Race/ethnicity reporting and representation in US clinical trials: A cohort study

Brandon E. Turner,[a,b,*] Jecca R. Steinberg,[a] Brannon T. Weeks,[a] Fatima Rodriguez,[c] and Mark R. Cullen[d]

[a]Stanford University School of Medicine, Stanford, CA, USA
[b]Massachusetts General Hospital, 55 Fruit Street, Lunder Building LL3, Boston, MA 02114, USA
[c]Division of Cardiovascular Medicine and the Cardiovascular Institute, Stanford University, Stanford, CA, USA
[d]Center for Population Health Sciences, Stanford University School of Medicine, Stanford, CA, USA

The Lancet Regional Health - Americas 2022;11: 100252
Published online 10 April 2022
https://doi.org/10.1016/j.lana.2022.100252

Summary

Background Systemic progress in improving trial representation is uncertain, and previous analyses of minority trial participation have been limited to small cohorts with limited exploration of driving factors.

Methods We analyzed detailed trial records from all US clinical trials registered in ClinicalTrials.gov from March 2000 to March 2020. Minority enrollment was compared to 2010 US Census demographic estimates using Wilcoxon test. We utilized logistic regression and generalized linear regression with a logit link to assess the association of possible drivers (including trials' funding source, size, phase, and design) with trials' disclosure of and amount of minority enrollment, respectively.

Findings Among 20,692 US-based trials with reported results (representing ~4·76 million enrollees), only 43% (8,871/20,692) reported any race/ethnicity data. The majority of enrollees were White (median 79·7%; interquartile range [IQR] 61·9–90·0%), followed by Black (10·0%; IQR 2·5–23·5%), Hispanic/Latino (6·0%; IQR 0·43–15·4%), Asian (1·0%; IQR 0·0–4·1%), and American Indian (0·0%; IQR 0·0–0·2%). Median combined enrollment of minority race/ethnicity groups (Black, Hispanic/Latino, Asian, American Indian, Other/Multi) was below census estimates (27·6%) ($p < 0·001$) however increased at an annual rate of 1·7%. Industry and Academic funding were negatively associated with race/ethnicity reporting (Industry adjusted odds ratio [aOR]: 0·42, 95% confidence interval [CI]: 0·38 to 0·46, $p < 0.0001$; Academic aOR: 0·45, CI: 0·41 to 0·50, $p < 0.0001$). Industry also had a negative association with the proportion of minority ethnicity enrollees (aOR: 0·69, CI: 0·60 to 0·79) compared to US Government-funded trials.

Interpretation Over the past two decades, the majority of US trials in ClinicalTrials.gov do not report race/ethnicity enrollment data, and minorities are underrepresented in trials with modest improvement over time.

Funding Stanford Medical Scholars Research Funding, the National Heart, Lung, and Blood Institute, NIH (1K01HL144607) and the American Heart Association/Robert Wood Johnson Medical Faculty Development Program.

Keywords: Racial disparities; Ethnicity; Clinical trials; Public Policy; Research funding; Industry funding; Government funding; Health Policy

Introduction

In 1993, the United States (US) Congress passed the National Institutes of Health Revitalization Act as part of an effort to improve enrollment of minority groups in clinical trials.[1] Multiple academic and government initiatives to increase inclusion followed.[2] Despite these efforts, more than twenty-five years later minority racial/ethnic groups remain underrepresented and racism remains an international public health crisis.[2,3] The Covid-19 vaccine trials underscored these disparities; Despite Black individuals representing 21% of Covid-19 deaths, they comprised only 3% of major vaccine trial participants.[4] Other minority populations were similarly underrepresented.[5]

The paucity of diversity in clinical trials generates a racial/ethnic data gap that skews medical evidence and innovation towards therapies with understudied efficacy and safety for minority populations.[2,6] Data generated from investigations that lack racial/ethnic diversity formalize a biased framework of "normal" and "diseased" biological variants which subsequently become

*Corresponding author at: Massachusetts General Hospital, 55 Fruit Street, Lunder Building LL3, Boston, MA 02114, USA.
E-mail address: brandon_turner@dfci.harvard.edu (B.E. Turner).

Research in context

Evidence before this study

US trials under-enroll racial/ethnic minorities and frequently do not report their minority enrollment at all. This disparity contributes to biased medical evidence and excludes minorities from the benefits of clinical trial participation. Little is known about changes in minority representation over time and how it relates to trial characteristics such as funding. We searched PubMed using the terms "race", "racial", "ethnic*", "minorit*", "divers*", "trials", and "ClinicalTrials.gov" to identify studies in any language analyzing minority participation in clinical trials, published from any date up to 24 January 2022. Effectively all studies examined small subsets of trials using varying sampling strategies. Most studies tabulated published journal results from a small subset of journals. Some studies searched institutional or third-party datasets of clinical trials. The few studies that used ClinicalTrials.gov were restricted to specific diseases and narrow time periods with small sample sizes.

Previous studies found widely varying levels of race/ethnicity reporting and representation. Few had sufficient size or scope to permit investigation of trends over time. Those that did investigate changes over time found inconsistent trends. Studies also found differing associations with industry and governmental funding. Studies seldom conducted multivariable analyses to control for other factors. The divergent methods and sampling strategies among studies prevent effective comparison between their findings.

Added value of this study

We analyzed both the reporting of race/ethnicity enrollment and the representation of diverse groups using all United States trials in the ClinicalTrials.gov registry from 2000 to 2020. To our knowledge, this is the largest study of racial/ethnic diversity in clinical trials. The longitudinal data source and large cohort enabled us to increase generalizability and to assess multivariable relationships along with temporal trends. Our analysis also included data from unpublished trials which are absent in previous analyses using journal samples.

Implications of all the available evidence

Our results show that race/ethnicity reporting is historically poor but has improved to a high level in recent years. We clarify that while enrollment of minority race/ethnicity participants remains poor, it is in fact modestly increasing. Positive trends in both race/ethnicity reporting and representation over time may reflect the impact of the various initiatives to improve minority recruitment. Industry-funded trials enrolled the least diverse participants and US government-funded trials enrolled the most diverse. However, after controlling for other factors, we show that the influence of funding varies across race/ethnicity groups and is most pronounced for Blacks. All stakeholders must commit to consistent and transparent results reporting to enable innovative solutions for the recruitment of cohorts that are representative of the US as a whole.

propagated through future research and precision therapies.[6,7] Furthermore, underrepresented populations lack access to the health benefits conferred through trial participation.[2] An international reckoning regarding racial discrimination has added urgency to current efforts to evaluate progress with improving diversity and inclusion within health research. Though randomized trials and trial meta-analyses comprise the gold standards for evidence generation, the overall state of diversity in trials, progress over time, and associated factors are poorly understood.

Two major barriers to adequate surveillance have included insufficient race/ethnicity enrollment reporting by trialists and the difficulty in collecting data from a sufficient number of trials to permit robust analyses.[2] Previous studies of racial/ethnic representation have employed varied sampling strategies, usually within subsets of journals and/or selected publications for a specific disease.[2,8] These approaches have produced divergent assessments of the race/ethnicity landscape including frequent discrepancies in the literature regarding minority enrollment trends[8–10] progress with race/ethnicity reporting[2,9–11] and whether industry and government funding influence these outcomes.[9,12] Effective comparison between previous studies is further undermined by their disparate data sources. Relatedly and potentially due to insufficient sample size, previous studies seldom controlled for other factors or performed multivariable analyses.[10]

In this study, we conducted an analysis using available data from all registered trials in the ClinicalTrials.gov registry from 2000 to 2020 that were conducted within the US. We aimed to investigate temporal trends in race/ethnicity reporting and enrollment in US clinical trials, compare racial/ethnic enrollment in clinical trials to the US population census, and identify trial features associated with greater reporting/representation of minorities.

Methods

Data sources

We used the Aggregate Analysis of ClinicalTrials.gov database to download records of clinical studies and all results submitted to ClinicalTrials.gov between 1 March 2000 and 9 March 2020.[13] We limited our analysis to trials conducted exclusively in the US to ensure consistent race/ethnicity definitions across trials and in acknowledgement of the unique US context of complex race/ethnicity relations that extends to the US history of medicine and clinical trials.[14,15] We referenced the 2010 US Census database for US population statistics.[16] We selected the US Census to capture the population with the potential to be afflicted by health issues in the United States. By selecting the entire US population, we attempt to avoid convenience sampling which often biases studies that only examine easily accessed populations (e.g. insured individuals, individuals within a

specific zip code). This study was reviewed by the Institutional Review Board at Stanford University School of Medicine and exempted from oversight as it was not human-subject research.

Definitions and study variables

"Race/ethnicity" was defined in accordance with US Census and Department of Health and Human Services guidelines (*Supplemental Panel* S1).[16] We detail our race, ethnicity, and gender extraction approach in the supplement (*Supplemental Panel* S1). Briefly, we selected enrollment data for five racial/ethnic categories, which we refer to as "all five groups": to align with the most common racial/ethnic groups reported in trials: White, Hispanic/Latino, Black, Asian (including Pacific Islander and Native Hawaiian), and American Indian (including Alaskan Native).

The reporting of clinical trial enrollees' race/ethnicity is not strictly required for all trials in ClinicalTrials.gov. The September 2007 passage of Section 801 of the Food and Drug Administration Amendments Act (FDAAA 801) and the confirmation of the Final Rule (effective January 2017) expanded the number of trials with mandated reporting.[17] We review race/ethnicity reporting regulatory requirements and practices in the supplement (*Supplemental Panel* S2). To avoid potential confounding, we present analysis for trials submitted after FDAAA 801, 27 September 2007, however none of our major findings change when also including the cohort from 2000 to 2007 (provided in the *Supplemental Appendix*).

Exposure variables

Our primary exposure variable was funding (*Supplemental Panel* S3). We also explored 11 other trial features and their associations with race/ethnicity reporting and enrollment: (1) primary purpose (defined on ClinicalTrials.gov as "The main reason for the clinical trial. The types of primary purpose are: treatment, prevention, diagnostic, supportive care, screening, health services research, basic science, and other."[18]; (2) phase; (3) number of arms; (4) enrollment (defined as the number of participants in a clinical trial) (5) blinding; (6) randomization; (7) placebo-controlled or controlled with no intervention; (8) active comparator; (9) oversight by a data monitoring committee; (10) number of sites; (11) year of trial submission. All exposure variables aligned with ClinicalTrials.gov definitions and pre-established categories and we classified funding consistent with previous analyses (*Supplemental Panel* S3).[19]

Outcomes

The primary outcome was reporting of any race/ethnicity data. Secondary outcomes were the reporting of race/ethnicity data for individual racial/ethnic groups and trial diversity (the combined proportion of enrollees belonging to minority race/ethnicity groups (Black, Hispanic/Latino, Asian, American Indian, and Other/Multi).

Statistical analysis

We assessed differences in the distribution of exposure variables between trials which did or did not report race using Fisher's exact test. We assessed trends over time using compound annual growth rates and the Mann-Kendall significance tests.[20] These analyses provided details on changes over time in race/ethnicity reporting and representation. For descriptive temporal statistics (e.g. growth rates, trend statistics) we included only complete calendar years for which there were at least 100 trials with submitted results (2003–2018) to exclude years with inadequate data for year-to-year estimates to prevent skewing results (represents the removal of 193 trials out of 20,692). When analyzing trials after FDAAA 801 (September 2007), we use only trials from 2008 to 2018.

In the descriptive analyses, to account for inconsistent race/ethnicity reporting as a potential source of bias, we performed a sensitivity analysis and generated four models for calculating demographic estimates (*Supplemental Panel* S4).

We conducted multivariable logistic regression analysis of the association between funding and the primary outcome, controlling for the 11 other trial features in the multivariable models given our exploratory design. We also completed an inductive hypothesis-generating analysis of the 11 other trial features. In the inductive hypothesis-generating analysis other trial features and their associations were explored without a specific a-priori hypothesis.

For our trial diversity outcome (the combined proportion of participants from minority race/ethnicity groups), we performed generalized linear regression with a logit link[21] to identify the influence of funding and other trial features. For each race/ethnicity group we included all trials that reported the number of enrollees from that group. For this analysis we report relative differences in the estimated proportion of enrollment when all other variables are held constant at their reference level (*Supplemental Table* S5).

We treated all features as confounding variables in multivariable analysis except the feature in consideration as the exposure variable.

All analyses were two-sided. We set statistical significance at $\alpha = \cdot 05$ level. We analyzed all data using R version 3.5.2.

Missing data in multivariable regression analysis

Because ClinicalTrials.gov does not require completion of all fields when submitting a trial record, some records have missing data. We assume missing values among

our 12 exposure variables are missing at random. The number of missing elements per variable varied between 0 and 7·9%. We handled missing data for our regression analyses using multiple imputation by chained equations using the mice 3.0 package and the "mice" function.[22] We generated 30 imputed data sets. We used Bayesian logistic regression to estimate missing binary data and Bayesian multinomial logistic regression to estimate missing categorical data (our sample did not contain missing continuous data). Parameter estimates were pooled using Rubin's rules.

Role of the funding source

The funding source had no role in the study design, data collection, data analysis, data interpretation, or writing of the report, nor have they had access to the analyzed data or completed manuscript. The funding sources supported the two manuscript authors JRS and FR. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results

Trial population and general characteristics

From the 328,452 clinical studies registered between 1 March 2000 and 9 March 2020, we identified 20,692 US-based clinical trials with reported results (representing over 4·76 million enrollees; Figure 1). From this cohort, 8,871/20,692 (43%; approximately 2·09 million enrollees) reported any race/ethnicity enrollment data, including 7792/16780 of trials since FDAAA 801 in 2007 (46%; 1·84 million enrollees).

Trials with and without race/ethnicity reporting had less than a 5% difference in trial feature distributions for 9 of our 12 features (Table 1). The greatest differences occurred in rates of US Government-funding (25% vs 16%) and multisite trials (39% vs 31%).

Race/ethnicity reporting

The proportion of trials that reported race/ethnicity data within ClinicalTrials.gov increased for trials registered after the mandated creation of the ClinicalTrials.gov results database in 2007 (Figure 2a, *Appendix Table* S1). From 2008 to 2018 reporting of any race/ethnicity enrollment data increased from 26% (599/2334) to 91% (194/213) (annual growth rate 13·5%), compared to only 11% (248/2334) to 41% (87/213) (annual growth rate 14·4%) for reporting all five groups (*Appendix Table* S2). We found similar growth patterns for each race/ethnicity group (*Appendix Figure* S2). Overall since 2007, 45% (8088/17,886) of trials reported some race/ethnicity data and 22% (3780/17,886) reported data for all five groups. In comparison, 98% (17,448/17,886) of trials reported participants' sex over the same period (Figure 2a).

The majority of trials that reported any race/ethnicity data reported White and Black enrollment (95% (8413/8871) and 92% (8134/8871), respectively; Figure 2b, *Appendix Table* S3). Latino enrollment was the least commonly reported (62% (5517/8871)), followed by American Indian (77% (6748/8871)) and Asian (84% (7411/8871)). Only 47% (4105/8871) of clinical trials with some race/ethnicity reporting reported on all five groups.

Representation of different race/ethnicity groups among enrollees

In trials reporting race/ethnicity enrollment for all five groups, the majority of participants were White, with a median White enrollment of 79·7% (Interquartile Range [IQR] 61·9–90·0%, Figure 3, *Supplemental Table* S4). This exceeded the US Census estimate of White representation among the US population (72·4%, $p < 0·001$). Latinos, Asians, and American Indians were all underrepresented compared to their US populations ($p < 0·001$), with the largest discrepancy observed for Latinos (median 6·0%, IQR 0·4–15·4%; US Census population 16·3%). The median enrollments of Blacks did not reach significance (median 10·0%, IQR 2·5–23.5%, Census 12·6%), however 21% (856/4105) of trials reported 0 Black enrollees (compared to 44% (1807/4105), 25% (1027/4105), 74% (3035/4105) and 2% (76/4105) for Asian, Latino, American Indian, and White enrollees, respectively). 10% (412/4105) of trials reported 100% White enrollment.

These findings were largely consistent across the alternative coding rules for race/ethnicity in our sensitivity analysis (*Appendix Panel* S5, *Figure* S2 and *Table* S8).

Trial features associated with race/ethnicity reporting

In multivariable regression analysis, the strongest association with race/ethnicity reporting was funding source: Industry and Academic trials had significantly lower odds of reporting race/ethnicity data compared to US Government-funded trials (Academic adjusted Odds Ratio (aOR) 0·47, 95% confidence interval [CI] 0·44 to 0·52, $p < 0.0001$; industry aOR 0·45, CI 0·41 to 0·49, $p < 0.0001$; *Appendix Table* S5).

Among other trial features, each additional year between 2008 and 2020 was associated with 1·39 (CI 1·38 to 1·41) greater odds of race/ethnicity reporting. Relative to Phase 2/3-3 trials, Phase 1 and Phase 1/2-2 trials had higher odds of reporting race/ethnicity (aOR 1·72, CI 1·54 to 2·16; aOR 1·27, 95% CI 1·21 to 1·44, respectively). In contrast, trials with no applicable phase and Phase 4 trials had lower odds of reporting race/ethnicity than Phase2/3-3 (aOR 0·77, 95% CI 0·67 to 0·87;

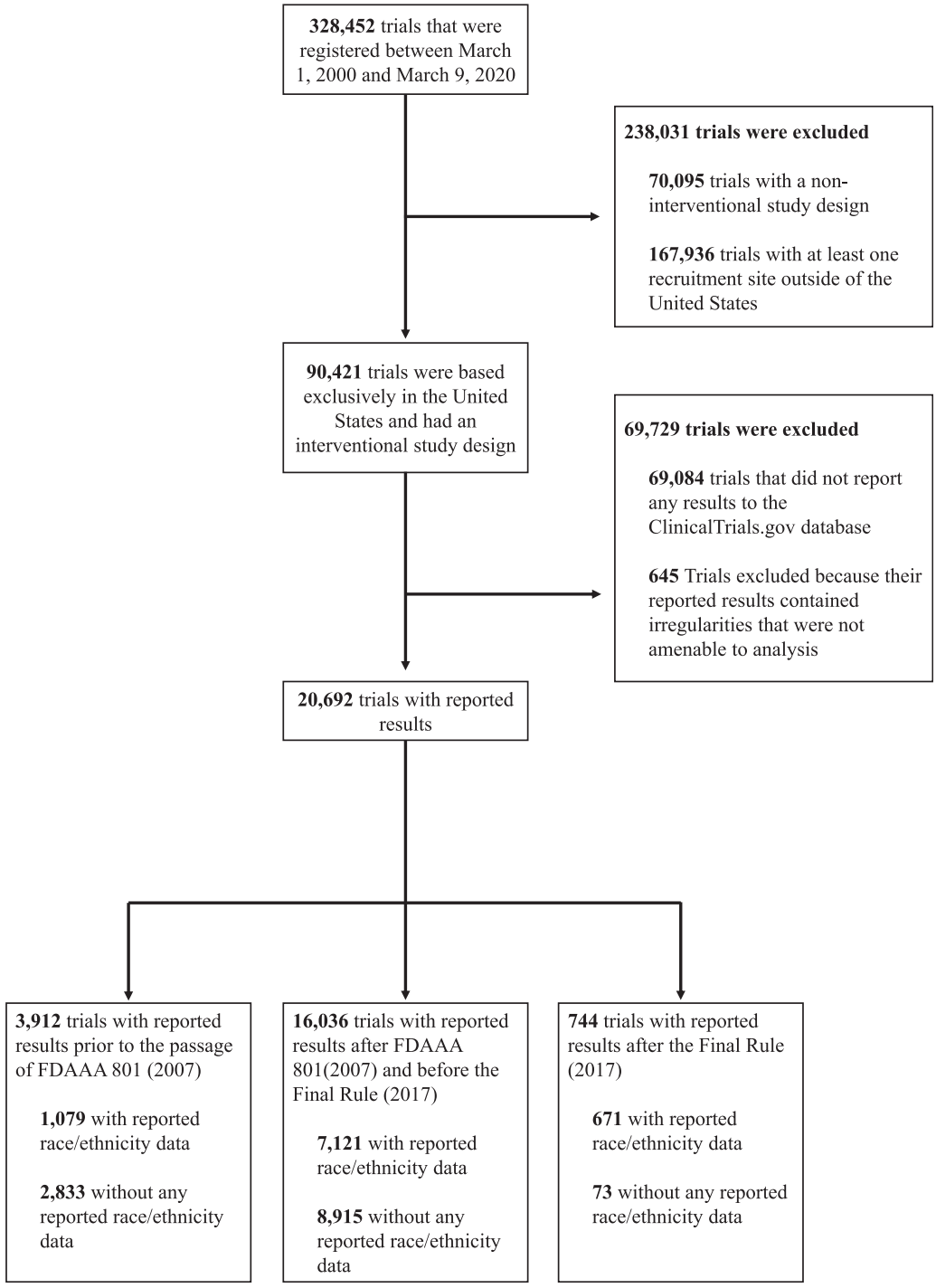

Figure 1. CONSORT Diagram of Clinical Trials Included in the Analysis.

The FDAAA 801 refers to the 27 September 2007 enactment of Section 801 of the Food and Drug Administration Amendments Act which required a subset of US phase 2-4 trials to report results in ClinicalTrials.gov and established reporting guidelines for clinical trials. The Final Rule refers to the 18 January 2017 Food and Drug Administration clarification and expansion of reporting requirements for clinical trials in Clinicaltrials.gov including race/ethnicity reporting. Irregularities that were not amenable to analysis included data entry errors and illogical data responses.

Trial Feature	Total* *n* (%)	Clinical trials with and without Race Reporting*,§ *n* (%)	
		No	Yes
Funding§			
Industry	7717 (46·0)	4345 (48·3)	3372 (43·3)
Academic	5669 (33·8)	3202 (35·6)	2467 (31·7)
US Government	3394 (20·2)	1441 (16·0)	1953 (25·1)
Primary Purpose			
Treatment	11,877 (70·8)	6361 (70·8)	5516 (70·8)
Basic Science	791 (4·7)	403 (4·5)	388 (5·0)
Prevention	1289 (7·7)	688 (7·7)	601 (7·7)
Other†	2319 (13·8)	1196 (13·3)	1123 (14·4)
Missing	504 (3·0)	340 (3·8)	164 (2·1)
Phase			
Not Applicable¥	5660 (33·7)	3132 (34·8)	2528 (32·4)
Phase 1	1316 (7·8)	563 (6·3)	753 (9·7)
Phase 1/2-2	5623 (33·5)	2913 (32·4)	2710 (34·8)
Phase 2/3-3	1844 (11·0)	979 (10·9)	865 (11·1)
Phase 4	2337 (13·9)	1401 (15·6)	936 (12·0)
Number of Trial Arms			
1	5185 (30·9)	2920 (32·5)	2265 (29·1)
2	8313 (49·5)	4331 (48·2)	3982 (51·1)
≥3	3200 (19·1)	1668 (18·6)	1532 (19·7)
Missing	82 (0·5)	69 (0·8)	13 (0·2)
Enrollment			
0—9	2243 (13·4)	1341 (14·9)	902 (11·6)
10—49	7280 (43·4)	4033 (44·9)	3247 (41·7)
50—99	2958 (17·6)	1541 (17·1)	1417 (18·2)
100—499	3499 (20·9)	1698 (18·9)	1801 (23·1)
500—999	465 (2·8)	219 (2·4)	246 (3·2)
≥1000	335 (2·0)	156 (1·7)	179 (2·3)
Blinding			
None	9500 (56·6)	5099 (56·7)	4401 (56·5)
Double	4946 (29·5)	2686 (29·9)	2260 (29·0)
Single	2325 (13·9)	1201 (13·4)	1124 (14·4)
Missing	9 (0·1)	2 (0·0)	7 (0·1)
Randomization			
Non-Randomized	6467 (38·5)	3601 (40·1)	2866 (36·8)
Randomized	10,242 (61·0)	5335 (59·4)	4907 (63·0)
Missing	71 (0·4)	52 (0·6)	19 (0·2)
Placebo-controlled or controlled with no intervention			
No	11,294 (67·3)	6055 (67·4)	5239 (67·2)
Yes	5404 (32·2)	2864 (31·9)	2540 (32·6)
Missing	82 (0·5)	69 (0·8)	13 (0·2)
Use of an Active Comparator			
No	11,020 (65·7)	5846 (65·0)	5174 (66·4)
Yes	5678 (33·8)	3073 (34·2)	2605 (33·4)
Missing	82 (0·5)	69 (0·8)	13 (0·2)
Oversight by a Data Monitoring Committee			
No	9041 (53·9)	4934 (54·9)	4107 (52·7)
Yes	6773 (40·4)	3491 (38·8)	3282 (42·1)
Missing	966 (5·8)	563 (6·3)	403 (5·2)
Number of Sites			
1	10,984 (65·5)	6193 (68·9)	4791 (61·5)
≥2	5796 (34·5)	2795 (31·1)	3001 (38·5)

Table 1 *(Continued)*

Trial Feature	Total* *n* (%)	Clinical trials with and without Race Reporting*,δ *n* (%)	
		No	Yes
Study Status			
Completed	13,358 (79·6)	7093 (78·9)	6265 (80·4)
Ongoing	338 (2·0)	76 (0·8)	262 (3·4)
Stopped early	3073 (18·3)	1812 (20·2)	1261 (16·2)
Unknown	11 (0·1)	7 (0·1)	4 (0·1)

***Table 1*: Characteristics of US-based clinical trials (*n* = 16,780) from September 2007 to March 2020 with and without race and ethnicity results reporting.**[Ω]

[Ω] September 27, 2007 aligns with the enactment on the FDAAA 801, the Food and Drug Administration Amendments Act which required that all US phase 2-4 intervention studies register in ClinicalTrials.gov and established reporting guidelines for clinical trials.

[δ] Fisher's exact test resulted in p-values < 0·001 for all trial features except for blinding ($p = ·008$).

[*] Percentages may not sum to 100 because of rounding.

[†] Other primary purposes include diagnostic, screening, supportive care, health services research and other.

[Ψ] On ClinicalTrials.gov "Not Applicable" is used to describe trials without Food and Drug Administration-defined phases, including trials of devices or behavioral interventions.

[§] Funding categories were determined with data on the sponsor and collaborators. Industry funding includes trials with an industry sponsor or collaborating agency US Government trials include remaining trials with a US Government sponsor or collaborating agency.

aOR 0·82, 95% CI 0·71 to 0·94, respectively). The presence of a data monitoring committee and multiple study sites were both associated with greater race/ethnicity reporting. Trials with larger enrollment showed an incrementally greater odds of reporting race/ethnicity, with an aOR 0·48 (95% CI 0·42 to 0·54) for trials with 0–9 enrollees and an aOR 1·30 (95% CI 1·08 to 1·56) for trials with 500–999 enrollees compared to the reference of 100–499 enrollees.

Trial features associated with enrollment of minority participants

The median combined proportion enrollees from minority race/ethnicity groups increased at an annual rate of 1·7% per year.

In multivariable regression, funding showed differing associations with enrollment by race/ethnicity (Figure 4; *Appendix Table* S6). Compared to Government trials, industry trials were associated with significantly less Black enrollment (adjusted relative difference [Δ] -26.0%; 95% CI -35.7% to -15.0%), and greater White (Δ 10.2%; 95% CI 6.5% to 13.5%) and Latino (Δ 16.8%; 95% CI -4.9% to 42.5%) enrollment. Academic trials showed a similar but less pronounced trend (Black Δ -12.4% 95% CI -23.4% to -0.2%; White Δ 2.3%, 95% CI -1.8% to 6.1%; Latino 18.5%, 95% CI -3.8% to 45.0%).

Most other trial features had similar associations with Black and Latino enrollment and the inverse relationship with White enrollment.

No features were clearly associated with the proportion of Asian or American Indian enrollment. However,

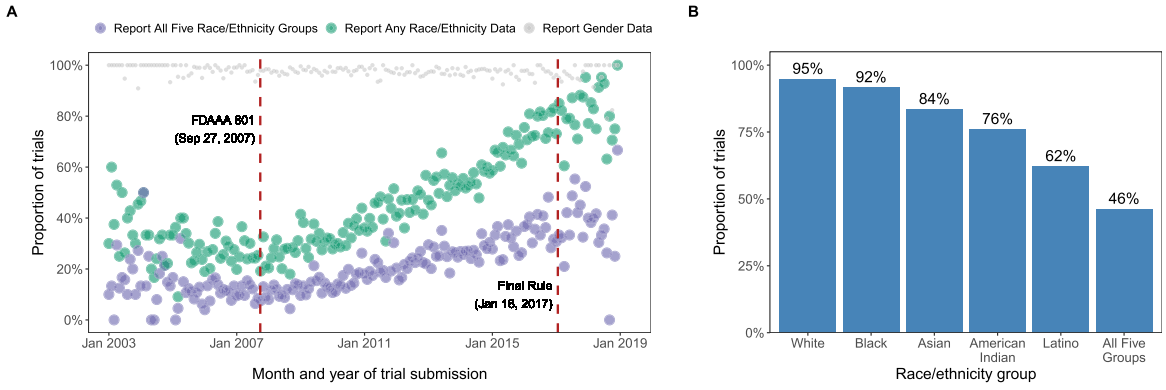

Figure 2. Race and ethnicity enrollment reporting in United States-based clinical trials registered on ClinicalTrials.gov.

Panel A shows change over time in proportion of trials reporting race/ethnicity enrollment data. All Five Race/Ethnicity Groups include White, Latino, Black, Asian (including Pacific Islander and Native Hawaiian), and American Indian (including Alaskan Native). *Panel B* shows the races/ethnicities that were reported among trials that included any race/ethnicity enrollment results data and the proportion of those trials that reported each individual race/ethnicity.

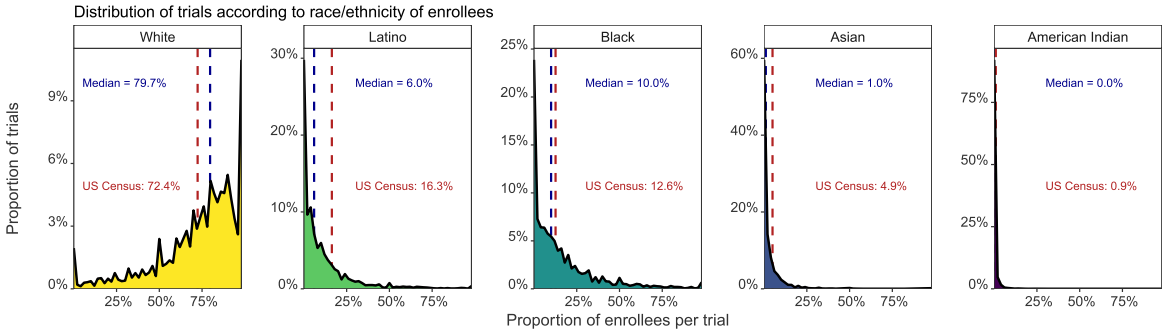

Figure 3. Race/ethnicity representation among all United States-based clinical trials with race/ethnicity enrollment data in Clinicaltrials.gov.

The graph shows the distribution of trials and the representation of each racial/ethnic group organized by racial/ethnic category. Distribution only includes trials that reported data for all five racial/ethnic groups. Census calculations reflect 2010 US Census data.

trials with small enrollments (<50) showed a trend toward greater enrollment for Whites and Asians, and lesser enrollment for Blacks and Latinos (Figure 4). Multi-centered trials were associated with greater White enrollment (Δ 5.3%, 95% CI 1.9% to 8.5%), and lesser Black enrollment (Δ -13.0%, 95% CI -22.8% to -2.3%).

Phase 1 and Phase 4 trials were associated with or trended toward greater enrollment of Blacks and Latinos and lesser enrollment of Whites relative to Phase 2/3-3 trials.

Discussion

In our analysis of two decades of data from over 20,000 USA-based clinical trials registered in ClinicalTrials.gov, we found that fewer than 44% of trials report any race/ethnicity data. Among trials that do report race/ethnicity, as a group minorities remain underrepresented compared to their US populations, though at the subgroup level median Black enrollment is not statistically below their US population. Even after controlling for other trial features and conducting several sensitivity analyses evaluating alternative encodings of race/ethnicity, we found that industry-funded trials were associated with less race/ethnicity reporting and with lower rates of minority race/ethnicity enrollment compared to US government-funded trials.

While the increase in race/ethnicity reporting is encouraging, it constitutes a low bar with fewer than 25% of trials reporting enrollment data for all five major race/ethnicity groups, in contrast with rates of sex reporting, which remained above 99% for all study years.[19] Reporting progress is further complicated by the large subset of completed trials (excluded from this analysis) that have not reported any trial results. Decades of research have highlighted the clinical and ethical concerns presented by selective reporting of trial results (despite legal reporting obligations for many trials).[23] Our data showed that smaller trials were less likely to report race/ethnicity data. While this may reflect trialists' assessment of the limited utility of tracking small populations within limited studies, the absence of transparent trial demographics obscures the generalizability of results and weakens the medical community's ability to identify gaps or opportunities for further research.

Our results suggest the lack of consensus among previous studies that examined race/ethnicity reporting rates (ranging 2-58% in one review[2]) may stem from a lack of power and/or differences among their sampled populations. For example, Kwiatkowski et al. found significantly improved race/ethnicity reporting over time in their study of 304 Phase 3, non-industry trials,[10] whereas Loree et al. found only minimal reporting changes in their study of 230 drug approval trials with 97% industry funding.[11] Our model showed industry funding was strongly associated with lower levels of reporting, which may explain the modest findings observed in the drug approval trials. Many studies do not state the funding mix of the trials in their samples, nor the mix of other features that may influence reporting rates. The large sample size and multivariable approach used in our study help resolve these discrepancies and may explain some of the discordance observed in the literature.

While we found a modest improvement over time in trial diversity, minorities remain underrepresented relative to their US populations. This paradigm persists despite many national and international efforts to increase and facilitate greater diversity in clinical trials including the NIH Revitalization Act,[1] Federal Drug Administration Race and Ethnicity Guidance[24] and International Conference on Harmonization Guidance.[25] Our sensitivity analyses showed our observation of minority underrepresentation over time was robust and most pronounced for Latinos and Asians: even in the most generous sensitivity models, median representation was less than half their respective US populations. Previous studies have highlighted the unique shortage of Latino enrollees in trials.[8] This may be

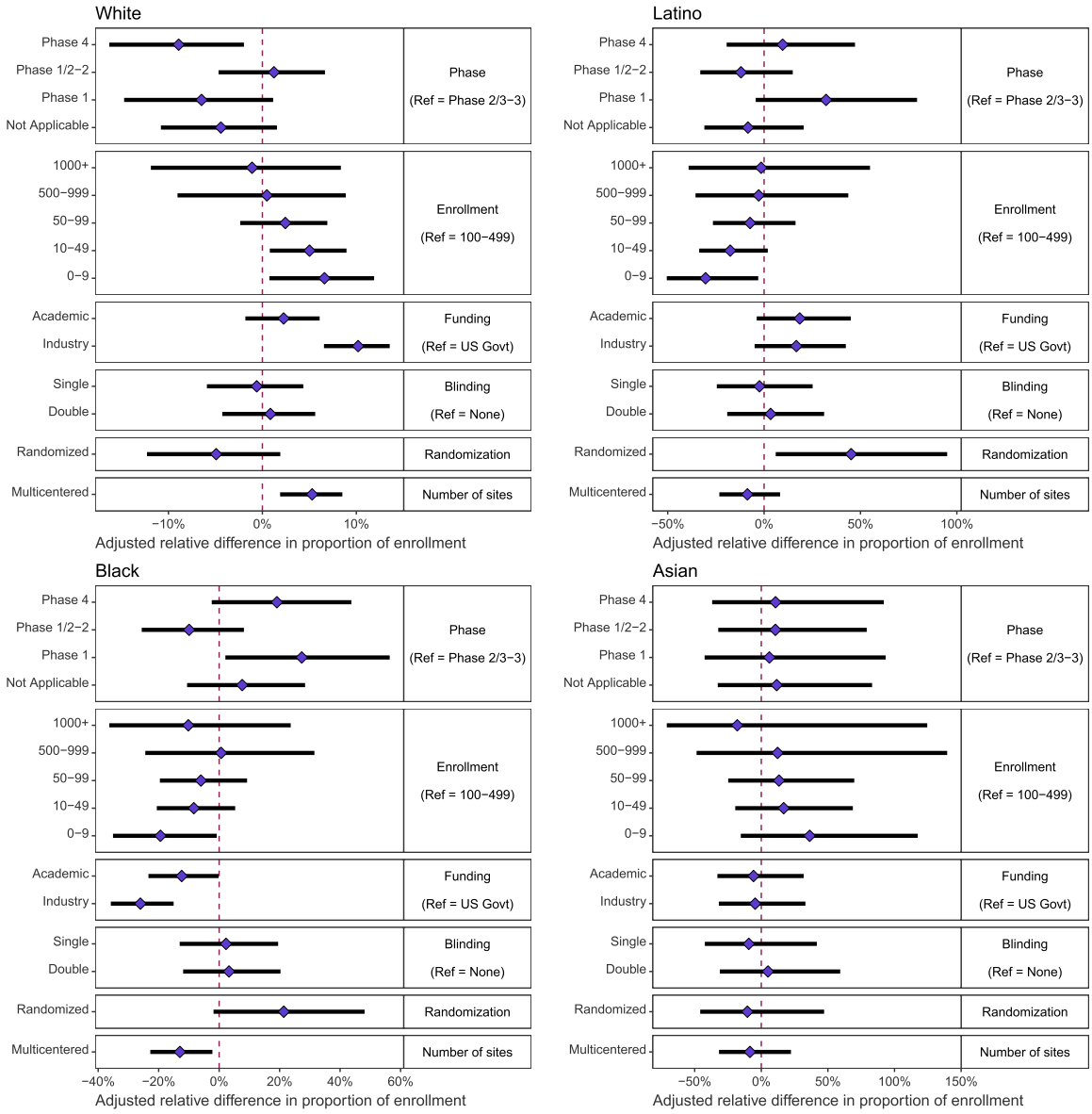

Figure 4. Association of trial features with enrollment of racial/ethnic groups.

Each unit shows the adjusted relative difference and the 95% confidence interval for that variable. On ClinicalTrials.gov "Not Applicable" is used to describe trials without Food and Drug Administration-defined phases, including trials of devices or behavioral interventions. Funding categories were determined with data on the sponsor and collaborators. Industry funding includes trials with an industry sponsor or collaborating agency US Government trials include remaining trials with a US Government sponsor or collaborating agency.

connected to healthcare distrust or the historical challenge of capturing Latino identity in administrative records.[26] We suspect the over-representation of Whites in our analysis understates the true overrepresentation of non-Latino Whites. Unfortunately, the standard demographic categories offered by ClinicalTrials.gov (and reported in many trials) lack precise race and ethnicity combinations and instead aggregate each separately, despite FDA recommendations to provide cross-tabulation.[24] Through the lens of median enrollment, Blacks were not underrepresented. This is similar to an FDA Snapshot analysis of trials leading to 231 new drug approvals from 2015 to 2019.[27] While Asians remained underrepresented (2% of US participants), Blacks and Latinos were not (16% and 15%, respectively). Trials leading to FDA drug approvals represent a particularly meaningful cohort in the context of concerns regarding discrepancies in drug efficacy or safety among

underrepresented populations. However, these trials comprise a narrow subset (total US participants were approximately 103,000, compared to the approximate 2,0878,000 in our study) and apply only to studies for new molecular entities or biologics (i.e. does not include trials for expanded indications, off-label use, or non pharmacologic interventions). In our cohort 21% of trials reported 0 Black enrollees, indicative of the heterogeneity that can be obscured in smaller samples.

Similar to reporting, the inconsistent results among previous studies assessing trial diversity is likely influenced by the varying mixture of trial features (e.g. funding) present within their samples. Detection of minority enrollment trends is also more challenging because only a fraction of trials report race/ethnicity enrollments (especially for all five groups) and the trends' true effect sizes are likely modest as demonstrated in our analysis. Our findings begin to resolve these inconsistencies, but without complete race/ethnicity reporting, all estimates remain susceptible to reporting bias.

The association of industry with greater White enrollment may result from both regulatory and financial incentives that indirectly reward homogeneity through the reduced risk of confounding from patient-related factors (e.g. more predictable therapeutic effects, decreased risk of adverse events).[7] As industry contributes the most to applied preclinical and clinical research in addition to production, marketing and distribution of new therapies,[28] this finding may have unique implications for disparities in treatment efficacy and access. Interestingly, industry and academic funding trended toward a positive effect among Latinos relative to government funding. We suspect this may reflect a unique barrier to recruitment of Latinos compared to other minorities among US Government trials. Relative to US Government-funded trials, the association of academic funding with race/ethnicity enrollment was similar to though typically weaker than that of industry. Academic trialists may experience a mixture of the regulatory and diversity influences within government agencies and the convenience and cost-incentives within industry.

The majority of studies we reviewed did not restrict their samples to exclusively US trials and also did not analyze differences between trials that enrolled within or outside the US. In addition to inconsistencies in racial/ethnic definitions across cultures, determining over- or under-representation of demographic groups is ambiguous without clear population references. Although racial prejudice and health disparities cross national borders, the relationship between research and systemic racism is formed within a cultural and socioeconomic context. In the US, systemic racial inequities and prejudices have produced severe health consequences.[15,29]

While demographically distinct, international trials frequently inform US drug approval and clinical practice. One study by Tahhan et al. found that the trend toward multi-centered, multi-regional trials increased minority inclusion.[12] However, this trend was driven primarily through increased enrollment of Asians within Asian countries, and the proportion of Blacks actually decreased in these studies. An analysis by Loree et al. that included non-US trials found that Asians represented 18% of trial participants and were "overrepresented", though they compared Asians representation only to their population within the US.[11] Khan et al. found that in European and multiregional trials Asians represented approximately 8% of participants while Blacks and Latinos represented <2%.[8] In contrast to these global studies, we find that among USA-based trials Asians are underrepresented and had a median enrollment of 1% compared to 10% and 6% for Blacks and Latinos, respectively.

Our study had several strengths. While other investigations limited their analyses to smaller samples of one specialty or subset of journals, we aimed to capture all US-based trials. We did not rely on published trials. Previous studies have demonstrated that journal data are less complete than ClinicalTrials.gov results and many trials are never published.[30] Our methods help mitigate the sampling bias observed in other studies and increase the generalizability of our findings. Our large cohort of trials permitted examination of temporal trends and multivariable analyses to account for potential sources of confounding and clarify potential sources of inconsistency among previous reports.

Our study should be interpreted in the context of some limitations. First, reporting of race/ethnicity was inconsistent across trials, particularly the handling of Latino populations, and thus estimates of Latino enrollment may be inaccurate. We attempted to account for this incomplete reporting through our sensitivity analysis. Second, because the majority of trials do not report race/ethnicity, our findings may not generalize to the rest of the database. Third, not all US trials are registered in ClinicalTrials.gov and thus our findings are biased towards those covered by FDAAA 801, though these trials are also the most clinically relevant. Fourth, while the National Library of Medicine does employ multiple quality checks, there remain incomplete or sometimes inaccurate data fields within ClinicalTrials.gov.[31] Our dataset is thus representative only of the data that trials voluntarily report. We attempted to account for this with multiple imputation and sensitivity analyses, but nonetheless the findings must be interpreted with this context. Fifth, we performed multiple comparisons and did not perform testing correction; some findings may have occurred by chance.

In conclusion, the reporting of participant race/ethnicity and the enrollment of diverse populations in USA-based trials are poor but improving. In an era of data-driven medicine, it is difficult to improve what we do not measure. Clear and consistent reporting represents an achievable goal which enables downstream innovation and accountability. A standardized system

with easier data entry and more detailed demographics (including cross-tabulations for race and ethnicity) may present initial steps. Straightforward tools to analyze race/ethnicity data within ClinicalTrials.gov could allow researchers to examine and learn from similar trials while anticipating likely challenges to diverse recruitment. Compulsory race reporting for funding or journal publication are straightforward and similar requirements have historically been effective (e.g. trial registration). Additional incentives and enforced regulations may be needed to ensure all sponsors are engaged and accountable for the recruitment of representative cohorts. All stakeholders must commit to consistent and transparent results reporting to enable innovative solutions for the recruitment of cohorts that are representative of the population as a whole.

Contributors

Dr. Brandon E Turner: Literature search, figures, study design, data collection, data verification, data analysis, data interpretation, writing.

Dr. Jecca R. Steinberg: Literature search, study design, data verification, data interpretation, writing.

Dr. Fatima Rodriguez: Data interpretation, writing.

Dr. Mark R Cullen: Study design, data interpretation, writing.

Dr. Brannon T Weeks: Study design, writing.

Data sharing statement

All data were downloaded from ClinicalTrials.gov/AACT or from US Census websites, which are free and open to the public.

Declaration of interests

MRC has been compensated for consulting services to Pfizer in the Global Epidemiology Unit and to the Bill and Melinda Gates Foundation as a Scientific Review Lead.

Acknowledgments

We acknowledge partial research funding from Stanford Medical Scholars and from the National Heart, Lung, and Blood Institute, NIH (1K01HL144607) and the American Heart Association/Robert Wood Johnson Medical Faculty Development Program.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.lana.2022.100252.

References

1 National Institute of Health (NIH). S.1 - National Institutes of Health Revitalization Act of 1993 Subtitle B-Clinical Research Equity Regarding Women and Minorities Part I-Women and Minorities as Subjects in Clinical Research Sec. 131. Requirement of Inclusion in Research. National Institute of Health (NIH); 1997. https://orwh.od.nih.gov/sites/orwh/files/docs/NIH-Revitalization-Act-1993.pdf.

2 Chen MS, Lara PN, Dang JHT, Paterniti DA, Kelly K. Twenty years post-NIH revitalization act: enhancing minority participation in clinical trials (EMPaCT): laying the groundwork for improving minority clinical trial accrual. *Cancer*. 2014;120:1091–1096.

3 Andrews K. Racism is the public health crisis. *Lancet*. 2021;397:1342–1343. (London, England).

4 Warren RC, Forrow L, Hodge DA, Truog RD. Trustworthiness before trust — Covid-19 vaccine trials and the black community. *N Engl J Med*. 2020;383:e121.

5 Flores LE, Frontera WR, Andrasik MP, et al. Assessment of the inclusion of racial/ethnic minority, female, and older individuals in vaccine clinical trials. *JAMA Netw Open*. 2021;4:e2037640.

6 Cahan EM, Hernandez-Boussard T, Thadaney-Israni S, Rubin DL. Putting the data before the algorithm in big data addressing personalized healthcare. *NPJ Digit Med*. 2019;2:78.

7 Nazha B, Mishra M, Pentz R, Owonikoko TK. Enrollment of racial minorities in clinical trials: old problem assumes new urgency in the age of immunotherapy. *Am Soc Clin Oncol Educ B*. 2019:3–10.

8 Khan MS, Shahid I, Siddiqi TJ, et al. Ten-year trends in enrollment of women and minorities in pivotal trials supporting recent US Food and Drug Administration approval of novel cardiometabolic drugs. *J Am Heart Assoc*. 2020;9:e015594.

9 Zhang T, Tsang W, Wijeysundera HC, Ko DT. Reporting and representation of ethnic minorities in cardiovascular trials: a systematic review. *Am Heart J*. 2013;166:52–57.

10 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;119:2956–2963.

11 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. https://doi.org/10.1001/jamaoncol.2019.1870.

12 Samman Tahhan A, 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. supplemental content clinical review & education JAMA cardiology | Review *JAMA Cardiol*. 2018;3:1011–1019.

13 Tasneem A, Aberle L, Ananth H, et al. The database for aggregate analysis of Clinicaltrials.gov (AACT) and subsequent regrouping by clinical specialty. *PLoS One*. 2012;7:e33677.

14 Washington HA. *Medical Apartheid: The dark History of Medical Experimentation on Black Americans from Colonial Times to Present*. 1st ed. New York, NY: Doubleday Books; 2007.

15 Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. *Lancet*. 2017;389:1453–1463.

16 United States Census Bureau. U.S. Census (2010). https://data.census.gov/cedsci/. Accessed 28 January 2020.

17 Zarin DA, Tse T, Williams RJ, Carr S. Trial reporting in ClinicalTrials.gov - the final rule. *N Engl J Med*. 2016;375:1998–2004.

18 Clinical Trials Transformation Initiative. 2018. https://www.ctti-clinicalrtials.org. Accessed 2 January 2019.

19 Steinberg JR, Turner BE, Weeks BT, et al. Analysis of female enrollment and participant sex by burden of disease in US clinical trials between 2000 and 2020. *JAMA Netw open*. 2021;4:e2113749.

20 McLeod A. Kendall Package: Kendall rank correlation and Mann-Kendall trend test. 2011. https://cran.r-project.org/web/packages/Kendall/Kendall.pdf.

21 Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *J Appl Econom*. 1996;11:619–632.

22 Buuren SV, Groothuis-Oudshoorn K. Mice : multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:128–129.

23 DeVito NJ, Bacon S, Goldacre B. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. *Lancet*. 2020;395:361–369. (London, England).

24 HHS, FDA. Collection of race and ethnicity data in clinical trials: guidance for industry and food and drug administration staff clinical medical preface public comment. 2016 https://www.fda.gov/media/75453/download.

25 ICH. ICH Harmonised tripartite guideline: ethnic factors in the acceptability of foreign clinical data. 1998 https://database.ich.org/sites/default/files/E5_R1__Guideline.pdf.
26 Guevara C, Cook C, Herback N, Pietrobon R, Jacobs DO, Vail TP. Gender, racial, and ethnic disclosure in nih k-award funded diabetes and obesity clinical trials. *Account Res*. 2006;13:311–324.
27 FDA. 2015-2019 drug trials snapshots summary report. 2020 https://www.fda.gov/media/143592/download.
28 Moses H, Matheson DHM, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research US and international comparisons. *JAMA J Am Med Assoc*. 2015;313:174–189.
29 Thomas JK. Medical Apartheid: The Dark History of Medical Experimentation on Black Americans from Colonial Times to the Present, by Washington H.A., Illustrate. Doubleday Books, 2009. https://doi.org/10.1526/003601109789037187.
30 Hartung DM, Zarin DA, Guise JM, McDonagh M, Paynter R, Helfand M. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. *Ann Intern Med*. 2014;160:477.
31 Tse T, Fain KM, Zarin DA. How to avoid common problems when using ClinicalTrials.gov in research: 10 issues to consider. *BMJ*. 2018;361:k1452.