 37.0%, respectively. COVID-19 studies exhibited a much higher proportion of Hispanics compared to non-COVID-19 studies (32.0% vs. 14.0%). Proportions of Hispanics were different between the study phases but not between the funding sources.

The lowest proportion of Females was observed in Cardiovascular studies, 36.6% and the highest in Urology and OBGYN, 95.1% and 98.1%, respectively. COVID-19 studies exhibited a lower proportion of Females compared to non-COVID-19 studies (44.6% vs. 60.2%). Proportions of Females were not different between the study phases and the funding sources.

We conclude that some of the high variability in racial, ethnic, and sex proportions in clinical trials could be explained by study-level characteristics, such as Location, Therapeutic area, COVID-19, and Funding sources. However, high heterogeneity still remains in the subgroups, as evident by large I^2 values. This may be due to local clustering of various populations in specific geographical areas from which participants are recruited.

Discussion

In an attempt to understand the racial and ethnic representation of clinical trials in the state of New Jersey, we conducted an exhaustive review of all registered clinical trials that: i. had results posted; ii. were initiated after the Common Rule mandate in 2017; and iii. had at least one study site located in the state of New Jersey. Our analysis included 481 clinical trials and revealed that since the regulatory changes took effect, the majority of applicable trials report sex (100% of identified trials), race (95.6%), and ethnicity (71.9%).

We divided the identified New Jersey-initiated trials into those that took place only in the US ($N = 229$) and those that had at least one international site ($N = 252$). We then compared the former group's demographic numbers to those of the 2020 US Census to better understand the trials' participants as compared to the US population as a whole.

Looking at the 229 US trials, 96.5% reported on race and/or ethnicity, and collectively included 76.7% White, 14.1% Black/African American, and 2.7% Asian participants. When ethnicity and sex were analyzed, these trials included 15% Hispanic and 58.6% Female participants.

Comparing to previous studies, the studies identified in this analysis all reported sex, compared to the 89.3% of COVID studies as described by Xiao et al. [22], and showed an improved reporting on race/ethnicity as contrasted to the results reported by Turner et al. [1], in which they

reported only 43% of studies provided race/ethnicity data.

We argue that the 2017 Common rule changes impacted reporting and that mandatory reporting - among other factors - has started to impact representation. In addition to these regulatory changes, the observed shift can also be attributed to the systematic efforts by the federal government to encourage minority representation in clinical trials [23,24].

Our argument is further supported when looking at the COVID-19 related trials. Within our search, we identified a sub-group of 32 studies, 31 of which (96.9%) reported race and 27 reported ethnicity (84.4%).

During the COVID-19 pandemic in the US, racial and ethnic minority groups were disproportionately impacted in terms of risk of exposure and infection, development of severe symptoms, need for hospitalization, and fatality rates [25–28] The largest disparities occurred between May and July 2020 in all census regions, with Hispanic/Latino individuals being overrepresented among hospitalized COVID-19 patients, a trend which continued through the end of 2020 [29]. However, the racial and ethnic disparities became less pronounced as the number of hospitalized White COVID-19 patients increased over time [29]. It is important to note that the COVID-19 clinical trials in New Jersey examined here all took place in 2020, and consisted entirely of drug trials. Of the total trials ($n = 32$), 25 (78.1%) started before the end of July; 2 (6.25%) in August; and the remaining 5 (15.6%) from October to December. Interestingly, the percentages of racially and ethnically diverse participants in COVID-19 trials (analyzed here) were higher than in non-COVID-19 trials; more importantly, the numbers indicate that the COVID-19 trials do not exhibit racial and ethnic disparities. This improvement is likely due to several factors, including the high numbers of COVID-19 patients from racial and ethnic minority groups during that period and the intentional efforts by pharmaceutical companies to ensure participation of these patient populations [30,31].

The lack of clarity regarding the funding source in many of the studies included in our analysis makes it difficult to assess if funding actually impacts either reporting, or trial inclusion. In our sample, neither parameter seems to be dependent on the funding source, or at least whether the source is industry or non-industry. This is consistent with other previously reported studies that lack consensus on how much funding or the type of funding affects enrollment [1,32,33].

Representation of different populations is inevitably affected by multiple factors such as the disease studied, location, or the level of involvement needed from participants. For example, states with a high diversity index might have a higher proportion of one population or another. These factors, or their combinations, could explain the high heterogeneity observed among clinical trials. However, when trials taking place across the US are considered collectively as a whole, the target representation should be the one coinciding with the diversity of the US population.

By virtue of its design, our study was limited by the dataset used and the original search criteria, i.e., at least one New Jersey site; nevertheless, the overall number of studies analyzed provides adequate information and suggests a potential shift with respect to clinical trial enrollment and diversity. The main limitation remains the nature of the data employed which reflects the way information is captured by Clinicaltrials.gov. This becomes especially problematic in the distinction of White-Hispanic versus White-non Hispanic populations as this differentiation is not captured in the way data is collected. As discussed elsewhere [1], the lack of cross tabulation makes it challenging to obtain accurate data on ethnicity. Similarly, inconsistencies with race/ethnicity reporting could have also affected our results with some of the studies not including all groups or reporting customized groups. Also, aggregated data as presented uses the "other" option to summarize information not fitting in available structured categories which could introduce a level of confusion. Finally, as Clinicaltrials.gov does not provide a detailed breakdown of enrollment by location, it was impossible to obtain baseline reference data for international studies - hence, those studies were excluded from the second part of the analysis. Needless to say, this does not suggest that international studies are less important sources of information about research enrollment. Rather, it remains unclear whether international studies are more inclusive, for instance, by having access to higher percentages of a certain population (e.g., Asians in Asian countries) or due to more systematic efforts to include those populations as suggested elsewhere [34].

Regardless of these limitations, the main strength of our study is that it demonstrates the progress that has been - until now - only anecdotally reported regarding both reporting and representation. Limiting information to published studies introduces the potential for the well-described (publication) bias which is why data originating from ClinicalTrials.gov is a more reliable source. In addition, the methodology used allowed us to control for some factors that would normally limit generalizability of the data.

Equitable representation of all populations (based on sex, race, and ethnicity) in well-documented and fully transparent studies will go a long way to assure trust in the results. Importantly, achieving such a status would encourage the various groups to remain engaged with the clinical research enterprise - a truly laudable goal. Statistically validated, generalizable results would, in every way, strengthen the health care providers' approaches to treating patients - and could lead to a wealth of downstream opportunities for further research. A process by which sustainable improvements in the overall health of people across racial, ethnic, sex, and socio-economic lines are realized may be within reach. Our data suggests that trends are moving in the right direction; nevertheless, true target representation in clinical trials, i.e., one reflecting the actual US population, has not yet been achieved.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloepi.2023.100134.

CRediT authorship contribution statement

Elli Gourna Paleoudis: Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Zhiyong Han:** Conceptualization, Data curation, Formal analysis, Project administration, Writing – original draft. **Simon Gelman:** Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Hernan Arias-Ruiz:** Data curation, Investigation, Formal analysis. **Destiney Carter:** Investigation, Data curation, Formal analysis. **Jovan Bertrand:** Data curation, Formal analysis, Investigation. **Nicole Mastrogiovanni:** Data curation, Formal analysis, Investigation. **Stanley R. Terlecky:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare no competing/conflicting interests.

Data availability

Data included herein available from the corresponding authors upon appropriate request.

Acknowledgements

The authors would like to thank the leadership of the Hackensack Meridian *Health* Research Institute and the Hackensack Meridian School of Medicine for their support of this project.

The authors acknowledge and thank Dr. Yen-Hong Kuo (Hackensack Meridian Health) for helpful comments and suggestions regarding data analysis. They would also like to acknowledge Mr. Michael Oppenheim for his expert editorial review and administrative support.

References

[1] Turner BE, Steinberg JR, Weeks BT, Rodriguez F, Cullen MR. Race/ethnicity reporting and representation in US clinical trials: a cohort study. Lancet Reg Health Am 2022;11(100252). https://doi.org/10.1016/j.lana.2022.100252.

[2] U.S. Food and Drug Administration. Food and Drug Administration Amendments Act (FDAAA) of 2007. https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/food-and-drug-administration-amendments-act-fdaaa-2007 [Accessed 2nd March 2023].

[3] Federal Register. Clinical trials registration and results information submission. https://www.federalregister.gov/documents/2016/09/21/2016-22129/clinical-trials-registration-and-results-information-submission [Accessed 2nd March 2023].

[4] Hwang TJ, Brawley OW. New Federal incentives for diversity in clinical trials. N Engl J Med 2022;387(15):1347–9. https://doi.org/10.1056/NEJMp2209043.

[5] Cohen NA, Silfen A, Rubin DT. Inclusion of under-represented racial and ethnic minorities in randomized clinical trials for inflammatory bowel disease. Gastroenterology. 2022;162(1):17–21. https://doi.org/10.1053/j.gastro.2021.09.035.

[6] King Jr TE. Racial disparities in clinical trials. N Engl J Med 2002;346(18):1400–2. https://doi.org/10.1056/NEJM200205023461812.

[7] U.S. Food and Drug Administration. ClinicalTrials.gov - Notices of noncompliance and civil money penalty actions. https://www.fda.gov/science-research/fdas-role-clinicaltrialsgov-information/clinicaltrialsgov-notices-noncompliance-and-civil-money-penalty-actions [Accessed 3rd March 2023].

[8] U.S. Food and Drug Administration. Civil money penalties relating to the ClinicalTrials.gov data bank. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/civil-money-Penalties-relating-clinicaltrialsgov-data-bank [Accessed 3rd March 2023].

[9] Piller C. FDA and NIH let clinical trial sponsors keep results secret and break the law. https://www.science.org/content/article/fda-and-nih-let-clinical-trial-sponsors-keep-results-secret-and-break-law [Accessed 3rd March 2023].

[10] Adashi EY, Cohen IG. The FDA initiative to assure racial and ethnic diversity in clinical trials. J Am Board Fam Med 2023;36(2):366–8. https://doi.org/10.3122/jabfm.2022.220290R1.

[11] Stanbrook MB, Salami B. CMAJ's new guidance on the reporting of race and ethnicity in research articles. CMAJ. 2023;195(6):E236–8. https://doi.org/10.1503/cmaj.230144.

[12] ICMJE. Clinical trials. https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html [Accessed 27th February 2023].

[13] United States Census Bureau. Explore census data. https://data.census.gov/ [Accessed 3rd March 2023].

[14] National Institutes of Health. Minority health and health disparities: Definitions and parameters. https://www.nimhd.nih.gov/about/strategic-plan/nih-strategic-plan-definitions-and-parameters.html [Accessed 3rd March 2023].

[15] National Institutes of Health. Office of Management and Budget (OMB) Standards. https://orwh.od.nih.gov/toolkit/other-relevant-federal-policies/OMB-standards [Accessed 3rd March 2023].

[16] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377–81. https://doi.org/10.1016/j.jbi.2008.08.010.

[17] Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208. https://doi.org/10.1016/j.jbi.2019.103208.

[18] Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Chichester: John Wiley & Sons, Ltd; 2009.

[19] R Core Team. R: A language and environment for statistical computing. https://www.R-project.org; 2020 [Accessed 3rd March 2023].

[20] Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019;22(4):153–60. https://doi.org/10.1136/ebmental-2019-300117.

[21] Wickham H. ggplot2: Elegant graphics for data analysis. Use R! 2nd ed. Houston: Springer Nature; 2016.

[22] Xiao H, Vaidya R, Liu F, Chang X, Xia X, Unger JM. Sex, racial, and ethnic representation in COVID-19 clinical trials: a systematic review and meta-analysis. JAMA Intern Med 2023;183(1):50–60. https://doi.org/10.1001/jamainternmed.2022.5600.

[23] U.S. Food and Drug Administration. Collection of race and ethnicity data in clinical trials: guidance for industry and food and drug administration staff. https://www.fda.gov/media/75453/download [Accessed 3rd March 2023].

[24] U.S. Food and Drug Administration. FDA takes important steps to increase racial and ethnic diversity in clinical trials. https://www.fda.gov/news-events/press-announcements/fda-takes-important-steps-increase-racial-and-ethnic-diversity-clinical-trials [Accessed 3rd March 2023].

[25] Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. JAMA. 2020;323(24):2466–7. https://doi.org/10.1001/jama.2020.8598.

[26] Gold JA, Wong KK, Szablewski CM, et al. Characteristics and clinical outcomes of adult patients hospitalized with COVID-19 — Georgia, March 2020. MMWR Morb Mortal Wkly Rep 2020;69:545–50. https://doi.org/10.15585/mmwr.mm6918e1.

[27] Garg S, Kim L, Whitaker M, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 — COVID-NET, 14 states, March 1–30, 2020. MMWR Morb Mortal Wkly Rep 2020;69:458–64. https://doi.org/10.15585/mmwr.mm6915e3.

[28] Gold JA, Rossen LM, Ahmad FB, et al. Race, ethnicity, and age trends in persons who died from COVID-19 — United States, May–August 2020. MMWR Morb Mortal Wkly Rep 2020;69:1517–21. https://doi.org/10.15585/mmwr.mm6942e1.

[29] Romano SD, Blackstock AJ, Taylor EV, et al. Trends in racial and ethnic disparities in COVID-19 hospitalizations, by region — United States, March–December 2020. MMWR Morb Mortal Wkly Rep 2021;70:560–5. https://doi.org/10.15585/mmwr.mm7015e2.

[30] Guardian News & Media Limited. AstraZeneca reviews diversity in trials to ensure drugs work for all. https://www.theguardian.com/business/2022/may/25/astrazeneca-reviews-diversity-in-trials-to-ensure-drugs-work-for-all#:~:text=AstraZeneca%20reviews%20diversity%20in%20trials%20to%20ensure%20drugs%20work%20for%20all,-This%20article%20is&text=The%20pharmaceutical%20giant%20AstraZeneca%20is,women%20is%20a%20particular%20challenge [Accessed 3rd March 2023].

[31] Pfizer Inc. Doing better: Increasing diversity in clinical trials. https://www.pfizer.com/news/articles/doing_better_increasing_diversity_in_clinical_trials [Accessed 3rd March 2023].

[32] Kwiatkowski K, Coe K, Bailar JC, Swanson GM. Inclusion of minorities and women in cancer clinical trials, a decade later: have we improved? Cancer. 2013;119:2956–63. https://doi.org/10.1002/cncr.28168.

[33] 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(10):e191870. https://doi.org/10.1001/jamaoncol.2019.1870.

[34] Tahhan AS, Vaduganathan M, Greene SJ, et al. Enrollment of older patients, women, and racial and ethnic minorities in contemporary heart failure clinical trials: a systematic review. JAMA Cardiol 2018;3(10):1011–9. https://doi.org/10.1001/jamacardio.2018.2559.