Modern Reflections: An Assessment of Diversity in U.S. Cancer Clinical Trials

Katherine Huerne

Biomedical Ethics Unit, McGill University, Montreal

Initial submission date: June 1, 2022

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Masters of Science.

Table of Contents

Abstract (English): Demographic diversity in clinical trials can be defined as explicitly including and thoughtfully characterizing the differences in response between subpopulations of a given sample of research participants. This thesis will provide an overview of the modern state of diversity by exploring the history of U.S. clinical trial diversity, offering ethical and epistemic reasons for diversity, and examining how diversity relates to greater social ideals of equity, representation, and inclusion within Western society. A cross-sectional study using U.S. Phase III trial data from the last 5 years was conducted with the aim of analyzing the state of demographic diversity in modern cancer clinical trials. It was found that women were slightly underrepresented in all cancer trials, racial/ethnic minorities were significantly underrecruited, especially for Black and Native American demographics, and the average age in clinical trial recruitment was 5.9 years below the real-world mean for all cancer types. The detailed demographic analysis presented in this thesis, alongside practical recommendations for increasing diversity, was presented with the aim to contribute to existing critical reflections of modern-day diversity in cancer clinical trials.

Abstract (French): La diversité démographique dans les essais cliniques peut être définie comme l'inclusion explicite et la caractérisation réfléchie des différences de résultats entre les sous-populations d'un échantillon de participants à la recherche donné. Cette thèse donnera un aperçu de l'état actuel de la diversité en explorant l'histoire de la diversité dans les essais cliniques aux États-Unis, en proposant des raisons éthiques et épistémiques pour la diversité et en examinant comment la diversité est liée à des idéaux sociaux plus importants d'équité, de représentation et d'inclusion dans la société occidentale. Une étude transversale utilisant les données des essais américains de phase III des 5 dernières années a été menée dans le but d'analyser l'état de la diversité démographique dans les essais cliniques modernes sur le cancer. Il a été constaté que les femmes étaient légèrement sous-représentées dans tous les essais sur le cancer, que les minorités raciales/ethniques étaient nettement sous-recrutées, en particulier les Noirs et les premières nations, et que l'âge moyen de recrutement dans les essais cliniques était de 5,9 ans inférieur à la moyenne du monde réel pour tous les types de cancer. L'analyse démographique détaillée présentée dans cette thèse, ainsi que les recommandations pratiques pour accroître la diversité, ont été présentées dans le but de contribuer aux réflexions critiques existantes sur la diversité moderne dans les essais cliniques sur le cancer.

Acknowledgements

The author would like to thank Dr. Jonathan Kimmelman and Dr. Phoebe Friesen for their supervisory efforts and guidance in the development of this thesis.

The author would like to thank Dr. Yann Joly, Dr. Carolyn Ells, and Dr. Jennifer Fishman for their support as members of the thesis advisory committee.

The author would like to thank Dr. Nicholas King for their support as examiners of the initial thesis.

The author would like to thank Dr. Nicole Palmour for her efforts in revising initial drafts of the thesis.

The author would like to thank Samuel Gregoire, B.Sc. McGill University, for his guidance in statistical analysis and development of the algorithm used for data extraction and visualization.

Contribution of Authors

Katherine Huerne was responsible for the drafting, writing, and subsequent revision of the thesis. There are no co-authors for this thesis.

List of Figures and Tables

List of Abbreviations

U.S. - The United States of America

FDA - U.S. Food & Drug Administration

ADRs - adverse drug reactions

NR3C1 - glucocorticoid receptor gene, nuclear receptor subfamily 3, group C, member 1

HPA axis - hypothalamic–pituitary–adrenal axis

FKBP5 - K506 binding protein 5

ggplot2 - an R package used for statistical computing & data representation in data visualization

NIH - U.S. National Institutes of Health

ANOVA - Analysis of Variance

HIV - Human Immunodeficiency Virus

ECOG - The Eastern Cooperative Oncology Group

OMB - U.S. Office of Management and Budget

LGBTQIA - lesbian, gay, bisexual, transgender, queer or questioning, intersex, and asexual

CIHR - Canadian Institute of Health Research

Rationale and Objectives

Rationale:

Cancer clinical trials comprise the largest type of clinical trials occurring today and act as a good indicator of clinical trial performance in general. However, U.S. clinical trials have historically lacked diversity for particular races, sexes, and age groups of participants. For example, a systematic review of Phase III cancer clinical trials conducted between 2001-2010 revealed that 82.9% of participants were White, while all other races were underrecruited. Other demographic studies suggest that contemporary cancer trials continue to overrepresent White, younger and male participants on average. Although various public health initiatives and federal regulations have sought to increase diversity in clinical trials over the past few decades, it is unclear if fair representation has successfully been achieved in modern cancer clinical trials. This project will offer an assessment of demographic diversity in U.S. cancer clinical trials from 2017-2021 to determine whether underrepresentation of certain populations still exists today.

Objectives:

- ➢ Present a qualitative review of modern clinical trial diversity for race, sex, and age demographics, including an overview of how diversity has evolved over the history of U.S. medicine.
- ➢ Provide an argument for ethical and epistemic reasons for diversity, including an examination of how social and biological factors can influence these reasons for diversity.
- ➢ Present an overview of race, sex, and age diversity in U.S. cancer clinical trials from 2017-2021 via quantitative data analysis.
- ➢ Perform a quantitative analysis of additional variables in clinical trials that can affect race, sex, and age diversity, such as trial size, sponsorship status and trial location. A qualitative analysis of clinical trial eligibility criteria will also be conducted.

Introduction

Demographic diversity in clinical trials can be defined as explicitly including and thoughtfully characterizing the differences in response between subpopulations of a given sample of research participants. In essence, diversity ties clinical research to social differences, as the usefulness of medical knowledge created is determined by the participants recruited. This section will provide an overview of the modern state of diversity by exploring the history of United States (U.S.) clinical trial diversity, offering ethical and epistemic reasons for diversity, and examining how diversity relates to greater social ideals of equity, representation, and inclusion within Western society. This section focuses on the U.S. history of clinical trial diversity to contextualize the diversity analysis of modern U.S. cancer clinical trials in the next section of the thesis.

The History of U.S. Clinical Trial Diversity

Tracing the social evolution of clinical trial history is critical to understanding diversity today. Until the 21st century, researchers were not always careful in collecting or presenting accurate demographic information of research participants, reflecting the notion that diversity may not have been an explicit concern in medical research (Epstein, 2008; Simon et al., 2014). At the same time, socially disadvantaged populations were differentially treated throughout clinical trial history through overrepresentation in risky trials (Fisher & Kalbaugh, 2011). In addition, these populations were subjected to experimentation without consent to develop new technologies, techniques and medical knowledge. Such was the case of Marion Sims who performed experimental operations on Black enslaved women from 1845-1849 (Wall, 2006).

Biological knowledge in Western medicine was historically framed as comparative differences between social groups. Medical claims of differences between social demographics in the 19th century within Europe and United States were used to justify racial hierarchies (that ‘Whites were more fit than Blacks’) and slavery in general (Epstein, 2008). At the same time, women were demarcated as biologically inferior to men, with femaleness and ‘women’s maladies’ viewed as inherently unhealthy and controlled by a woman’s reproductive organs (Cayleff, 1988; Epstein, 2008). Notions of biologically inferior populations persisted into the

20th century, where White physicians often believed non-White races (in particular Blacks and Chinese immigrants) posed a special risk of infection to the 'mainstream' White U.S. population (Epstein, 2008; Fofana, 2013). Although medical knowledge was directed towards the everyday adult European man, institutionalized populations such as soldiers, prisoners, mentally ill people, neurodivergent children, and poor people in general were primarily used as research participants in medical studies until the mid-20th century (Epstein, 2008; Strauss et al., 2021; Welch et al., 2015). In essence, the demography of the research subject, and demography who would benefit from the resulting medical knowledge were unequal and shaped by social prejudices and discriminatory policies.

The 1960's witnessed a series of high-profile exposés, such as the Tuskegee Syphilis Study, showcasing clear abuse of socially disadvantaged participants in U.S. medical experiments (Epstein, 2008; Reverby, 2000). By the 1970's, prominent publication of these cases drew a national wave of legislative reform, designating research participants to have formal, legal protection in experimental studies. This changed the paradigm of research participation to be that of a burden which must be distributed as equally as possible in society (Epstein, 2008). The National Research Act of 1974 mandated trials to be approved by institutional review boards (IRBs) to ensure participants would not be placed at undue risk, while formalizing a process of obtaining informed consent from participants (Epstein, 2008; Kraybill, 2004). The Belmont Report published by the U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1979, outlined the main ethical principles to guide medical research with human subjects: justice, respect for persons, and beneficence (Artal & Rubenfeld, 2017; Epstein, 2008). This report officially defined 'vulnerable populations' as children, prisoners, the poor, and the mentally infirm, to be in need of special protection from harm (Epstein, 2008; Office of the Secretary, 1979). In addition, the 1950's onwards saw widespread use of the thalidomide drug in pregnant women which subsequently caused birth defects in the U.S. and many other parts of the world (J. H. Kim & Scialli, 2011). This resulted in women of childbearing potential to be subsequently viewed as another vulnerable population in need of protection. Accordingly, the U.S. Food & Drug Administration (FDA) instituted a policy in 1977 to exclude women of childbearing potential from drug trials out of concern that an

experimental drug might bring the fetus harm if a woman became pregnant during a clinical trial (Uhl et al., 2007).

Although these protective measures were initially made in good faith, the added systemic barriers and fear of harm resulted in these demographics becoming unavailable participants. This led to the underrepresentation of these vulnerable populations (women, children, the elderly, people with mental illnesses, racial and ethnic minorities) in modern clinical trials altogether (Epstein, 2008; Noah, 2003; Shenoy & Harugeri, 2015; Uhl et al., 2007). For example, although the FDA rescinded the 1977 policy of banning women of childbearing potential from clinical trial participation in 1993, 'women of childbearing status' continue to be listed as a category of exclusion in most clinical trial eligibility criteria (Office on Women's Health, 2019). In combination with the historically poor quality of recording diverse participant demography, diversity in clinical trials has now become a matter of ensuring fair representation and establishing justice of inclusion for these particular populations (Duma et al., 2018). Thus, it is important to revisit the ethical and epistemic reasons for establishing diversity in modern clinical research.

Reasons for Exclusion Within Key Demographic Categories

The three main demographic categories tracked in modern clinical research are race/ethnicity, sex, and age, which will be the focus of the thesis. Here, I identify the reasons for differential treatment of socially disadvantaged populations within each demographic category. It is critical to demarcate these issues, as each demographic has a different context to consider.

Differential Treatment Between Various Races and Ethnicities

Differences in race and ethnicity diversity is contextualized by how various races and ethnicities were treated throughout North American society. Historically, physicians have used race as a factor in the clinical decision-making process. Whether intentionally done, the consideration of race has reduced ideas of race to a biological or genetic concept - a notion captured by the phrase 'race essentialism'. When taken to the extreme, the profiling of individuals based on race has historically been used to reinforce simplistic assumptions of

biological and psychological differences of the human body, as if race denoted essential differences between groups of people (Braun, 2017; Epstein, 2008). The visible, physiological, or behavioral differences between races and ethnicities were believed to be due to innate biological or genetic differences. These ideals were further used to justify differing treatment of each race/ethnicity in drug use or medical interventions. Comparisons of difference became heralding Whiteness as the standard of health against other racial and ethnic minorities. For example, the biological superiority of White populations compared to Blacks was taught in U.S. medical schools until the mid-20th century (Byrd & Clayton, 2001). Black people were likened to rats and mice - as carriers of disease to White folks, while the 'alien' Chinese were blamed for bringing diseases such as syphilis and the bubonic plague to White populations (Epstein, 2008; Jedwab et al., 2021; Wailoo, 2006). While these beliefs were rooted in larger issues of racial discrimination in Western society, a major outcome of these injustices was that it segregated the clinical relevance of medical knowledge to be predominantly about White populations at the expense of others.

Furthermore the differential distribution of resources at clinical trial centers results in various races being included and served differently. Most clinical research is conducted in well-funded and sophisticated hospitals in urban centers (Baquet et al., 2006; Borno et al., 2018; Mudaranthakam et al., 2022). In the U.S., insufficient private health insurance coverage (which predominantly affects undocumented and low-income racial minorities) limits these populations from accessing the hospitals (Baquet et al., 2006; Bartlett et al., 2005; Unger et al., 2013). Meanwhile, the community hospitals that serve racial minorities often do not conduct trials, which will result in the overall biases in recruitment (Byrd & Clayton, 2001; Murthy et al., 2004; Williams, 1997). A study conducted by Joseph Galen revealed that the organizational climate of clinical trial centers can impose institutional barriers to recruitment of racial and ethnic minorities (Joseph & Dohan, 2009). These barriers include accessibility issues with the clinic/hospital (e.g. clinic hours, patient assignment method); the lack of interdisciplinary and continuous care; competing provider priorities of clinical care versus research; staff, funds and institutional limitations to facilitate clinical trials; and linguistically and literacy-appropriate resources for participants (Joseph & Dohan, 2009).

Along similar sentiments, racial minorities have received different standards of treatment at different stages of clinical research, or were recruited to risky trials with unfavorable conditions. For example, Jewish populations were used as subjects for Nazi Germany medical research, and often subjected to lethal and painful conditions. Chinese research subjects were used as replacements to prosecute Japanese war missions (Nie, 2002). Black enslaved individuals were sold and used as "specimens" of research subjects to improve the institution of slavery (Savitt, 1982). Later on, Black populations were often overrepresented in early-phase riskier trials, but underrepresented in late-phase safer trials (Fofana, 2013; Osborne & Feit, 1992). Meanwhile, both Black and Indigenous communities have historically been used for medical testing while overlooked in healthcare services (Srikanth, 2020). The treatment of these populations as available test subjects has resulted in their general mistrust of the medical community to conduct research in the interests of these populations (Kennedy et al., 2007). Instances of poorly-executed differential treatment toward Black populations was furthered by attempts to mitigate racial discrimination, such as the BiDil controversy, which was a campaign approved by the FDA that targetted congestive heart failure drug only for Black populations (Brody & Hunt, 2006). The problem with BiDil was that it sought to appeal to 'race biology' to create the impression that the best way to address health inequities was through drug development, rather than examining other social determinants of health associated with race and racism (Brody & Hunt, 2006).

Likewise, Asian and Hispanic populations have also been overlooked in late-stage clinical research, with differing reasons for their exclusion. Today, the medical community holds widespread disagreement about whether Asians are actually underrepresented in clinical research, possibly due to confirmation bias of the model minority myth or the perceived cultural proximity of Asian populations to White privilege (Alexander et al., 2000; Nguyen et al., 2021; Shah & Kandula, 2020). Meanwhile, the consensus is clear that Hispanic populations are generally underrepresented in clinical research, but an issue remains in delineating a clear definition of the 'Hispanic' demographic category apart from the overrepresented White race (Parra et al., 2014). Nonetheless, these populations of 'visible minorities' were still removed

from considerations of Western clinical knowledge which has historically championed 'White biology' at the forefront of medicine (Hamel et al., 2016).

Today, there remains considerable ethical dilemmas posed by racial classification in health research (Duster, 2015; Ellison et al., 2008; Kaufman, 1999). Namely, there have been diverse types of racial classification proposed over time, with a blurred line whether these categories are socially or biologically defined, and if these categories are clinically warranted. While the forthcoming epistemic section will explore the implications of racial classification in health research as a biological category, diversity in race/ethnicity nonetheless remains ethically warranted based on historical means. Ensuring a fair distribution of races and ethnicities means examining the prejudices that have fueled the differential treatment of various races and ethnicities from clinical research.

Differential Treatment Between Various Sexes and Genders

Differences in sex and gender diversity is contexualized by how women and sexual minorities have been ignored or policed in clinical research, and how gender roles have been used to justify the exclusion of women from participation.[1] Western medicine has historically portrayed the male body as the standard of human health, with women understood to be equivalent to men in essential ways except for differences in reproductive organs (Cayleff, 1988; Epstein, 2008). In addition, historical medical knowledge denoted male and female to be the only two sexes of relevance, with intersex and sexual minorities seen as diseased or disfunctional forms of the binary sex (Matsuno & Budge, 2017). The evolution of how sex differences are understood in medicine forms the basis for justifying modern clinical knowledge between the sexes. The main argument for having sex diversity in clinical trials rests on the belief that men and women differ in anatomical, physiological, genetic, chromosomal, and hormonal aspects (Geller et al., 2011). These differences explain the disparities observed in therapeutic outcomes, like drug response differences between men and women that take the same dosage, which reinforces the reason to study different sexes in their own right. Emphasis on sex differences in

[1] Sex is defined as the set of biological attributes such as physical, genetic, and physiological features, whereas gender is defined as the socially constructed roles, behaviours, expressions and identities of girls, women, boys, men, and gender diverse people (Canadian Institutes of Health Research, 2015).

medicine has become part of a larger trend of assuming the overriding difference of biology and genetics to justify the different behaviors of males and females (Epstein, 2008).

However, genetic studies reveal that men and women are more similar than distinct: within the approximate 31,000 genes of the human genome, men and women differ only in the two sex chromosomes (X/Y) and by only a few dozen genes (Epstein, 2008). It becomes hard to justify that all sex disparities are based solely on biological difference - which is where gender as a social determinant of health could be useful to explain possible health differences seen between the sexes. Yet, due to the popularity of biological sex profiling, the effect of gender on therapeutic outcome has been overlooked to the point that gender is not explicitly tracked in clinical research today as a separate demographic category outside of sex (Matsuno & Budge, 2017). Although binary sex and gender may largely overlap, gender-diverse identities are becoming commonly expressed, thus expanding the difference between sex and gender populuations, albeit a minor difference for now (Matsuno & Budge, 2017). In addition, sex in medicine typically only encompasses biological, genetic, or chromosomal differences, whereas gender encompases behaviour and social differences.

Alongside the historical exclusion of women of child-bearing potential discussed prior, gender biases have influenced the perceived value that men and women bring to medical knowledge. The exclusion of women's autonomy in medicine meant that medical technologies and knowledge developed throughout North American history were seen through a patriarchal and restrictive lens, where the utility of women's bodies was primarily for childbearing, caregiving, and domestic duties (Feuerstein et al., 2018). ‘Healthcare about women’ was historically focused on preserving or restoring a woman’s childbearing ability rather than in the interests of women’s health in itself. Although the awareness of women’s inclusion in clinical research has increased in modern times, the advocacy of women’s interests outside of childbearing has not necessarily received proportional attention (Feuerstein et al., 2018). Thus, diversity in sex/gender means taking into account a woman’s own autonomy beyond her childbearing potential, while also including sex and gender minorites (such as intersex and transgender individuals) in the trial itself.

Differential Treatment Between Various Age Groups

Differences in age diversity is contextualized by the differential treatment of age groups in clinical trials. Two age groups have been historically underrepresented in clinical trials for different reasons: children and the elderly. Children were largely excluded from participating in medical research due to fear of causing them undue harm and their inability to provide legal consent (Office for Human Research Protections, 2016). This fear began from stories of Nazi war crimes presented at the Nuremberg Trials, highlighting the medical experimentation on children, as well as subjecting institutionalized or orphaned children to invasive surgery or intentional infections (Nellhaus & Davies, 2017). Such an example was Dr. Krugman who deliberately injected institutionalized children with hepatitis at Willowbrook State School to study hepatitis in the mid-1900's (Diamond, 1973; Robinson & Unruh, 2011). These instances positioned children as a special demographic with high liabilities of potential safety and autonomy compromises (Laventhal et al., 2012; Welch et al., 2015).

Furthermore, children were not seen as paradigmatic examples of biological standards for human beings. The diverse and fast-changing biological variation at the age ranges of child growth meant they served as less suitable research subjects in many types of clinical research against the comparably stable standard of the adult human body (Caldwell et al., 2004; Cooter et al., 1992). The international ethical guidelines introduced throughout the 20th century (The Declaration of Geneva, The Nuremberg Code, The National Research Act, The Declaration of Helsinki, The Belmont Report) formally introduced children as a vulnerable demographic needing special protections. Weak incentives to conduct research effectively dissuaded clinicians from performing medical research on children altogether (Office for Human Research Protections 2016). However, the desire to safeguard children from dangerous medical interventions exacerbated the problem of inclusion: researchers could not obtain knowledge on children's health and the safety of medical interventions if medical research was not conducted on children in the first place. As a result, pediatric trials are highly scrutinized by Institutional Review Boards (IRBs), making them scarcer in number and more difficult to implement than adult trials (Laventhal et al., 2012).

For elderly adults, their underrepresentation in clinical trials corresponds to the tendency of researchers to minimize adverse reactions while optimizing reports of safety and efficacy/effectiveness. The diminishing biological fitness of elderly populations is typically thought to compromise these goals (Aapro et al., 2005). However, people 65 and over are more affected by chronic health issues like cardiovascular disease or cancer, and rely on prescription drug use compared to younger age groups - in other words, they are often the intended subjects of approved clinical trial drugs and therapies (Granger, 2020; Shenoy & Harugeri, 2015). Paradoxically, trials (Phase III cancer trials in particular) often have a 65 and over cut-off age for participants without explanation (Granger, 2020; Pruitt et al., 2017; Watts, 2012). Clinical trials also tend to exclude participants with comorbid conditions, polypharmacy, communication issues, and physical immobility, implicitly excluding older populations as many of these conditions apply to them (Granger, 2020; Shenoy & Harugeri, 2015). Other operational challenges can disproportionately affect elderly participants, like difficulty in communication, gaining informed consent, economic constraints, and transportation issues, which will create issues in retention (Granger, 2020). For elderly populations and children, ensuring equity means their enrollment in clinical trials should be essential to confirm dosage, safety, adverse events and effectiveness for their population. For children in particular, it means addressing systemic barriers that decrease the rate and quantity of pediatric trials.

As a culmination of all the aforementioned mechanisms of differential treatment, medical research has historically been based on that of the 18-65, White/European man. It is ironic then, that in the interests of protecting these ‘vulnerable populations’ from over-recruitment, medical knowledge becomes represented by the same white, male and young adult demographic. Thus, establishing diversity becomes a way to challenge the systemic barriers that have prevented other populations from inclusion in clinical research, with diversity becoming an effective way to heterogenize the research participant pool. It is important for research participants from different demographic backgrounds to be fairly represented, as they are all deserving of medical knowledge created *about* them. This prompts a consideration of the ethical and epistemic reasons for diversity in clinical trials, which I discuss next.

Ethical Considerations of Diversity

Equity is one of the commonly cited reasons for clinical trial diversity. Equity is defined as fairness and justice in the way people are treated. Here I explore equity in the context of how vulnerable or socially disadvantaged populations are treated in clinical research. In the last section, it was discussed how these populations came to be, which was based on their historical mistreatment in medicine. Clinical trial diversity is thus for the sake of restorative justice within the historical context of social prejudice or discrimination against these populations, which has either led to their exploitation, underrepresentation or exclusion. This section delineates the two main ethical consequences of this historical context, which encompasses the greater goal of diversity considerations in clinical research: equal access to novel medical technologies, and justice in future representation of these populations.

Access

Access to clinical trials is a concern when disadvantaged populations (racial minorities, women, children, etc.) are underrepresented due to the systemic barriers they face related to enrolling or staying in a trial. Due to the larger systemic inequities in North America, these populations are often correlated to being a visible minority or another 'vulnerable population' category. For example, racial bias from clinicians, or distrust of the racially-prejudiced medical system, can be one of the reasons that visible minorities are ignored or dissuaded from recruitment efforts in trials (Clark et al., 2019). Trials with frequent follow-ups might prevent older adults and women with caregiving duties from consistently participating in the trial (Mahmud et al., 2018). The location of the trial itself, the language of conduct, or transportation demands can deter those living in remote areas, non-English speakers, or people with mobility issues from participating (Baquet et al., 2006). The result of these constraints mean that it is easier for remaining populations - aka White, male, younger, wealthier, or educated participants in urban areas - to enroll in these trials instead (Baquet et al., 2006). Meanwhile, it can be easier to mitigate these trial-related systemic barriers than the larger societal barriers to medical access. For example, clinicians can design the trial to be in multiple languages, or in rural locations, to increase accessibility of the trial before it is implemented. Because these barriers

disproportionately affect the various ‘vulnerable’ demographics, intentionally designing and implementing accessible trials will coincidentally work to ensure access to these socially-disadvantaged populations.

Access is important to consider in clinical trial design as it can reveal systemic barriers behind the underrepresentation of various demographic populations which may have previously been overlooked by trial investigators (Borno et al., 2018). Addressing the accessibility concerns forces investigators to critically revisit how trial design, implementation, and knowledge may have excluded certain populations in the past, while giving underrepresented communities a chance to participate in the knowledge-making process. In the greater context, access issues also persist within the healthcare industry, which prevent particular populations from receiving or knowing about medical technologies or research in the real world. The culmination of these systemic barriers means that low socioeconomic, non-English speaking, and remotely-located populations face issues of access to medical technologies in the real world (Baquet et al., 2006). On the other hand, access to late-phase clinical trials may be one of the only avenues for these populations to use novel medical technology (Baquet et al., 2006). Although clinical research should be distinguished from healthcare in theory, in reality, participating in clinical research may be the only reasonable avenue for socially disadvantaged populations to have fair access to medical technology or health check ups with a physician. Ensuring diversity in clinical trials thus becomes a way to ensure fair access to medical technology on a large scale.

Justice

The Belmont Report defines ‘justice’ as distributive justice, meaning fairness in distribution between social categories, with injustice arising from social, racial, sexual and cultural biases institutionalized in society (Institute of Medicine (US), 1994; Office of the Secretary, 1979). In the clinical trial context, the ‘distribution’ is of the benefits and burdens that a trial brings across the various social identities. Although it can be argued that there are few individualistic benefits to participation in clinical research, one benefit is access to the trial itself, where participants benefit from the early application of medical technologies directly before its release to the general public (National Institute on Aging, 2017). Participants also have the

opportunity to engage in more collaborative communication with clinicians over their healthcare needs (National Institute on Aging, 2017). Clinical trials can also offer incidental or collateral benefits, as sometimes clinicians may detect or treat underlying conditions that may have otherwise gone unnoticed. Individuals with chronic health conditions, rare diseases, or conditions with no standard cure can benefit from individualized therapy, psychological reassurance and individual regular monitoring of their health conditions by participating in clinical trials (Mahipal & Nguyen, 2014). Especially for participants in the global south or remote communities who do not have access to regular healthcare systems, participation in clinical trials might be the few times that individuals gain access to a physician.

However, for most populations, trials often hold additional burdens that outweigh the benefits of the trial when compared to standard, non-experimental care. Especially for early phase trials, clinical trials can involve the use of experimental drugs, therapies or surgeries that can negatively affect the participant in undetectable or unknown ways (Office of the Secretary, 1979). Research risks also range from minor inconvenience, to discomfort, to actual harm or pain, either acute or chronic (Office of the Secretary, 1979). The participant can also face burdens from trial participation, such as stopping the use of existing medication for their health conditions, refraining from becoming pregnant, or making lifestyle compromises just to maintain their participation in a trial. Sometimes, trials offer no direct benefit for participants - instead the benefit of the trial comes from creating generalizable knowledge for the medical community (Mahipal & Nguyen, 2014). As the purpose of a clinical trial is to assess the risks and benefits, the participant ultimately assumes all of the risks and harms that may arise, but with variable benefit for themselves.

Although justice implies a fair allocation of risks and benefits of research between social categories, we have seen that certain demographic populations (poor, homeless, Black or other vulnerable populations) have been over-recruited in early-phase trials, where the risks often outweigh immediate benefits (Strauss et al., 2021). On a global scale, the allocation of risks and benefits have thus not been fairly distributed among these populations, even if the risks and benefits might differ for each individual and trial (Office of the Secretary, 1979). In addition, historical attempts to be gender or race-blind in clinical trial recruitment has resulted in

detrimental health consequences for women and racial minorities, such as creating adverse drug reactions (ADRs) in women with male-standardized dosages (Franconi et al., 2007). Results from studies reveal that ADRs are 2x more likely to be seen in women, with the subsequent 'overmedication' of women posing safety concerns in adverse reactions as well as increased cancer risk (Zucker & Prendergast, 2020). The role of sex as a biological factor in ADRs are also poorly understood, due to the underrepresentation of women's bodies in clinical trials (Zucker & Prendergast, 2020).

Beyond a simple categorical inclusion of diverse populations, justice implores clinicians to ensure that the *distribution* of medical knowledge is fair. In other words, the knowledge generated should be accessible to, propagated by, and addresses the concerns of each demographic group, especially if the distribution of medical knowledge has historically underrepresented their bodies. This goal is needed to fill knowledge gaps like the study of sex biology on women's ADRs, which requires a shift in priorities on women's health research beyond their reproductive utility as childbearers. Thus, the moral imperative of justice becomes two-fold: (1) to ensure that undue burdens are not overly placed on these historically vulnerable populations, while (2) including populations that have been historically overlooked or underrecruited in beneficial clinical research.

To summarize, the ethical reasons for diversity in clinical trial participation stem from a fair distribution of harms and benefits between socially-disadvantaged and advantaged populations. Access in this sense represents one aspect of distributing justice: to ensure fair access to the benefits from research results in a fair distribution of risks and harms. Beyond considerations of equity, access and justice, there are other scientifically-inclined reasons for including these populations, such as the value that diversity brings to generalizable medical knowledge. The next section explores the epistemic reasons for diversity and the impact of diversification on the scientific process.

Epistemic Considerations of Diversity

The gold standard in clinical research is often said to be the double-blind, placebo-controlled trial - a form of empiricism that embodies the ideal method of generating

knowledge from specific findings about small samples to larger populations (Kaptchuk, 2001). Likewise, an implicit standard of the human body also exists in clinical research - the pinnacle of human physiology has historically been that of the heterosexual, adult European white man (Epstein, 2008). These standards imply that diversity affects epistemic goals of clinical trials, where it must not inhibit the scientific rigor of the sample population. This section will briefly trace how this ideal standard of human physiology came to be and how demographic diversity affects the generation of knowledge and ideas of biological essentialism within the greater scientific community.

The Ideal Human Body

The ideal clinical trial participant (or who was considered the standard 'patient' in Western medicine) has been embodied by many concepts throughout history - the 'Vitruvian Man', the 'normative standard', 'standardized patients', or lately, the 'Average American Man', coined by Dr. Bernadine Healy, former director of the National Institutes of Health (Epstein, 2008; Keating & Cambrosio, 2002). Precisely, it is the 60 kilogram white male, 35 years of age, often heterosexual and in relatively 'good health' (as in no chronic disabilities or co-morbidities)(Epstein, 2008; Viby-Mogensen et al., 1996). The history with how this ideal came to be aligns with the history of clinical trial development. As standards of clinical trial practice became increasingly ethically scrutinized in the 20th century, clinical trials subsequently became regulated and standardized under the promise of ensuring fairness and rigor in medicine (Keating & Cambrosio, 2002). Along the way, the push for universalization began from movements to reform clinical trials as a means to develop uniform, evidence-based guidelines for patient care, alongside efforts by the FDA and pharmaceutical industry to standardize the drug approval process across national borders (Arpinelli & Bamfi, 2006; Epstein, 2008).

The 'one size fits all' approach to clinical trials thus entailed extending its ideology to the object of the trial itself - the patient - who was conceived of in relatively standard and universal terms (Epstein, 2008; NIH Office of Research on Women Health, 2003). It echoed the basic tenets of empiricism and the scientific process: that the production of trustworthy knowledge presumes the standardized control of experimental objects, whose essential characteristics are

uniform and can be replicated from one laboratory to another (Epstein, 2008; NIH Office of Research on Women Health, 2003). In doing so, the scientific process justifies the knowledge generated from the clinical trial itself: rigor from the experimental process translates to rigor of the knowledge generated. The inclusion/exclusion criteria, formal rules designating eligibility of trial participation, also implicitly serve to formalize this demographic standard. Strict inclusion/exclusion criteria became designed to: (1) exclude 'protected demographic populations' from participating, while (2) creating a standardized and homogeneous study population (Epstein, 2008; Seidenfeld et al., 2008). The latter reason rests on the epistemic premise that reducing the number of variables will make it easier for researchers to deduce causal relationships, more precisely detect 'signal' from 'noise', and increase statistical power in the generalizability of its findings (Epstein, 2008; Kaptchuk, 2001). Thus, having a diverse demographic of participants would seem contradictory to the goal of reducing variability. However, I will show how diversity can actually serve to ensure more realistic forms of knowledge generation.

The Relation of Diversity to Knowledge Generation

As mentioned, the ultimate goal of a clinical trial is to create generalizable knowledge justified by statistically significant data, in a controlled environment through the empirical scientific process. Within this ideal, uniformity would increase the likelihood of detecting the variable of interest (signal), while any remaining variability (noise) is accounted for by diluting, factoring out or isolating (aka 'washing out') its influence from the variable (Epstein, 2008). In other words, having variability in the population is accepted so long as the variable of interest remains uniformly characterizable. According to this premise, having uniformity in demography would increase the generalizability for that chosen demographic, while decreasing or ignoring the generalizability for other demographics. However, the issue is that these signals (in racial diversity for example) would not be generalizable outside of the specific racial demographic that is overly recruited.

Even if the White racial demographic was not explicitly chosen by researchers as a variable of interest, conducting clinical trials according to this standardized process implicitly

increases this White demographic as a variable of interest (signal) at the expense of other racial demographics that are considered noise. This is an issue if demographics like race have an effect on clinical outcomes measured in trials. For example (Fazal et al., 2021), a retrospective study of 100 Black patients found that they had a 3-fold higher risk of cardiotoxicity compared with non-Black patients when treated with doxorubicin drug (Cooter et al., 1992). Another study of 216 patients with breast cancer showed that even after controlling for baseline cardiovascular risk factors, Black women were 2 times more likely to develop cardiotoxicity from trastuzumab drug use (Laventhal et al., 2012). It is also well-established that social determinants of health *do* have a measurable impact on health outcomes at the biological level. Percieved racial discrimination can also exacerbate health disparities and contribute to overall worse health outcomes through biological stress mechanisms (Williams & Mohammed, 2009). In this sense, racial differences become a useful variable of interest in clinical trials if the health outcomes can be attributed directly to racial difference.

Furthermore, it is also known that bodies of knowledge differ between social demographics, as *what* is prioritized in medical knowledge or *how* the knowledge is utilized can be drastically different between demographics. The most evident example is pharmacokinetic differences between age and sex, where young pre-menopausal women and older post-menopausal woman differ in biological characteristics (fat distribution, metabolic rate, hormone cycles, etc.) such that clinical parameters like drug response need to be independently defined (Gray, 2007). Average weight differences between races can also affect dosage and metabolic rates of drugs, changing the way dosage may be standardized for the 'average adult population' (Schwartz, 2007). Thus, it is important to include a balanced diversity of subpopulations within a specific demographic category of interest (e.g. race) to ensure that the clinical knowledge for individual races can be generalized in a statistically significant manner.

As the ultimate goal of the clinical trial process is to create generalizable knowledge, the same epistemic gap translates to the kind of knowledge created. Especially when the demographic being studied has historically benefited from being the main demographic of focus in Western medicine, the exclusion of other demographics becomes a further way to exacerbate this epistemic gap. Besides race, the same logic applies to other demographic categories, such as

sex, where sex diversity means having a statistically significant sample population that is generalizable to all sexes. For situations where human differences may plausibly cause different outcomes, then it should entail that the sample population in clinical trials must be diverse and representative of all people.

Social Versus Biological Factors of Clinical Outcome

Another consideration to epistemic reasons for diversity involves reframing the tension between social versus biological categorizations of diversity. When researchers think of the types of diversity categories, they may think of social delineations of diversity such as culture and ethnicity, or biological delineations of diversity such as age. Some researchers believe that social variables should be separated or excluded from clinical research for the sake of examining discrete, biological variables instead (Bartlett et al., 2005). However, taking a blinded approach to social variables does not mean that these social factors do not exist, nor that they do not impact the clinical knowledge generated. Efforts to factor out social diversity for the sake of isolating the 'true cause or effect' of the variable of interest effectively blinds researchers to the impact that these social demographics may have on clinical outcome, even for early phase trials that focus on efficacy. For example, it is well established that race and ethnicity correlates with health differences: in the U.S., Black and Hispanic populations report higher rates of undiagnosed diabetes than White populations, while Asian populations had increased rates of undiagnosed hypertension and diabetes (E. J. Kim et al., 2018). American Indian and Alaska Native adults are 60 percent more likely to have a stroke than White adult populations (Russell, 2010). The same correlations can be made according to other social factors like gender or wealth: women develop Alzheimer's Disease, depression and Osteoarthritis at higher rates compared to men (Buvinić et al., 2006), while lacking health insurance is associated with increased odds of undiagnosed diabetes and hyperlipidemia (Russell, 2010).

At the aggregate level, health disparities between social demographics show correlations to clinical outcomes or variables of interest in clinical trials, such as blood pressure, spirometry, and drug responsiveness, whether the trial itself is concerned with efficacy or effectiveness. For example, ACE inhibitors are generally less effective in treating heart failure in Black patients

compared with White patients, likely due to low pre-existing activity of the renin-angiotensin system in Black patients (Ramamoorthy et al., 2015; Taylor & Ellis, 2002). However, it is important to note that racial categorization acts as only correlational markers for multi-faceted factors responsible for individual responses to drug therapy (Taylor & Ellis, 2002). The presence of untreated co-morbidities, race/ethnic or gender differences in self-perception and behavioral norms can also influence the report of symptoms and clinical metrics like pain, mental wellbeing, sensitivity to adverse drug reactions, etc (Sharath et al., 2016; Valdes-Stauber et al., 2013). For example, health issues like depression or autism have varying rates of prevalence between men and women. Sex and gender-associated stigmas and behaviours affects how these conditions are manifested, diagnosed, and treated (Franconi et al., 2007). The inclusion of all demographic identities becomes important to integrate in clinical research, as individual clinical outcomes become intricately influenced by biological and social factors. Even if the condition is primarily seen in one demographic, its markedly different response warrants the inclusion of a diversity of populations for the generation of comprehensive knowledge.

It is important to emphasize that these social disparities are *correlational* - they do not necessarily *cause* the actual health conditions, and any identified differences may not apply to all members of each demographic (Taylor & Ellis, 2002). For example, it is not the case that the underdiagnosis of cardiovascular health conditions in Black populations is caused by every Black person's genetics or biology that makes their condition somehow harder to detect. Rather, it could be that past ethical violations led to mistrust of medical authorities in Black communities, or systemic discrimination limited their ability to afford health insurance thereby preventing them from seeking medical care for chronic health issues (Sullivan, 2020). The reality is that these social identities intersect and create multi-faceted differences in health outcome, and it is not easy to discern the consequence it has on a person's physiology, behavior, psychological state - factors all relevant to a clinical context (Haslam et al., 2005). These sentiments are echoed by scholars like Quayshawn Spencer, who argue that social demographic classifications of race can be useful in medicine, as race-based effects for example do not behave in a mutually exclusive manner between solely biological or social causes (Q. Spencer, 2015a).

In lieu of this epistemic knowledge gap, the best way to account for these variables would be to include a diverse set of social demographics in samples recruited for clinical trials, thereby implicitly integrating the effect of demographic variance within the framework of analysis. In doing so, researchers are able to increase the generalizability and thus contextualize the applicability of their knowledge outside the silo of the artificially-controlled clinical trial setting, which is the ultimate purpose of the clinical trial anyway.

The Essentialism of Social Demographics in Clinical Knowledge

Ethics aside, some researchers have argued that establishing diversity in clinical trials may not be an essential goal for scientific research (Epstein, 2008; Rivara & Finberg, 2001). Having participant diversity in race, for example, does not bridge any knowledge gap about how biological processes function in all humans - if a biological process occurs in White populations, we can assume that the same process should happen in people of other races because 'Whiteness' is not a determinant of this biological response (Epstein, 2008). On the other hand, necessitating racial diversity implicitly legitamizes the idea of racial profiling along biological or genetic basis. In other words, stratifying clinical knowledge based on race implies that race (whether socially or biologically-defined) is a causal determinant of this clinical knowledge in some way, which is problematic as it can justify race essentialism and discriminatory medical practises based on race in reality (Epstein, 2008; Osborne & Feit, 1992). The same argument can be about other demographic delineations like sex or age.

My sentiments against this belief are two-fold. Firstly, some demographic categories like sex (biologically defined as XX, XY, or any variation of X and Y chromosomes) or age *are* biologically or genetically determined. Sex differences often translate to differences in metabolism, physiology and thus clinical parameters like drug response between sexes (Epstein, 2008). For example, women tend to have increased insulin secretion compared to men based on postprandial insulin and C-peptide levels, likely due to the differences in estradiol regulation between the two sexes (Basu et al., 2006). In a similar manner, age roughly delineates different levels of biological, physiological and genetic fitness, thereby translating to differences in metabolism and thus drug response between ages as well (Epstein, 2008). Therefore, including a

diverse population along these biological determinants is needed to establish robust, sound and generalizable knowledge for the benefit of all types of humans.

Secondly, for demographics like race, there is mixed evidence that race has discernable biological differences (Bierer et al., 2021; Kamin Mukaz et al., 2020). Even so, it is not necessary to commit to a biological or genetically-determinant view of health to justify epistemic reasons for social diversity. For reasons explained in the previous section, we have seen that social categories like race or wealth correlate with differences in health prevalence and outcome. Thus, race as a clinical metric is only useful insofar as we recognize that the 'effect' of race is actually a reflection of these other social determinants of health (like wealth, geographical location, education level, cultural practices) that interact with the treatment effects of drugs and public health interventions.

Epigenetic Perspectives on Diversity

So far I have discussed the ethical reasons for diversity, where the historical context of exclusion combined with providing access and justice for these populations justifies the need for a diverse population. These reasons complement epistemic reasons for diversity, as it allows clinical research to account for biological and social factors. In this section, I use epigenetics to show how the human epigenome can reshape our understanding of the social and biological basis of clinical knowledge to ultimately integrate ethical and epistemic reasons for diversity.

The ethical and epistemic reasons for diversity alludes to a question of whether social or biological determinants of health have a quantifiable impact on clinical outcomes. If social determinants of health can be detected within a person's biology, then the ethical and epistemic reasons for diversity become the same and diversity can be seen as a way to integrate these social and biological differences. However, there are differing opinions among experts if social factors actually cause, be detected, or have an influence on biological differences, and vice versa (Hahn & Stroup, 1994; Williams, 1997). For example, David Williams, Robert Hahn, and Donna Stroup argued that science relied on outdated constructs of race that leads to a lack of validity, consensus, and reliability in its use (Hahn & Stroup, 1994; Williams, 1997), while modern scientists increasingly touted the clear trends in biological states and biomarkers that differed by

continental race (Martinson et al., 1997; Maugh II, 1996; Wald, 2000). The philosopher of biology Quayshawn Spencer argued that social concepts of race are becoming shaped by population genetics of race theory and vice versa. For example, the population geneticists' use of racial groups like 'Native American', 'Asian', and 'Pacific Islander' to describe trends in certain genetic population clusters is not the same as the Native Americans, Asians, and Pacific Islanders social race categories understood by ordinary Americans, though scientists often seek to justify the latter with the former (Q. Spencer, 2015b).

It is important to know that up until the early 21st century, biological delineations of race, sex, or age were mainly justified based on the genetic paradigms of the time. Namely, that differences in race, sex or age are constituted within the DNA of an individual as polymorphic markers or variations from some ideal reference genome, where single changes in the genetic code cascades into the greater biological differences seen. If differences in sex, race and age could be substantiated entirely based on genetics, then the social determinants of sex, race, and age would be insiginificant or ignored entirely in research. However, as we've seen it is difficult to separate the biological from the social determinants of sex, race, and age. As a middle ground, it appears that epigenetics is emerging since 2010 onwards as a novel field of knowledge which can bridge the disjunct between social and biological factors of health outcome. Epigenetics is defined as the chemical modifications of DNA (such as DNA methylation) and the nuclear positioning of higher-order DNA structures and histone modifications, acting in concert with regulatory elements such as non-coding RNA to register, signal, or perpetuate activity states of DNA (Joly et al., 2021). Epigenetics is also becoming increasingly used for biological profiling of age (i.e. loss of histones or methylation pattern as an indicator of age), sex (i.e. genomic imprinting of methylation patterns), or biological diseases like cancer (Biliya & Bulla, 2010; Piferrer, 2021; Sharma et al., 2010).

Interestingly, emerging research suggests that social disparities may be detectable within a person's epigenome. For example, a U.S. study found perceived racial discrimination was discernable as DNA methylation differences in African American women (Barcelona de Mendoza et al., 2018). Furthermore, gender differences in disease susceptibility may be due to epigenetic-mediated activity associated with gendered behavioral differences, such as smoking

(which predominantly affects males) causing cancer development, lung, cardiovascular, and fertility changes detectable within the epigenome different from female smoking habits (Fragou et al., 2019; Kaminsky et al., 2006). Other findings suggest that gender (defined as a social component and involves differential expectations or treatment by others) may leave an epigenetic imprint in the brain apart from sex differences (Cortes et al., 2019). Finally, on the permanence of epigenetics, these markers can be inherited but can also change over time in response to shifting environmental signals (Joly et al., 2021; Lacal & Ventura, 2018; Zhang & Sirard, 2021).

Emerging evidence suggests that these epigenetic changes, whether due to social or biological inputs, can also be 'inherited' through several generations between parents and children (Zhang & Sirard, 2021). A recent study provides evidence for transgenerational and intergenerational inheritance of epigenetic states in humans between familial generations, even without direct biological exposure to the epigenetic modifier (Breton et al., 2021). For example, there is an association between the grandmother's exposure to famine during pregnancy and the grandchild's development of poor cardiometabolic disorders like type 2 diabetes and hyperglycemia by adulthood (Breton et al., 2021). Parental exposed trauma can also lead to metabolic disorders: post-traumatic stress disorder from exposure to the Tutsi genocide has been associated with NRC31 epigenetic modifications in the HPA axis (the pathway that modulates stress and regulates processes like the immune, digestion, and endocrine systems) of both mothers and their offspring (Breton et al., 2021). Adult offspring of Holocaust survivors show reduced DNA methylation and increased gene expression in FKBP5, a gene associated with glucocorticoid sensitivity (Breton et al., 2021). These findings indicate that the social injustices and mistreatment of particular populations, even when they have occurred historically in previous generations, can produce downstream measurable, biological effects.

Although epigenetics is still a novel field, these preliminary findings ground ethical reasons for diversity with measurable biological markers that impact the epistemology of clinical knowledge-making. Epigenetics has shown that social factors, such as differential treatments of races or genders, can interact and modulate the epigenetic state of an individual, and vice versa. Just like in genetics, the exact epigenetic state can then go on to influence a person's outcome in clinical parameters like drug response, metabolic rate, physiological fitness, etc. For example,

epigenetic differences can affect drug treatment by modulating the expressions of key genes involved in the metabolism and distribution of drugs as well as drug targets, thereby contributing to interindividual variation in drug response (Ivanov et al., 2012). By extension, the precise epigenetic difference is influenced by exposure to environmental, psychological or social factors, such as stress, smoke/air quality, diet, etc., which can change as the social factors change. Epigenetics thus reinforces the idea that social conditions warrant explicit consideration, inclusion, and analysis in scientific and clinical measurements as they have some quantifiable effect on biological outcome. Ensuring a demographically diverse set of participants is a way to account for this epigenetic difference in generating clinical knowledge.

Conclusion: Revisiting Diversity in Clinical Research

This section has explored the historical demographic evolution of U.S. clinical trials and provided some ethical and epistemic reasons for diversity. To actualize diversity, it entails notions of representation and equity within demographic subpopulations. Representation is recognizing, including, and characterizing the categorical differences between subpopulations of a given demographic category. For example, taking a sex and gender-sensitive approach to sex/gender representation means intentionally recruiting intersex and transgender individuals to be part of the study so that their biological, psychological or behavioural differences are represented in the sample. Fair representation also includes avoiding instances of token diversity, where one person of a demographic is only represented once for the sake of including that demographic. For example, recruiting one or a few African individuals in a genomics study to generalize all the African people's genomic profile is not adequate representation (Perry et al., 2018; Saey, 2021). Instead, fair representation means including individual populations to a statistically significant extent such that generalizable knowledge becomes epistemically valid for that demographic.

Equity entails thinking about how the representation of demographic sub-populations are contextual to histories of injustice, exclusion or harm experienced by that particular demographic. For example, an equitable approach to clinical trial recruitment of Black populations entails an explicit interest to increase their participation in Phase III trials as they

have been historically underrecruited in the past for these trials, while ensuring that they are not overrepresented in early-phase trials. Equity also entails fairness in clinical knowledge translation, by identifying gaps in knowledge which have generally been lacking or misunderstood for a specific demographic.

Diversity reveals multi-faceted ethical and epistemic issues with clinical trials participation. Solving issues in diversity inadvertently addresses the systemic, underlying reasons for underrepresentation, such as: strict eligibility criteria, clinician bias, challenges with cultural competence; access challenges related to cost of care and insurance; lack of trust between participants and clinicians and the health system as a whole; linguistic, cultural, and literary-related barriers; and factors with family and community engagement (Sternberg, 2020). Likewise, ensuring fair representation inadvertently implies restoring distributive justice for those who assume the risks and benefits of a trial, while increasing the generalizability of clinical knowledge by rooting the knowledge created in the social or biological realities of each demographic. Notably, there have been recent efforts advocating for 'pragmatic' clinical trials, which aims to assess effectiveness in real-world settings in real-world populations (Califf & Sugarman, 2015; Mentz et al., 2016). Ensuring diversity is thus a way for clinical trials to reflect a more accurate demographic composition of real-world populations alongside this sentiment.

Diversity in Cancer Clinical Trials

Cancer clinical trials often consist of drug or surgical interventions where the risk factors, incidence, prevalence, and progression of the cancers studied are all factors that can vary between patients and by extension, affect the outcome of the drug or intervention under study. These epidemiological indicators are intricately related to social determinants of health - many studies show racial-economic disparities between risk or prevalence of certain cancers, or that cancer incidence differs between sexes, for example. Thus, cancer clinical trials represent one of the main types of clinical research where accurately representing demographic data and ensuring a diversity of participants is critical to the integrity of the knowledge generated.

However, cancer trials have historically overrepresented young, White, wealthy, and male participants.[2] For example, a systematic review of Phase III cancer clinical trials conducted between 2001-2010 found that 82.9% of participants were White, while other races were underrecruited (Hamel et al., 2016). Other demographic studies suggest that contemporary cancer trials continue to overrepresent white, male participants below 65 years old, on average (Nazha et al., 2019). While there has been greater awareness in diversifying cancer clinical trials in the past decade, it is unclear if diversity has been successfully achieved on a systemic and widespread level today (Begun, 2021; Freestone, 2021; Stanford Medicine Cancer Institute, 2022).

The remainder of this thesis is dedicated to presenting an overview of diversity for race, sex, and age demographics for Phase III cancer clinical trials in the last 5 years. Phase III cancer trials were chosen because it is representative of the standard type of clinical research: cancer trials are the most common type of clinical trials, while the often large participant sample in Phase III trials offer statistically significant approximations for measuring the proportions of representation between demographic populations. The ultimate goal is to capture a snapshot on the state of diversity in modern U.S. clinical trials in order to retroactively assess current efforts in diversification and inclusion, contextualized by the goals of equity and justice for historically mistreated populations.

Methods

There has been recent interest in assessing the representation of various demographic populations in clinical trials (Eshera et al., 2015; Murthy et al., 2004; Poon et al., 2013; Unger et al., 2013; Varma et al., 2021), with this project being a continuance of these efforts for the most common type of clinical trial in health research. The aim of the study is to describe the demographic diversity in modern cancer trials by conducting a cross-sectional study using U.S. Phase III trial data from the last 5 years.

[2] In the cancer context, a young cancer patient is considered someone who is under 65.

Objectives

1. Collect demographic data on cancer clinical trial participants in U.S. Phase III trials occurring Jan 2017 - June 2021 for the top 10 most common cancers: Breast, Lung, Prostate, Colorectal, Melanoma, Lymphoma, Kidney, Leukemia, Pancreatic, and Liver.
2. Collect relevant demographic data on U.S. real-world cancer incidence[3], grouped by race, ethnicity, sex and age, for each cancer type studied.
3. Assess the diversity of various demographic indicators using statistical analysis to compare the representation of each demographic to incidence, for each cancer type and all cancers within the dataset. Demographic indicators include: race, ethnicity, sex, age, sponsorship, trial location, trial size. Other types of demographic categories, such as income level, health insurance status, and presence of coexisting morbidities, will be checked to see if they are reported as inclusion/exclusion criteria.

Hypotheses

Based on existing published literature on cancer trial demography and a preliminary analysis of pilot data, it is hypothesized that (1) a slight under-recruitment of women compared to men will occur, (2) the mean age of cancer trial participants will be skewed towards patients younger than the real-world average, and (3) the dominant race/ethnicity reported will be White, with under-recruitment and/or a lack of precise reporting in race and ethnicity data for visible minorities. In particular, American Indian/Native, Asian/Pacific Islander and Hispanic populations may be inaccurately reported, or not reported altogether.

Outcomes

The primary outcome is to generate representation scores and assess whether representation significantly deviates from real world prevalence for the age, sex, race and ethnicity of each cancer type studied, by normalizing demographic data from the clinical trials to the real-world reported data. These scores will be visualized in graph form for each cancer type.

[3] In epidemiology, 'incidence' is the term that refers to newly identified cases of a disease or condition per population over a specified timeframe.

The secondary outcome is to assess if other clinical trial factors can impact diversity of recruited participants. The following factors will be examined: particular type of lead funding status (NIH, Industry, or Other) correlates to more diverse participants, grouped by representation scores for sex, race, ethnicity and age. Using the same method, multi-center versus single-center trial location in the U.S. will be assessed to see if trials conducted at multiple trial locations correlates with greater diversity of participants. The total number of participants per trial will be grouped into n = 0-99, n = 100-999, and n = 1000+ categories to see if trial size impacts representation. Lastly, other types of demographic categories, such as income level, health insurance status and co-existing morbidities/chronic disabilities, will be checked to see if they are reported in clinical trial demographic data.

Search Strategy

Clinicaltrials.gov was used as the reference database as it primarily reports U.S./North American trials with adequate demographic detail.

(1) Eligibility Criteria

The eligibility criteria for clinical studies included the following: (a) the study condition was breast cancer, lung cancer, prostate cancer, colorectal cancer, melanoma, lymphoma, kidney cancer, leukemia, pancreatic cancer, or liver cancer, (b) Phase III trial, (c) the results were posted from 2017-2021, (d) at least one country's site location was in the U.S., and (e) studies were complete with posted results.

Cancer clinical trials were chosen as the clinical trial field of study as they represent the highest volume of clinical trials. The cancer focus was restricted to the top 10 most common types of cancers occurring in the U.S. population, chosen based on the U.S. prevalence data (NIH National Cancer Institute, 2021). Trials were grouped by cancer type to see if representation scores differ between different types of cancers. Phase III clinical trials were chosen as these trials often enroll a large number of participants, thereby increasing the statistical power for demographic analysis. Search results were restricted from 2017 onwards to focus on the most recent clinical trials data. The location was restricted to the U.S. as it has a diverse real-world

population that is representative of the greater context to restore justice and health equity for historically underrepresented populations. Due to federal diversity requirements and previously-focused research on this context, there is also an abundance of U.S. data for real-world comparison. Clinical studies were restricted to 'completed studies with results' to ensure that demographic data would be included in the report.

(2) Data collection

A search was conducted for each of the 10 aforementioned cancer types, leading to 10 searches with 246 total hits. The exact search parameters, URLs and hits are listed in Appendix A (see supplemental material). A pilot search was performed on June 25, 2021 to gauge the quantity of trials available and perform a preliminary data analysis. The search was performed again for data collection and extraction on July 19, 2021.

A PDF of each search result was downloaded on July 19, 2021 for recordkeeping and details of each individual trial were saved as an individual XML file. Baseline characteristics of demographic data relevant to the primary outcomes were extracted from the XML files using a Python script on August 13, 2021. A second data extraction was performed on September 2, 2021 for data relevant to the secondary outcomes, namely the funding status and number of enrolling locations. Other demographic categories (income level, health insurance status and co-existing morbidities/chronic disabilities) were manually extracted on Sept 20, 2021. All data were tracked on Excel for analysis. The script used for data extraction is given in Appendix B.

(3) Exclusion Criteria

The exclusion criteria for screening extracted studies were: (a) duplicate studies, (b) the trial's primary location was outside the U.S., (c) trials with less than 10 participants (d) trials with less than 3 studies per cancer type, (e) pediatric trials, and (f) breast and prostate cancer trials in sex analyses only. Trials that only focused on a specific race/ethnicity/age group were excluded for race/ethnicity/age analysis but included in sex analysis since representation scores are calculated on a per-trial basis and they would show up as outliers. Using trial identifiers, duplicate studies were removed. Given that the clinical data analysis would only be compared to real-word U.S. populations, studies were also removed if over 50% of trial centers were located

outside the U.S. (for example, trials that were sponsored in the U.S. but had over 50% of trial locations in another country). Studies with less than 10 total participants per trial, and studies with less than 3 trials per cancer type were excluded, as they would not have provided enough statistical power for analysis by a single cancer type. This resulted in kidney cancer and melanoma trials being excluded in assessments of individual cancer type as there were only 2 and 1 studies for each type respectively. However, the kidney and melanoma cancer studies were included in demographic analyses for all cancer types combined. To avoid detection bias and data skewing, pediatric cancer trials and trials targeting only one age group (e.g. those over 65 years old) were excluded. In similar fashion, one breast cancer trial was excluded because it only recruited Black women. All breast cancer and prostate cancer studies were excluded for the sex analysis, because these trials recruited primarily one sex, and thus would offer no significant comparison of demographic diversity between the sexes. Given how small the male proportion of US breast cancer patients is, an analysis of their representation would likely give wildly varying results, even in studies including a single man. Thus, it was judged that performing the sex analysis on breast cancer trials would give unreliable results. After applying the exclusion criteria, a total of 38 studies remained for sex analysis, 53 studies remained for race analysis, 29 studies remained for ethnicity analysis, and 35 studies remained for age analysis. A total of 85 studies remained for 'all cancer' pooled analysis. A summary of the total number of studies included for each analysis is provided in the table below.

Table 1: Characteristics of Included Studies

Type of Analysis	Total # of Included Trials	Total # of Participants (Sample Size)
Sex	38	10,622
Race	53	17,979
Ethnicity	29	10,121
Age	35	33,053
All Cancer Types	85	50,419
- Breast Cancer	25	29,527
- Colorectal Cancer	5	2,245
- Kidney Cancer	2	1,063
- Leukemia	4	1,155
- Liver Cancer	3	634

- Lung Cancer	7	2,099
- Lymphoma	5	1,224
- Melanoma	1	185
- Pancreatic Cancer	2	145
- Prostate Cancer	21	8,448
- Multiple Cancer Studies	10	3,694
Lead Sponsorship	85	50,419
Trial Location	85	50,419
Sample Size per Trial	85	50,419

Data Analysis

(1) Data Analysis for Primary Outcomes

The following baseline characteristics were extracted for each trial:

- Trial identifier
- Cancer type, based on the search parameters of breast, lung, prostate, colorectal, melanoma, lymphoma, kidney, leukemia, pancreatic, or liver cancer
- Total number of participants
- Sex data, given as the number of enrolled males and females, or any other sex designation as reported by the trial
- Race data, given as the number of enrolled White, Asian, Pacific Islander, Black or African American, Native, Other, or any other race designation as reported by the trial
- Ethnicity data, given as the number of Hispanic and Non-Hispanic, or any other ethnicity designation as reported by the trial
- Age data, given as the number of participants according to the age brackets reported by the trial, or the mean/median age or age range of participants
- Any eligibility criteria pertaining to age, if reported

A summary of the extracted data is listed in Appendix C.

The real-world U.S. cancer incidence for each type of cancer, stratified along race, ethnicity, age and sex was taken from (U.S. Cancer Statistics Working Group, 2021). Extracted characteristics for each cancer type include:

- Race and ethnicity data, given as case count for 'White', 'Black', 'American Indian and Alaska Native', 'Asian and Pacific Islander', or 'Hispanic'
- Sex data, given as case counts for 'Male' or 'Female'
- Age data, given as case counts grouped by 5-year age brackets, but extracted as mean age
- Total number of cases for Race/Ethnicity, Sex and Age

A summary of the extracted data is listed in Appendix D.

All statistical analyses were performed using *R* (R Core Team, 2021) and plots were generated using the *ggplot2* package (Wickham, 2016)**.**

(1a) Sex Analysis

The sex analysis, as described by Poon et al. (2013), was completed by comparing the proportion between the clinical trial data and real-world incidence. First, the proportion for clinical trials was calculated as:

$$Proportion\ (trial)\ =\ \frac{number\ of\ males\ in\ the\ trial}{total\ number\ of\ enrolled\ participants};$$

with the process repeated for females, and calculated for each clinical trial. Next, the proportion for real-world incidence was calculated as:

$$Proportion\ (world)\ =\ \frac{number\ of\ male\ cases\ in\ the\ real\ world}{total\ number\ of\ cases};$$

with the process repeated for females, and calculated for each cancer type. Lastly, a representation score was calculated by comparing the proportions of each trial to the real world:

$$Representation\ score\ =\ \frac{proportion\ (trial)}{proportion\ (world)}$$

Statistical analysis was performed to test for significant difference between the trial data and real-world incidence. A t.test($representation_scores, "mu"=1) function was computed in *R*, with p-values listed for each cancer type.[4] Another t-test for overall sex representation was performed for all cancer types, by pooling the representation scores for all trials together. All results are given in Appendix E.

(1b) Race & Ethnicity Analysis

The race analysis was conducted in the same manner as the sex analysis, using the race categories of 'White', 'Black', 'American Indian and Alaska Native', 'Asian and Pacific Islander'. The ethnicity data was also conducted in the same manner, using the ethnicity category for 'Hispanics'. The categories of "Unknown" or "Other" race/ethnicities were reported in Appendix F but a representation score was not calculated as there lacked real-world incidence data to compare these demographics to. The t.test(representation_scores, "mu"=1) function was computed in *R*, with p-values listed for each cancer type. Another t-test for overall race & ethnicity representation was performed for all cancer types, by pooling the representation scores for all trials together. All results are given in Appendix F.

(1c) Age Analysis

As age was reported in different ways for each trial with varying degrees of detail, the age analysis could only be performed for trials that reported the mean age. The mean age for real-world cancer incidence was calculated for each relevant cancer type, calculated as:

$$Mean\ Age\ (USA)\ =\ \frac{0.5*Count_0+2.5*Count_1+\sum_{i=1}^{17}(i*5+2)*Count_{i+1}}{\sum_{i=0}^{18}Count_i};$$

[4] The t-test compares the mean and standard deviation of clinical trial representation scores, given as a p-value. A p-value less than 0.05 (the typical significance threshold) signifies that there is less than 5% chance that the difference between the mean representation score of the clinicals versus real-world population (with an ideal score of 1) is equal to 0 (in other words, the results are due to chance). This means the calculated representation score is a statistically significant difference from the ideal score of 1.

where $Count_i$ = # of patients within age group i

As the age of real-world incidence was only reported in sub-groups of 5-year age brackets, it was assumed that the age of individuals within each age bracket was equal to the mean age of the bracket. For the 85+ group, this age was assumed to be 87 in statistical analysis.

The mean age difference was computed as:

$$Mean\ Age\ Difference\ =\ Mean\ Age\ (trial)\ -\ Mean\ Age\ (USA)$$

The t.test($mean_age_difference, "mu"=0) function was computed in *R*, with p-values listed for each cancer type. Another t-test for overall age representation was performed for all cancer types, by pooling the representation scores for all trials together. All results are given in Appendix G.

(2) Data Analysis for Secondary Outcomes

The following baseline characteristics for each trial were extracted for secondary analysis:

- Lead funding status, given as industry (signifying private funding), NIH (signifying public/government funding) or other (often university-funded)
- Location type, given as multi-center (>1 trial locations) or single-center (1 trial location)
- Any mention of other demographic categories, such as income level, health insurance status and co-existing morbidities/chronic disabilities in the inclusion/exclusion criteria or demography section of the trial

A summary of the extracted data is listed in Appendix H.

(2a) Lead Sponsorship Analysis

All cancer trials were pooled together, and grouped by sponsorship status (NIH, Industry, Other). The ANOVA test[5], aov($representation_scores ~ $sposorship_type), was computed in *R*,

[5] Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.

comparing the representation between each pair of sponsorship categories for each demographic category delineated by race, sex and age. For example, one t-test was run for the 'male' sex category, to see if the 'representation score' differed significantly between NIH and Industry-sponsored trials, then repeated for the 'female' sex category. The results from these two tests would indicate if a specific funding status had an impact on sex representation. Another t-test test was run for the age category, to see if the 'mean age difference' differed significantly between NIH, Industry, and Other-sponsored trials. The p-values for each demographic category comparison are listed in Appendix I.

(2b) Trial Location Type Analysis

All cancer trials were pooled together, and grouped by trial location type (single-center, multi-center). The t.test(representation_score_single_site, representation_score_multi_center) function was computed in *R*, comparing the representation between location types for each demographic category delineated by race, sex and age. For example, a t-test was run for the 'male' sex category, to see if the 'representation score' differed significantly between multi-centered and single-centered trials, and repeated for the 'female' sex category. The results from these two tests would indicate if location type had an impact on sex representation. Another t-test was run for the age category, to see if the 'mean age difference' differed significantly between multi-centered and single-centered trials. The p-values for each demographic category comparison are listed in Appendix J.

(2c) Sample Size per Trial Analysis

Each trial was sorted into one of three categories based on the total number of participants recruited for each trial: n = 0-99 participants, n = 100-999 participants, and n = 1000+ participants. The analysis of variance, aov($representation_score ~ $paticipant_count) function was computed in *R*, comparing the representation between the three participant categories for each demographic delineated by race, sex and age. For example, an ANOVA was run for the 'male' sex category, to see if the 'representation score' differed significantly between trials with <100 participants, 100-999 participants and >1000 participants, then repeated for the

'female' sex category. The results from these two tests would indicate if the relative size of the trial had an impact on sex representation. Another ANOVA test was run for the age category, to see if the 'mean age difference' differed significantly between sample size per trial. The p-values for each demographic category comparison are listed in Appendix K.

(2d) Other Demographic Category Analysis

Using Python, every trial was searched for keywords and phrases within the inclusion and exclusion criteria that may explicitly or implicitly exclude certain populations from the trial. The conditions/keywords searched were:

- Women must not be pregnant, breastfeeding/nursing, or of childbearing potential
- No other malignancies
- No mental condition, psychiatric or addictive disorders
- No other concurrent investigational drug use
- No other […] condition
- Negative HIV status
- Performance scores (ECOG, Zubrod, Karnofsy)
- Unable to read or speak English
- Insured or insurance status
- Life expectancy of [...] years or months
- Dependent on wheelchair/walker for mobility

A summary of the results was compiled in Appendix L.

Results

(1) Primary Outcomes: Sex, Race, Ethnicity and Age Data

(1a) Sex Diversity

Using values from Appendix E, a box plot was generated from the sex proportions and representation scores, comparing the diversity of sex categories for each cancer type:

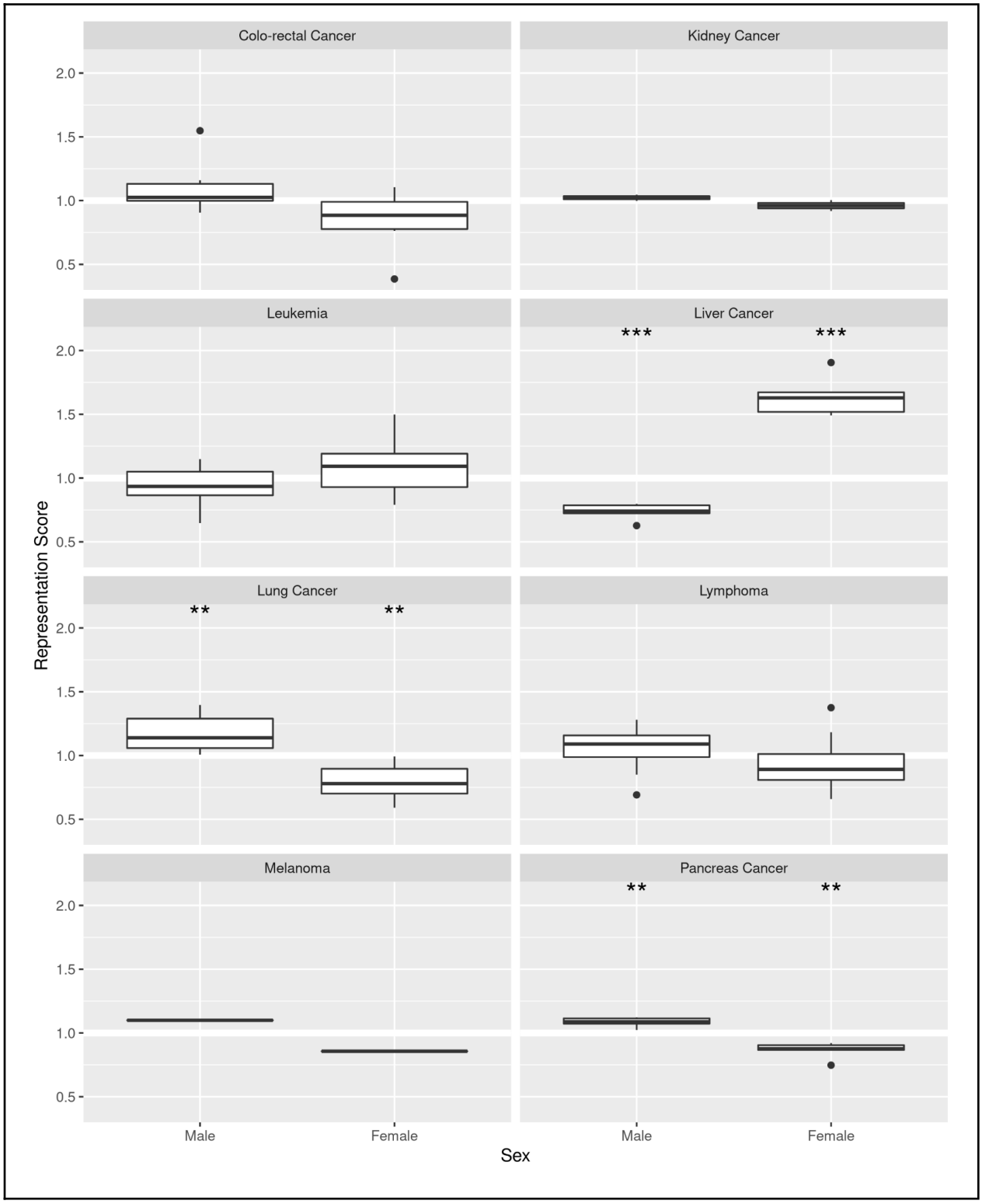

Figure 1: Sex Representation Scores, Stratified by Individual Cancer Type

Box plots show the distribution of the representation scores for each category, plotted along the y-axis. The boxes and vertical lines show the total range of representation scores (boxes: 2nd & 3rd quartile, vertical lines: 1st & 4th quartile) and single dots represent the outliers in a given dataset. Sample means were compared to the target value of 1 using t-tests (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$).

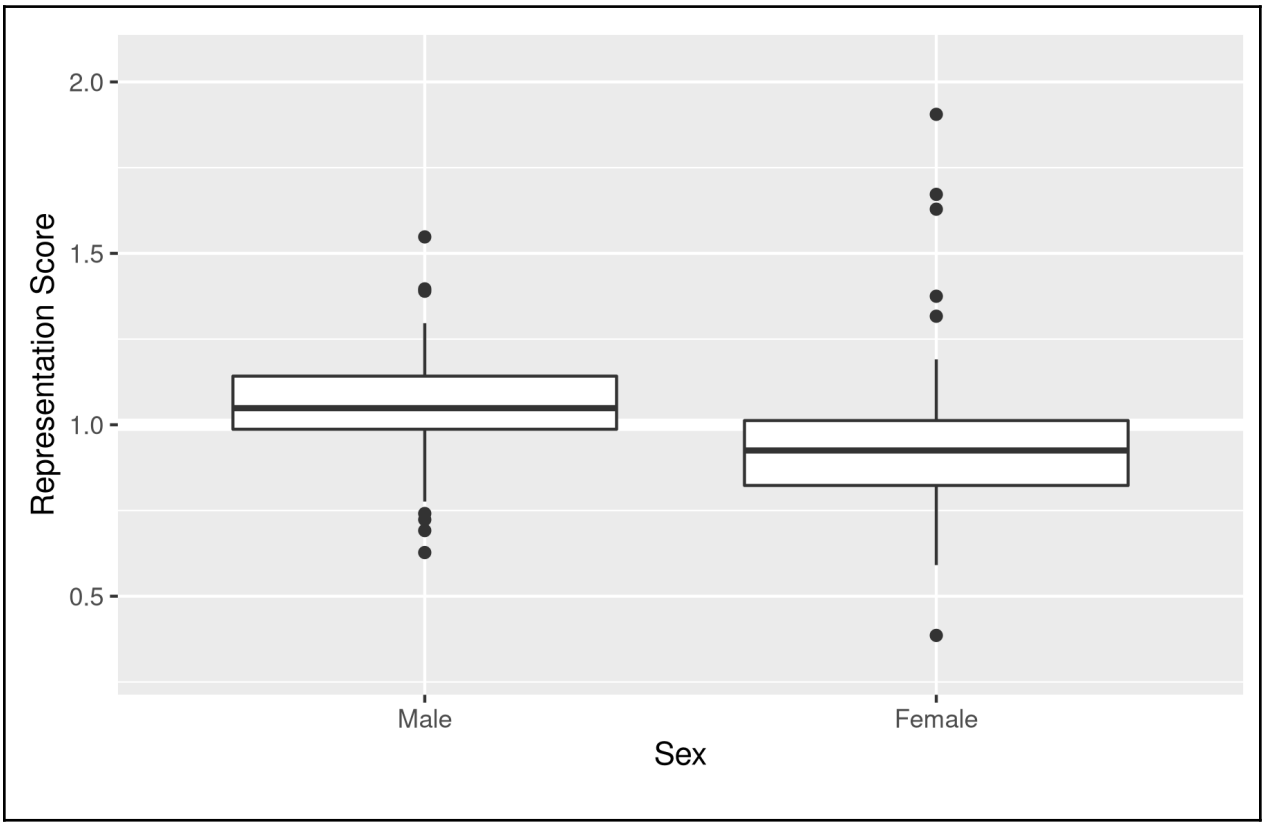

Figure 2: Sex Representation Scores, for All Cancer Types Combined
Box plots show the distribution of the representation scores for each category, plotted along the y-axis. The boxes and vertical lines show the total range of representation scores (boxes: 2nd & 3rd quartile, vertical lines: 1st & 4th quartile) and single dots represent the outliers in a given dataset. Sample means were compared to the target value of 1 using t-tests (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$).

Based on the results, males are significantly overrepresented in lung and pancreas cancer trials, whereas females are overrepresented in liver cancer trials. Sex representation isn't significantly different from 1 when combining all trials together, but trends towards male overrepresentation. No other categories for sex or gender designation were reported by the trials.

(1b) Race & Ethnicity Diversity

Using values from Appendix F, a box plot was generated from the race/ethnicity proportions and representation scores, comparing the diversity of race/ethnicity subcategories for each cancer type:

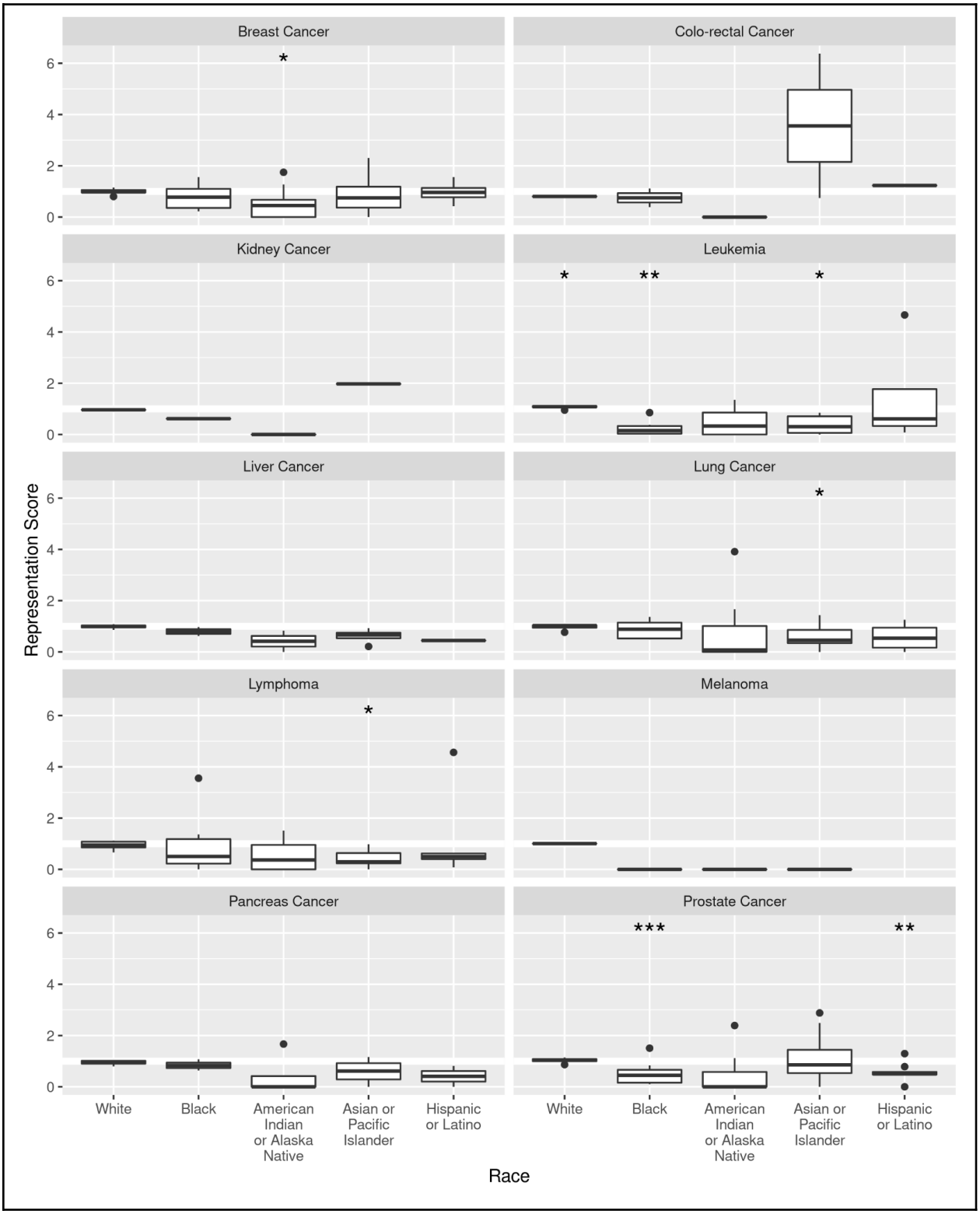

Figure 3: Race & Ethnicity Representation Scores, Stratified by Individual Cancer Type
Box plots show the distribution of the representation scores for each category, plotted along the y-axis. The boxes and vertical lines show the total range of representation scores (boxes: 2nd & 3rd quartile, vertical lines: 1st & 4th quartile) and single dots represent the outliers in a given dataset. Sample means were compared to the target value of 1 using t-tests (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$).

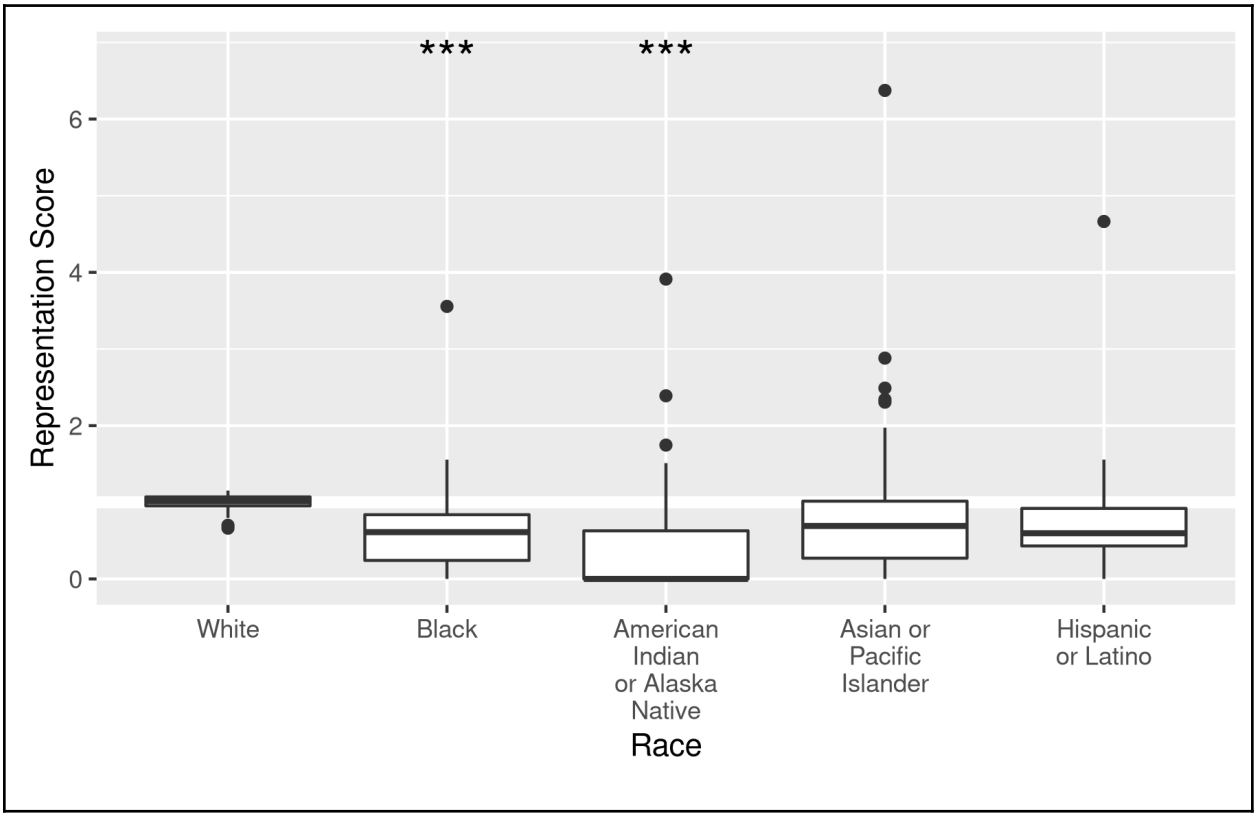

Figure 4: Race & Ethnicity Representation Scores, Stratified by All Cancer Type
Box plots show the distribution of the representation scores for each category, plotted along the y-axis. The boxes and vertical lines show the total range of representation scores (boxes: 2nd & 3rd quartile, vertical lines: 1st & 4th quartile) and single dots represent the outliers in a given dataset. Sample means were compared to the target value of 1 using t-tests (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$).

Based on the results, American Indian/Alaska Native and Black populations are underrepresented in all cancer types, while Asian/Pacific Islanders are underrepresented in leukemia, lung, lymphoma, melanoma, and pancreas cancers. General trends indicate Asian/Pacific Islanders are overrepresented in kidney, colorectal and prostate cancers, although the p-values for breast, colorectal, populations, liver, melanoma and prostate cancers are not statistically significant. Hispanic populations are underrepresented across all cancer types, although the trend was only statistically significant for prostate cancer. White populations are either fairly represented[6] or overrepresented for each cancer type, although the differences in proportions are not statistically significant. For all cancer types, all non-White populations are underrepresented, with American Indian/Alaska Natives having the worst representation

[6] 'Fairly represented' is when the proportion of a certain demographic population (i.e. White race) in clinical trials matches the proportion of the same demographic in the real world.

followed by Black populations. These results are consistent with the hypothesis that racial minorities are generally underrepresented compared to White populations with the accuracy of reporting becoming an issue for calculating statistically significant representation scores.

(1c) Age Diversity

Using values from Appendix G, a box plot was generated to compare the mean age difference for each cancer type:

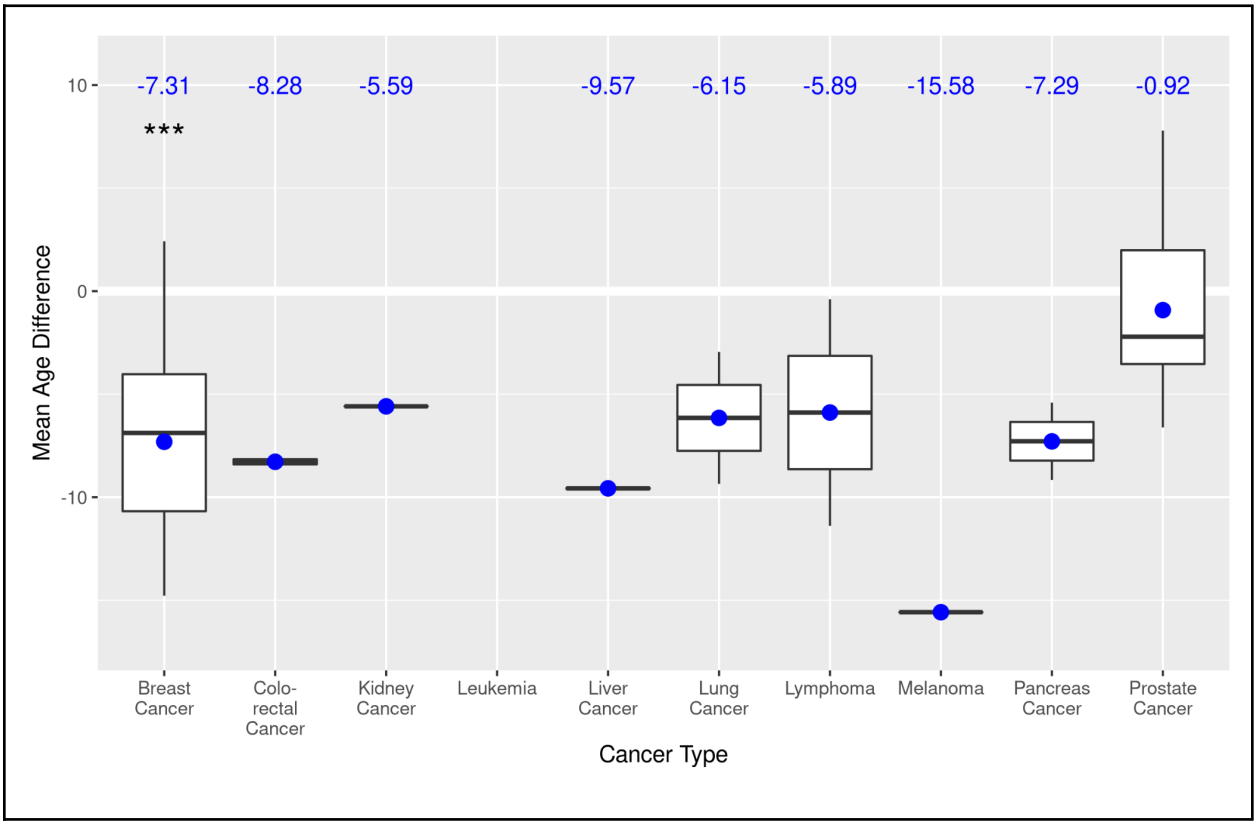

Figure 5: Mean Age Differences, Stratified by Individual Cancer Type

Box plots show the distribution of the difference between the mean age of enrollment in trials and the real-world mean age of onset for each category, plotted along the y-axis. Negative values indicate that trial participants are younger than the average age of onset for a given category. The boxes and vertical lines show the total range of age differences (boxes: 2nd & 3rd quartile, vertical lines: 1st & 4th quartile) while the blue dots and numbers show the mean age difference in a given dataset. Sample means were compared to the target value of 0 using t-tests (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$).

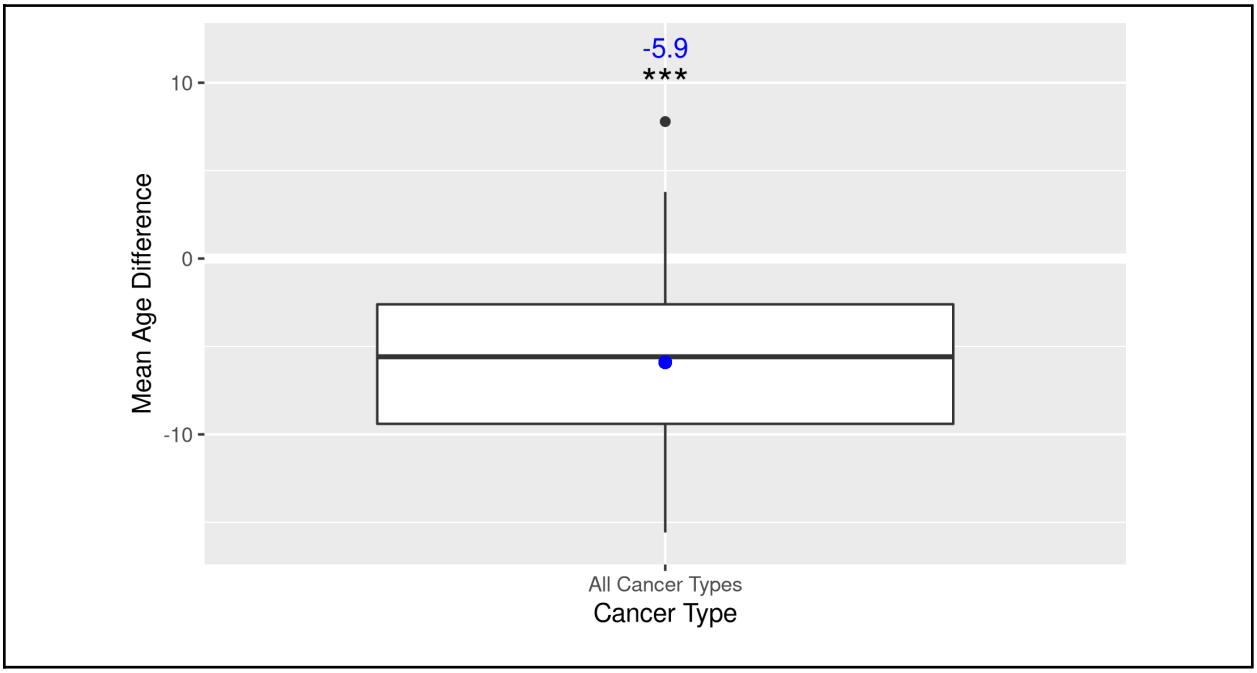

Figure 6: Mean Age Differences for All Cancer Types
Box plots show the distribution of the difference between the mean age of enrollment in trials and the real-world mean age of onset for all trials, plotted along the y-axis. Negative values indicate that trial participants are younger than the average age of onset for their particular condition. The boxes and vertical lines show the total range of age differences (boxes: 2nd & 3rd quartile, vertical lines: 1st & 4th quartile) while the blue dots and numbers show the mean age difference in a given dataset. Sample means were compared to the target value of 0 using t-tests (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$).

Based on the results, participants are on average younger than the mean real-world cancer incidence for each cancer type, with an average participant mean age of 5.9 years younger than the real-world incidence for all cancers.[7] Notably, melanoma cancer trials recruit the relatively youngest patients (15.58 years below the real-world average) while prostate cancer trials recruit the best age-balanced patients (0.92 years below the real-world average). Other trials recruit about 5-10 years younger than the real-world average. This is consistent with the hypothesis that cancer trial populations are skewed towards younger patients.

(2) Secondary Analysis: Sponsorship Status, Trial Location Type, Sample Size and More

(2a) Lead Sponsorship Status

Using values from Appendix I, a box plot was generated to compare the diversity of sex, race/ethnicity and age based on funding status:

[7] Pediatric cancer trials and trials which only recruited older patients (65+ y.o.) were excluded from the age analysis.

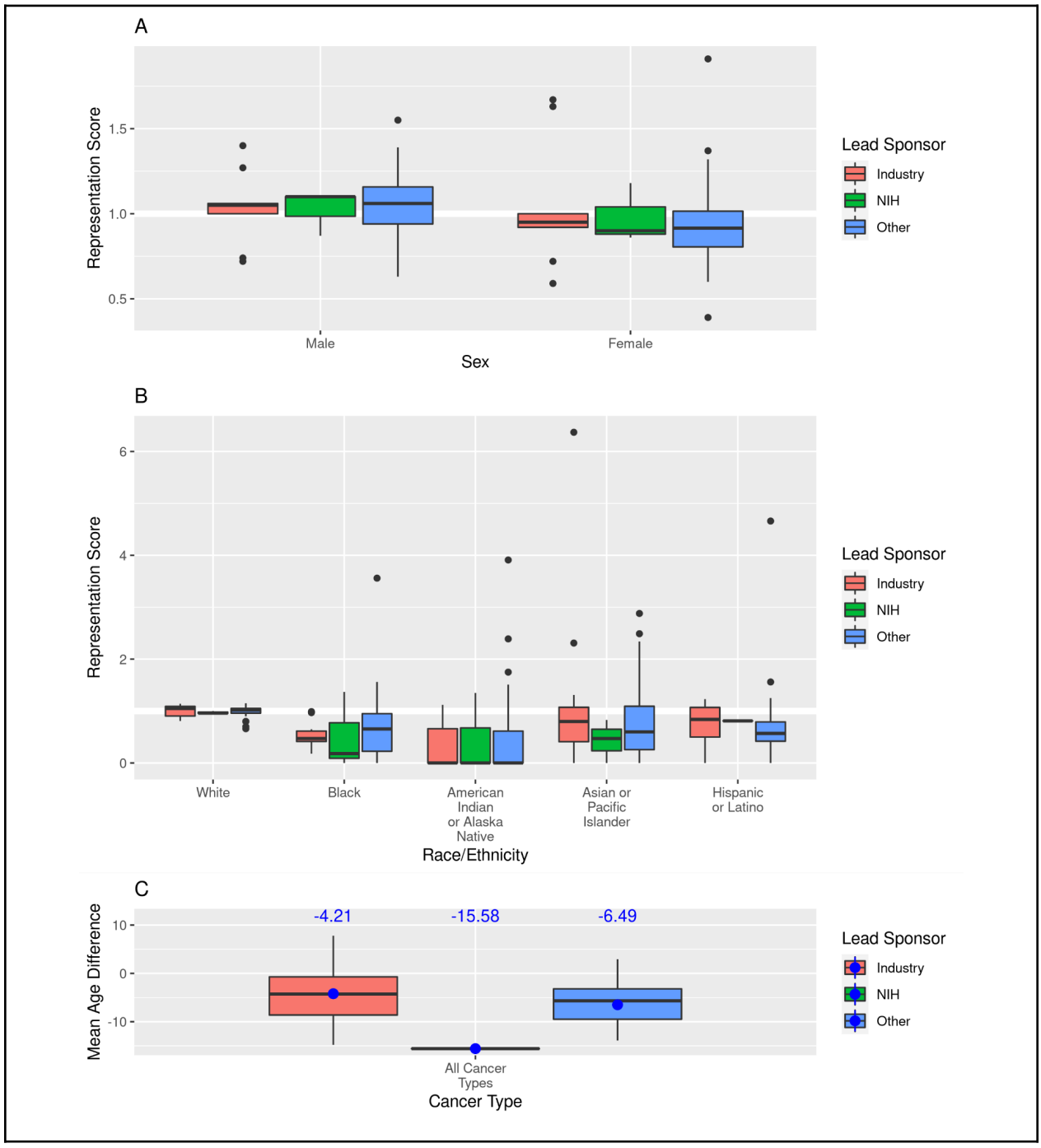

Figure 7: Sponsorship Representation Stratified by Sex, Race, and Mean Age Difference
Box plots show the distribution of the representation scores for each category, plotted along the y-axis. The boxes and vertical lines show the total range of representation scores (boxes: 2nd & 3rd quartile, vertical lines: 1st & 4th quartile) and single black dots represent the outliers in a given dataset. Blue dots and numbers show the mean age difference. Differences in the distribution between sponsorship status were analyzed using one-way ANOVA (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$). A: Comparison of sex demographics representation scores; B: Comparison of race/ethnicity demographics representation scores; C: Comparison of mean age difference

Based on the results, sponsorship status does not have any significant effect on sex, race, or age representation. Women are generally underrepresented between all sponsorship types. For race data, White populations are fairly represented across all sponsorship types. Black and American Indian/Alaska Natives populations are underrepresented in all types of trials. Trends seem to indicate that Asian/Pacific Islanders are more fairly represented in industry-sponsored trials than in NIH-sponsored ones. Hispanics are underrepresented regardless of sponsorship status. Of note is the low number of reports of participants ethnicity in NIH trials. For mean age difference, the data suggests that the 'Other' sponsorship has a slightly younger mean age recruitment in comparison to industry trials (6.49 years younger for "Other" versus 4.21 years younger for "Industry"), with NIH-sponsored mean age data being generally unavailable.

(2b) Trial Location Type

Using values from Appendix J, a box plot was generated to compare the diversity of sex, race/ethnicity and age based on trial location type:

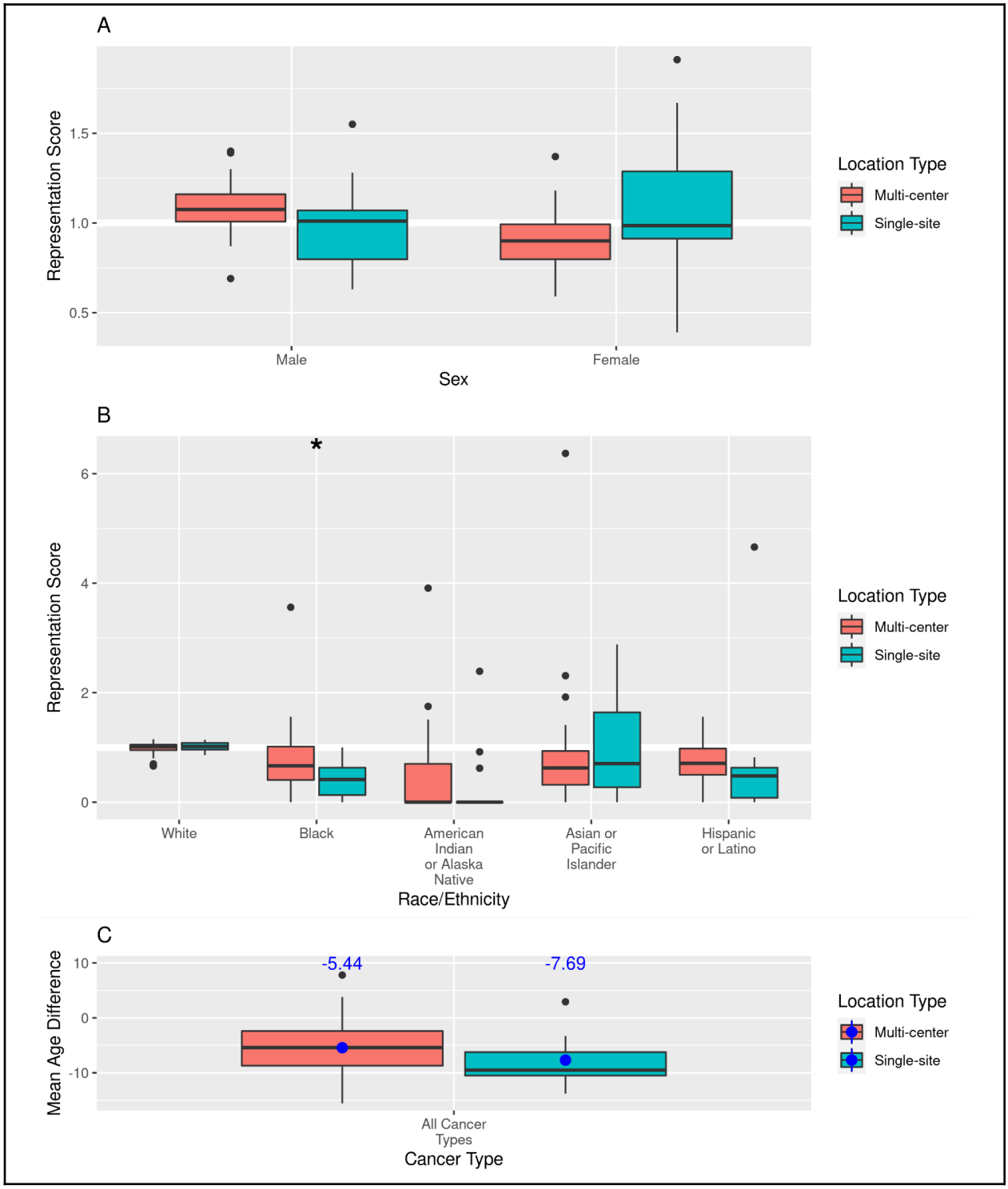

Figure 8: Trial Location Representation Stratified by Sex, Race, and Mean Age Difference
Box plots show the distribution of the representation scores for each category, plotted along the y-axis. The boxes and vertical lines show the total range of representation scores (boxes: 2nd & 3rd quartile, vertical lines: 1st & 4th quartile) and single black dots represent the outliers in a given dataset. Blue dots and numbers show the mean age difference. Differences in the distribution between location type were analyzed using t-tests (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$). A: Comparison of sex representation scores; B: Comparison of race/ethnicity representation scores; C: Comparison of mean age difference

Based on the results, the only significant difference between representation for single- and multi-site trials is with regards to Black participants, whose representation is worse in single-site trials than multi-site ones. Trends indicate a slight overrepresentation of females in single-site trials, in comparison to an underrepresentation of females for multi-center trials. Males are fairly represented in all trial location types, with a slight overrepresentation in multi-center trials. White populations are fairly represented across all trial location types.American Indian/Alaska Native populations are underrepresented across all trial location types, with a trend towards worse representation scores reported for single-site trials. On the other hand, Asian/Pacific Islander and Hispanic populations are also underrepresented regardless of trial location types. The average age recruited for single site trials (7.69 years younger than the real-world mean) is also slightly younger than multi-site trials (5.44 years).

(2c) Sample Size Per Trial

Using values from Appendix K, a box plot was generated to compare the diversity of sex, race/ethnicity and age based on the sample size (number of total participants) per trial:

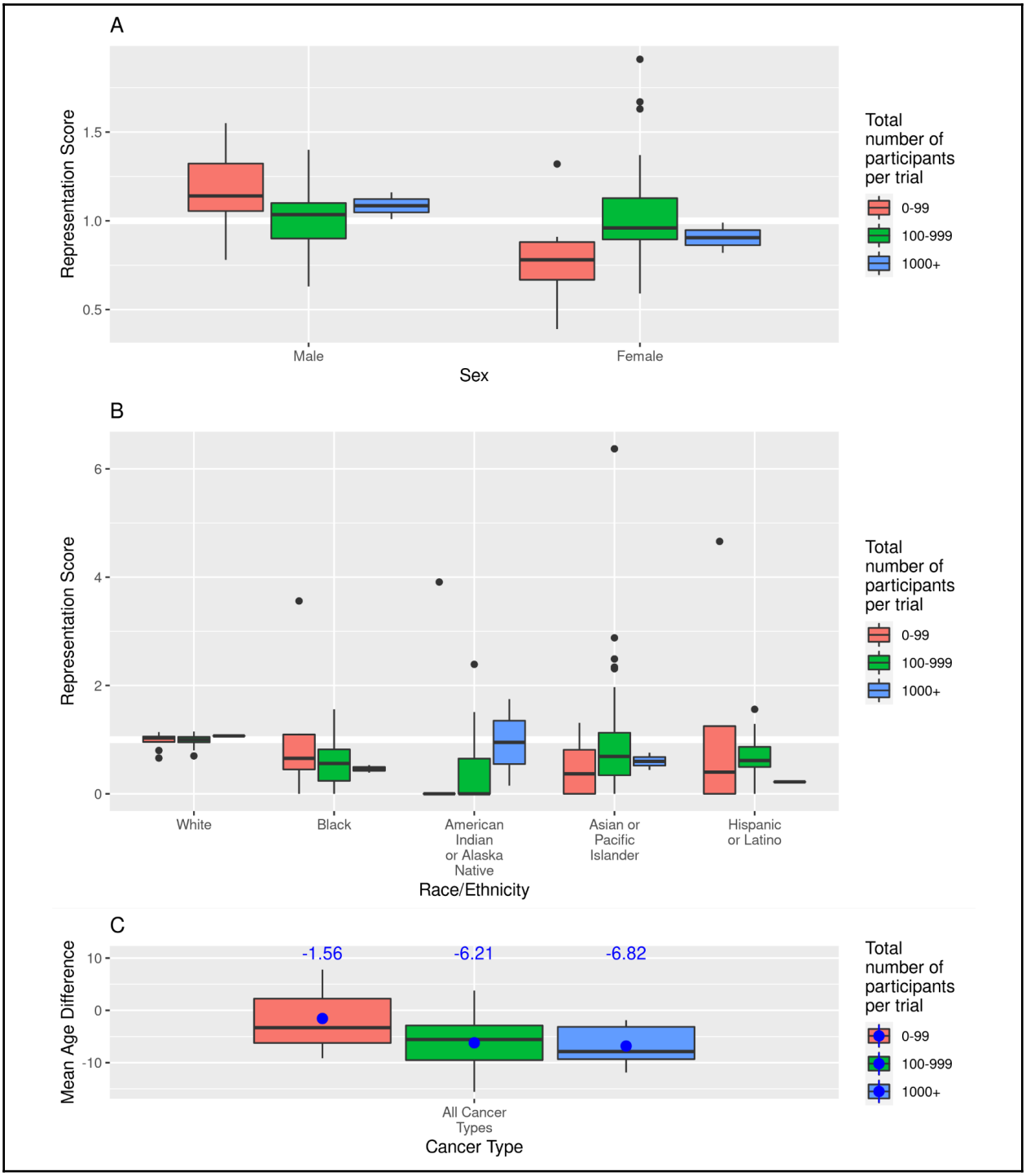

Figure 9: Trial Size (Based on Participants per Trial) Representation Stratified by Sex, Race, and Mean Age Difference

Box plots show the distribution of the representation scores for each category, plotted along the y-axis. The boxes and vertical lines show the total range of representation scores (boxes: 2nd & 3rd quartile, vertical lines: 1st & 4th quartile) and single black dots represent the outliers in a given dataset. Blue dots and numbers show the mean age difference. Differences in the distribution between trial sizes were analyzed using one-way ANOVA (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$). A: Comparison of sex demographcs representatio scoresn; B: Comparison of race/ethnicity demographcs representation scores; C: Comparison of mean age difference

Based on the results, trial size does not make a significant difference on sex, race, or age representation. Trends seem to indicate males are overrepresented in trials with less than 100 participants and over 1000+ participants, while females could be underrepresented in these trial sizes. For race data, White populations are fairly represented across all participation counts. Black populations are underrepresented in all types of trials, regardless of the number of participants, with a trend towards lower representation in trials with more participants. American Indian/Alaska Natives are also underrepresented in all types of trials, but are notably absent in trials with less than 100 participants. Trends suggest poorer representation of Asian/Pacific Islanders for trials with less than 100 participants. Hispanics and Latinos have greater variance in their representation in trials with less than 100 participants, but are consistently underrepresented overall. For mean age difference, the data suggests that trials with less than 100 participants recruit patients closer in age to the real-world prevalence data (mean age difference of 1.56 years younger than the real-world average). Larger trials over 100 participants tend to recruit comparatively younger patients (6.21 years younger for n = 100-999 trials, and 6.82 years younger for n = 1000+ trials).

(2d) Eligibility Criteria: Other Demographic Data

Lastly, after searching for keywords in the inclusion and exclusion criteria of each cancer trial, a bar chart was created to showcase the number of trials that excluded participants based on certain demographic populations. It should be noted that the purpose of this analysis was to present a qualitative overview of the kinds of additional criteria that may appear in clinical trial eligibility criteria, not to provide a discussion or assessment of whether there were defensible reasons to exclude individuals with these characteristics from particular trials.

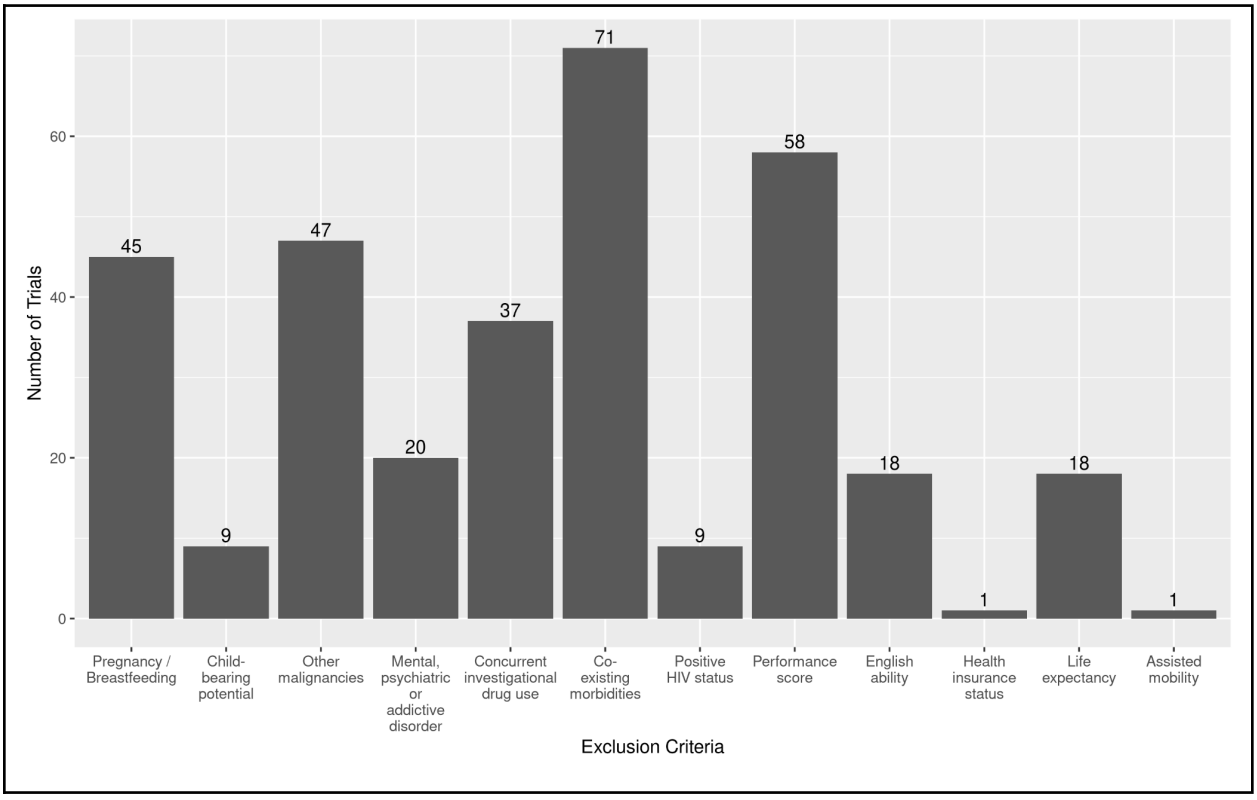

Figure 10: Proportion of Trials That Exclude Various Demographic Populations
Quantity of eligibility criteria that mention the exclusion based on pregnancy or childbearing status, presence of other malignancies, mental health conditions, investigational drug use, coexisting morbidities, HIV status, performance score (ECOG/Zubrod/Karnofsky), English ability, private health insurance status, life expectancy, and mobility.

Out of the 63 studies that included females, 45 excluded pregnant or breastfeeding participants and 9 excluded participants of child-bearing potential. When considering all 85 trials, 47 excluded participants that had other or prior malignancies, 20 excluded participants with a history of mental, psychiatric or addictive disorder, and 37 excluded participants undergoing other investigational drug treatments. The most common exclusion criteria was the presence of coexisting morbidities, which was present in 71 of the 85 trials. HIV-positive individuals were excluded from 9 of the 85 studies. Additionally, 58 out of 85 studies used some functional performance test (ECOG, Zubrod, Karnofsy) to exclude participants outside of a particular threshold. 18 of the 85 studies excluded participants who could not communicate or understand English. 1 trial excluded participants who did not have private health insurance. Finally, 8 out of 85 studies excluded participants with a life expectancy below a particular

threshold, and 1 study excluded participants requiring the use of a wheelchair or walker for mobility.

Discussion

Sex and Gender Representation

The representation of women was slightly below that of men, and in some instances comparable. This may be due to greater awareness on the inclusion of women in clinical research and sex diversity policies introduced by the FDA and NIH (Poon et al., 2013). However, most notable was the absence of non-binary, trans or 'other' categorical designations of sex/gender demographics, in either the reported demographics or the inclusion/exclusion criteria. Furthermore, many studies excluded patients who use hormonal drugs and/or birth control for other means, which can prevent transgender and intersex patients from participating in the trials (Hao et al., 2016). Although not explicitly listed in the exclusion criteria, this may include hormone replacement therapy or puberty blockers - interventions which will primarily apply to sex and gender minorities. Some studies also exclude people with HIV-positive status, which is known to disproportionately affect homosexual men and other sexual minorities (Grulich & Kaldor, 2008).

As a starting point, clinical research should work towards a fair inclusion of sex and gender minorities, and not just binary sex as it currently stands in the clinical trial data. Intersex, transgender, and non-binary populations are a smaller demographic relative to the proportion of cisgender men and women in society, but can still add value to clinical research nonetheless. Both sex and gender differences can have an impact in clinical outcome, as they differ on a biological, genetic or psychological basis that can impact individual responses to drug dosage, surgical outcome, pain experience, to name a few examples (Epstein, 2008). From a scientific standpoint, the inclusion of sex and gender minorities should be encouraged to gain knowledge about populations that have been underrepresented in medicine, while helping researchers discern sex differences apart from gender. From an ethical standpoint, sexual and gender minorities have historically been on the fringe of clinical research due to social stigma, as is the case of exclusion based on HIV-positive status which predominantly applied to homosexual male populations (Hao et al., 2016). These populations deserve the opportunity to take part in research

which can potentially benefit them, or at the very least, deserve not to be implicitly excluded from participating because of their gender or sex identity.

A potential reason for why women were underrepresented in many clinical trials can be due to the various inclusion/exclusion criteria governing the kind of women that could enroll in these trials (Macklin, 2010). Almost every single trial reported excluding pregnant or breastfeeding women, while women of childbearing potential were also often excluded. Likewise, many trials often reported only including women who have reached menopause. The exclusion of pregnant or breastfeeding women may be justified in the interests of protecting potential fetuses or infants from possible harms of an experimental drug or intervention, but excluding women of childbearing potential (which is often undefined in the trial criteria) has a less ethically-sound basis. It hints at underlying paternalistic attitudes that seek to protect unconceived children, without considering the best interests of the woman (Fletcher, 1993). By not defining the threshold for women of childbearing potential, it removes the choice for women whether they want to participate in a trial that could potentially help their lives, or protect their bodies for the sake of a potential baby. Even if women wish to avoid pregnancy during a trial, excluding women who take birth control further assumes that women are not responsible enough to use birth control consistently.

By contrast, none of the trials included or excluded men based on their fertility or childbearing potential. The only noticeable mention of male-related eligibility criteria was in some trials which asked the male partners of female clinical trial participants to use contraception to avoid the woman getting pregnant during the trial procession. Although the real-world average age of women with cancer are likely past the age of childbearing, the stated differences in the eligibility criteria of women and men show that there are still remnants of gender inequalities and policing between the two demographics, which go beyond eligibility criteria for cancer trials. At best, the strict requirements of women's eligibility criteria reflects a shortsighted way to remove potentially adverse events that clinicians want to avoid.

Lastly, recall that women are underrepresented in lung, pancreas, colorectal, kidney, lymphoma and melanoma cancer trials based on the given data. This is consistent with other diversity studies which showed female underrepresentation in lung, pancreas, colorectal, kidney,

lymphoma and melanoma cancer drug approval trials, as well as thyroid cancer trials not covered in this analysis (Duma et al., 2018; Dymanus et al., 2021; Ludmir et al., 2020; Mendis et al., 2021). Interestingly, the Mendis *et al.* study reported that women are slightly underrepresented in leukemias, while this thesis study found women are overrepresented in leukemias (Mendis et al., 2021). However, this discrepancy is likely due to the fact that the results were not statistically significant for leukemia comparison.

In addition, both this thesis study and other studies noted that women seem minimally underrepresented on an overall (all cancers) scale, but they have more dramatic differences in representation between individual cancer types (Mendis et al., 2021). A possible explanation for this disparity could be differences in diagnosis, prevalence and mortality between the sexes and cancer types (i.e. solid cancers typically tend to be under-examined in women compared to men)(Dymanus et al., 2021; Ludmir et al., 2020; Mendis et al., 2021). Another reason could be the additional eligibility barriers women are subjected to, which may vary between cancer fields. A recent meta-analysis of different cancer types showed that lack of trial availability and patient ineligibility account for 77% of women's exclusion in trials (Unger et al., 2019).

In other literature, general barriers that women face to clinical trial participation include lack of awareness, transportation difficulties and economic considerations intersecting with family responsibilities and caregiving obligations (Borno et al., 2018; Brown et al., 2000; Young, 2010). This includes gender role differences between men and women, where women as the predominant household caregivers have less time and resources to participate in trials themselves, but are able to support and transport their male partners in trials. The sex and gender differences demonstrated in this thesis thus show that despite the increased awareness and regulatory mandates from the FDA, little has changed in the past few decades, and diversity in trial enrollment still has to address critical areas for large-scale improvement (Mendis et al., 2021).

Race and Ethnicity Representation

A greater variance in representation was seen for race and ethnicity populations. This may be due to the way race and ethnicity intersect with clinical outcomes. Unlike sex and gender

profiling, race/ethnicity profiling (or the categorization of a population based on race and ethnicity identities) has no consensus in its use. Researchers may be less inclined to ensure an equitable representation of different races and ethnicities if it has no actionable value to clinical data (Epstein, 2008). Nonetheless, it is still an important value to uphold as a lack of fair representation in clinical trials translates to a lack of awareness and access for the same therapies in the real-world (Baquet et al., 2006). The inclusion of race and ethnicity data thus serves mainly to rectify historical injustices, such as ensuring that historically underrepresented populations like Black and American Indian/Alaska Native populations now are fairly represented proportional to their prevalence in society. Race and ethnicity data also hold predictive value in revealing underlying social determinants of health outcome (Bierer et al., 2021).

Another reason for variability in race and ethnicity data can be that they are difficult demographics to categorize and accurately represent (Yanow, 2015). Although it is often debated whether race and ethnicity are socially or biologically constructed, both have predictive value to health outcome as a result of the intersection between biological and social factors (Benjamin, 2019; Yanow, 2015). Furthermore, the effect of biological and social factors are not mutually exclusive, as demonstrated by the separate reporting of Hispanics vs non-Hispanics ethnicities and further separated from the other racial groups. For example, a mixed-race White and Asian adopted person raised in a Spanish-speaking household could identify simultaneously as White, Hispanic and Asian. These ambiguities between ‘actual’ versus self-reported race/ethnicities complicate the task of determining whether particular groups were indeed misrepresented. Nonetheless, it is worthwhile to discuss some trends observed in the race/ethnicity data in the hopes of improving equitable representation in the future.

In this study, White populations were fairly represented across all trials, regardless of the type of funding, the location type, or the trial size. This demographic also had much less variability and much more statistical certainty due to their high sample size, compared to other demographics of races which often report none or less than 10 participants per trial. The lack of variance across all trials represents an ideal distribution that clinicians should strive to ensure for other races and ethnicities.

On the other hand, Black populations were consistently underrepresented in all trials, with representation becoming significantly worse for single-center trials and as trials increased in the number of participants per trial. A possible reason for multi-center trials having better representation is that certain trial locations might be located in cities with a high Black population, or there were less enrollment barriers at that location (Rivers et al., 2013). On the other hand, the poor representation seen in single-center trials indicates that the U.S. still needs to improve their recruitment of Black populations. There are also systemic inequities that can implicitly impact clinical trial recruitment, such as lack of good-quality healthcare access leading to later age of cancer diagnosis, the presence of more coexisting morbidities leading to trial ineligibility, or Black populations being predominantly served by community hospitals that do not conduct trials.

Similarly, American Indian/Alaska Natives were also consistently underrepresented in all trials, with representation becoming worse for single-center trials. A possible explanation for this is that single center locations (which are only located in the U.S.) reflect long-standing issues in U.S. outreach and health support to American Indian/Alaska Natives communities (Yuan et al., 2014). The representation of this demographic also improved as individual trials increased in the number of participants. This further reflects that there may be issues in resource allocation for outreach and recruitment initiatives to Native communities, as most cancer trials are conducted in city centers (Đoàn et al., 2019).

A common belief is that Asian/Pacific Islanders are not underrepresented or even over-recruited in clinical research, which separates them from other visible minorities that face systematic discrimination or marginalization in healthcare (Fashoyin-Aje et al., 2017). From the analysis, there was mixed representation of Asian/Pacific Islanders compared to Black populations for example, who were consistently underrepresented in all studies. Interestingly, for studies which reported an overrepresentation of Asian/Pacific Islanders, the p-values were not calculated, meaning there were not enough trials in that cancer category to make a generalization of representation data that would be statistically significant. On the other hand, Asian/Pacific Islanders were consistently underrepresented in all cases where race data was statistically significant (i.e. reported a p-value of less than 0.05). This means we can discern with a high

degree of certainty that Asian/Pacific Islanders are underrepresented in many trials, but we cannot confirm with the same degree of certainty if they are overrepresented in other trials.

The secondary data analyses can shed light on the possible reasons for this disparity. Multi-center locations have greater variance in Asian/Pacific Islanders representation than single-center trials, although both show an overall trend of underrepresentation. A possible explanation for this is that single-center trials tend to be located in urban centers, and the majority of the U.S. Asian/Pacific Islander population (like most non-White racial populations) also tends to live in urban centers (Cromartie, 2018; Population Reference Bureau, 2004). The poor representation of Asian/Pacific Islanders in small trial sizes supports other literature findings that clinical trials often recruit 0 or very few Asian/Pacific Islanders, which is more likely to happen when the sample size per trial is low (Barba, 2020; Nguyen et al., 2021). In fact, because the distribution of the Asian/Pacific Islanders matches closely with typical trial locations, one might expect the representation of Asian/Pacific Islanders in clinical trials to exceed the real-world average.

It was expected that the Hispanic or Latino demographic would be underrepresented, however, the analysis has revealed high variability and a lack of statistical significance for most trial types. A possible reason for this variance could be due to its demographic being inconsistently or ambiguously reported as either a racial or ethnic category in each trial (Fashoyin-Aje et al., 2018). This is reflective of real-world ambiguities, where some may consider Hispanic ethnic populations to be a White or ‘other’ racial category. That trial data may not be statistically significant should not deter from the general observed trend of this demographic being underrepresented. In fact, the results can be seen as an indicator that existing social structures and systemic barriers can influence the type of people enrolled and thus impact the clinical outcomes or the value of who clinical knowledge applies to (Epstein, 2008). These results are an indicator that more work should be done to accurately report and develop uniform guidelines for the recording of Hispanic/Latino participation, or to examine the underlying factors in their underrepresentation. For example, although the FDA strongly recommends collecting Hispanic/Non-Hispanic data according to the Office of Management and Budget

(OMB) standards and guidelines, it is undisclosed in the demographic data if these guidelines were followed (U.S. Department of Health and Human Services et al., 2016).

Lastly, while the ‘other’ race/ethnicity category did not undergo statistical analysis because it did not have a real-world incidence counterpart to compare it to, it does not mean that the category itself has no value. The “other” category indicates that the existing race/ethnicity delineations are far from comprehensive, and that improvements can be made to accurately report these subpopulations (Murthy et al., 2004). Most importantly, the value of racial and ethnic demographic data might reveal the causal influence of other social factors on participant recruitment, such as socioeconomic status, education level, insurance status, and overall trust in the medical system. We know that these social factors differ between different races, due to the systemic inequities in society (Gracia, 2020; O’Connell, 2012). For example, Hispanis/Latinos, Black, and American Indian/Alaska Natives tend to live in poverty in comparison to the rest of the U.S. population (Mogull, 2011) and Black populations tend to have a distrust in the medical system, thereby lowering the likelihood of these populations from participating in clinical trials (Sullivan, 2020). However, there is rarely any report of these social factors in clinical trials, with only one trial reporting insurance status and no other trials reporting the socioeconomic status or education level of participants. This is why recording and analyzing race/ethnicity data in clinical trials is only useful insofar as these other social factors are studied alongside it.

Age Representation

For all cancers types, it was evident that cancer clinical trials tend to recruit younger participants, with a global mean age of 5.9 years and median age of 6.72 years below the real-world average age of diagnosis. Likely, the way that eligibility criteria are designed can explain this phenomenon. Most clinical trials reported age inclusion to be “All ages" or “18 and above”, but many trials also reported cut-offs at “60 years and under”, or at 70, 75, 80, or 80 years and under, thereby limiting the top age of participants. In many of these trials, the reason for such exclusion was unexplained, which is confounding since 60% of incident cancer cases and 70% of deaths occur in people over age 65 in the U.S. (Mone & Mehl, 2020). Another contributing factor may be due to other specificities in the inclusion and exclusion criteria, which

often include a long list of clinical characteristics, and necessitate patients to have a life expectancy of over 10-15 years. These clinical characteristics are often very narrowly defined in terms of acceptable clinical biomarker ranges, and patients that meet all the eligibility requirements would thus be more along the 'healthier' physiological state of cancer progression, and therefore, more likely to be younger.

There were two cancer types with mean cancer ages that significantly diverged from the global mean of (-)5.9 years: prostate cancer trials recruited only 0.92 years younger than the real-world mean age, while melanoma trials recruited 15.59 years younger. A possible explanation for melanoma trials recruiting much younger patients is that being of younger age is a primary determinant for preventing melanoma treatment resistance (Mone & Mehl, 2020). As patients become older, there is a higher chance of genetic mutations changing their biological profile to develop drug resistance compared to other cancers. Clinical trial technicians would thus likely want to recruit the youngest patients to ensure the best outcomes for their intervention (Aapro et al., 2005). For prostate cancer trials, a possible reason for its closer match to the real-world age is that prostate cancer only typically develops for people 60 years and older, with incidences of prostate cancer below 60 years old being very rare (Li et al., 2012). Thus, if there is a lack of availability in the pool of younger participants with prostate cancer, then naturally the average clinical age would match the real-world age better.

Another interesting finding was in the secondary data analysis, where the mean age was significantly lower for NIH-grouped trials (15.58 years younger) compared to Industry and Other trials (4.21 and 6.49 years younger, respectively). In similar fashion, trials with less than 100 participants had much higher mean age (1.56 years younger) than trials with 100-999 or 999+ participants (6.21 and 6.82 years younger, respectively). A possible explanation for NIH trials having a lower mean age was that there were only 3 of these trials out of the 85 total trials studied, whereas there were 19 Industry trials and 63 Other trials, thus compromising the statistical significance for NIH trials. The same cannot be said for the analysis by trial sample size, as there were enough trials in each category: 12 total trials with 0-99 participants, 63 trials with 100-999 participants, and 10 trials with 999+ participants. A potential explanation for why small trial sizes recruit participants closer to the real-world mean age compared to medium and

large trials could be that smaller trials are often single-centered. This means they have more time and resources to review individual patient records and give more thought to the recruitment and trial process such as implementing adaptive design, thereby being able to match closer to the real-world distribution of cancer incidence (Coffey, 2010; Hackshaw, 2008).

A comparison of this study's age analysis results to other studies reveal a similar age discrepancy. A U.S. study of 261 FDA-approved oncology drug trials from July 2007 to June 2019 revealed that these trials recruited participants who were 5.1 years younger than the real-world average (95% CI −6.4 to −3.7 years, p-value = 0.04) (Jayakrishnan et al., 2021). The Jayakrishnan study also tracked the mean age difference by cancer group and showed similar age discrepancies between cancer types in our study: melanoma trials reported a mean average age of 9.5 years below the real-world average ($p<0.005$) while prostate cancer trials reported a mean age 4.6 years above the real-world average ($p<0.005$). For comparison, recall that our study indicated the mean age to be -5.9 years for all trials, -15.59 years for melanoma trials, and -0.92 years for prostate cancer trials. These findings support other studies which reveal participants above 65 years old have been consistently underrecruited in cancer clinical trials over the past decade (Gopishetty et al., 2020; Murthy et al., 2004). It is important to note that the FDA has released a guidance document outlining practical suggestions to mitigating systemic barriers of age bias in clinical trials (U.S. Department of Health and Human Services et al., 2022). Notably, this document is addressed to many stakeholders, including sponsors and institutional review boards, all who are responsible for establishing fair representation in cancer clinical trials.

Inclusion and Representation of Other Demographic Categories

Although other demographic data beyond sex, race and age were not explicitly recorded in the primary characteristics section of clinical trial data, the eligibility criteria can be used to infer the kinds of social demographics that are excluded from clinical trial participation. In the interests of eliminating confounding variables and prioritizing the healthiest participants, it is typical for clinical trials to exclude participants who have coexisting morbidities not related to the trial's interest, or include participants with the best prospective cancer outcome (Boyd et al., 2012). For example, it was mentioned that almost all trials required participants to have a life

expectancy of 10 years or over. As the average Phase III clinical trial length is 15 months, it is unlikely that the life expectancy was set to ensure that participants do not die before the trial's end. Instead, this requirement could symbolize a general metric of (relative) good health, which in turn may implicitly favor younger or early-stage cancer patients (Townsley et al., 2005).

The exclusion criteria of psychiatric/addictive disorders or mental conditions can be a form of mental health disability discrimination. The most commonly cited reason for this criteria was to ensure that eligible participants could provide informed consent. However, not all trials provide this justification for this exclusion (some trials do not provide any justification at all), while the definition of 'mental condition' is very loosely defined. As such, there is no standardized threshold for establishing cognitive capacity, meaning that the thresholds for what is considered a mental condition can vary by each trial. Although unlikely, it is possible that people who have mild or situational depression from having cancer could be excluded from the trial (Humphreys, 2013).

The exclusion of participants based on criteria like performance scores, co-existing conditions, malignancies or concurrent/investigational drug use can be considered a form of physical disability or polypharmacy discrimination. The exclusion of polypharmacy is typically justified to avoid potentially confounding drug interactions. However, in the context of late-stage clinical trials where effectiveness is a key goal in clinical knowledge, having this criteria can inadvertently ignore the real-life biological or social complexities of participants, or prevent common drug interactions from being detected in monitored settings. Comorbidity and polypharmacy restrictions also imply that cancer patients with chronic pain requiring drugs to manage, advanced cancer patients that have already taken other drugs, or older patients with general cognitive or mobility issues may be ineligible for the study (Ridda et al., 2008). Often, these conditions are vaguely defined and listed in the "Other" section of the eligibility criteria, as opposed to the main eligibility section. This makes the exact clinical exclusion parameters harder to discern, and harder to assess if they were reinforced with the same rigor, especially for multi-centre trials in different countries led by different clinicians.

Lastly, there are some trial requirements which seek to exclude participants solely based on social limitations, such as being unable to read or speak English, or not having private health

insurance. These restrictions limit access to clinical trials from populations such as immigrants/foreigners who may not speak English, or low income individuals who cannot afford private health insurance. Most surprising was that some trials excluded those who were dependent on wheelchair or walkers for mobility, possibly due to the accessibility limitations of the clinical trial center, but mobility should not directly affect the clinical outcomes. This would be a clear instance of disability discrimination, but it is also a form of age discrimination, as older participants tend to require mobility assistance devices. Although social factors like English literacy, comorbidities/disability status, and socioeconomic status were not reported in every trial, it would be important to track this demographic data as a first step to see if clinical trial participants are underrepresented based on these fronts.

Limitations of Study

The real-world incidence data taken from (U.S. Cancer Statistics Working Group, 2021) was based on a 5-year limited duration, collected on January 1, 2018. This means that the real-world incidence data may not be an accurate representation of the exact demography for 2017 onwards. As there lacked more recent alternatives, this was the best database available for reporting modern real-world cancer incidence. Another limitation is that the race and ethnicity demographic data was combined as one demographic category in the U.S. real-world cancer incidence database, whereas race and ethnicity are treated as two separate demographic categories. As the proportion comparisons of race/ethnicity thus did not use two mutually exclusive populations between the trial and real-world, the calculated representation scores may not be entirely accurate for Hispanic populations. Furthermore, some multi-centre trials had trial locations outside the U.S., meaning that the sample population did not compose of entirely U.S. proportions. However, the representation score was calculated based on real-world U.S. demographic proportions. Despite this discrepancy, the total number of non-U.S. trial locations was negligible compared to U.S. locations, so the population of non-US participants can be considered insignificant. Finally, although the FDA has produced detailed guidelines on reporting demographic information in clinical trials, individual trials may not necessarily uphold these standards (U.S. Department of Health and Human Services et al., 2016). In addition, the

database used in this study (clinicaltrials.org) may not report detailed demographic information up to the FDA standard, thereby limiting the depth of analysis. For example, many trials were excluded in the age analysis because age data was inconsistently reported across trials, thereby compromising the accuracy and precision of age analysis. Age was variably represented as either numerical age ranges, mean age with standard deviation, median age (with or without standard deviation), or not reported in enough detail for analysis.

Future Directions

Due to time constraints, a more detailed analysis of demographic data could not be performed. Ideas for future studies to improve the quality of demographic analysis can be to:

(1) Stratify sex, race and age data by study design, intervention type, drug type, and randomization style to see if these factors have a correlation to participant representation score and mean age difference. For example, a study could be done to assess if the mean age difference differs or improves for first, second, and third-line drugs (usually second line and higher drugs and are typically used in older cancer patients).

(2) Identify poor-quality or risky studies and analyze if the resulting underrepresented populations (women, Black, American Indian/Alaska Natives, etc.) are disproportionately underrepresented or overrepresented in these studies. Examples of these studies include: experimental drugs used as the sole intervention without a conventional treatment arm; inadequate training of clinical personnel; low funding studies; discontinued studies, etc., as defined by the Good Clinical Practices (GCP) Handbook (Food and Drug Administration, 2010).

(3) Perform a more detailed and precise analysis of trial location (such as if the trial was conducted in city centers or rural areas) to see if it can affect the demographic composition of particular races, sexes or ages. For example, as we know the majority of Asian Americans/Pacific Islanders tend to live in urban city centers, an in-depth comparison of precise trial location can provide a better representation score for these populations.

(4) Conduct a qualitative study of the enrollment and participation process, to see if any systemic barriers were in place to deter underrepresented populations from enrolling. This could mean

conducting interviews or collecting survey feedback from trial participants and the general public to identify personal barriers they may face in clinical trial participation.

Recommendations

The following recommendations are provided to improve diversity in clinical trials.

(1) Explicitly recruit and track the participation of sex and gender minorities, such as non-binary, transgender and intersex individuals

Sex and gender minorities constitute a systemically underrepresented demographic in medicine. Most clinical trial demographic data does not include intersex or transgender designations or report "other" as a sex demographic category. In addition, the current exclusion criteria of barring hormonal drugs implicitly excludes many individuals of the sex and gender minority community who use these drugs on a daily basis. There is also an absence of literature investigating the impact of these exclusion criteria on the recruitment and participation of sex/gender minorities. Recall that clinical research about sex and gender minorities is valuable for the sake of generating unbiased knowledge for populations that have been historically ignored or discriminated against in medicine. If clinical research is justified based on sex differences (whether physiological, genetic or otherwise), then research about sex minorities like intersex is needed to obtain a holistic understanding of sex-based effects. This can be used to inform comparisons of sex differences in sex minorities but also between binary sex (male/female) as well. Studying a diversity of sexes in medical research can help us gain a better understanding of sex-mediated regulation and developmental changes in gene expression, epigenetics and human physiology.

To improve diversity, clinicians should reassess eligibility criteria for conditions that may inadvertently exclude sex and gender minorities. When a criteria comes in conflict with the inclusion of sex/gender minorities, an evidence-based justification should be provided as to why it is necessary to the overall research objective. Clinicians should also make explicit efforts to notify participants in LGBTQIA+ communities, through social media networking and community-targeted recruitment strategies, and accurately report their participation apart from the 'Other' category (Frohard-Dourlent et al., 2017). Suggestions for diverse sex and gender

categories could be “intersex”, “transgender man”, “transgender woman”, or “non-binary” designations (Frohard-Dourlent et al., 2017). Finally, for trials that measure effectiveness or stratify differences in health outcomes, clinicians should report and analyze gender effects apart from sex. It is noted that the number of sex/gender minorities may be too small to analyze in individual trials, but the data can be compiled for aggregate demographic analysis similar to the methods employed in this thesis. Furthermore, within the greater medical community, sex and gender are often conflated or misunderstood as separate but interdependent concepts (Phillips, 2008). Fortunately, there are initiatives to educate researchers about sex and gender nuances, such as the Canadian Institute of Health Research (CIHR) training modules and workships for sex/gender-based analysis (Canadian Institutes of Health Research, 2015). Therefore, in addition to creating awareness on sex/gender nuances, regulatory bodies should revisit training programs and practical guidelines on the report and analysis of sex and gender data in clinical trials.

(2) Reconsider the exclusion criteria definition of ‘women who are of childbearing potential’ and increase the accessibility of trials for women

As mentioned, the exclusion criteria of ‘women of childbearing potential” is ill-defined and implicitly results in the exclusion of most healthy and younger women, including potentially women who take birth control. There needs to be a critical re-assessment of the reasons for this criteria, including rethinking biological or ethical justifications that are not based on blanket paternalistic ideals for the sake of protecting a woman’s reproductive ability. When it is justified, regulatory bodies or clinical trials should define the exact parameters of this criteria such that it can be uniformly applied. For example, ‘women of childbearing potential’ may be better described as women who are actively trying to conceive, with an age range (i.e. under 40), but exclude those on birth control and those who do not wish to become pregnant. In addition, evidence drawing from sex-based biology research should be used to identify where an exclusion of women based on childbearing potential is indeed needed. Clinical trial investigators may also benefit from equity training on sex and gender differences in the healthcare needs of clinical trial participants (Agency for Healthcare Research and Quality, 2010; Wizemann & Pardue, 2001). This can facilitate awareness, recognition, and reduction of health disparities that women face to

better integrate the interests of women in the clinical trial process. This will ensure a greater focus of women's issues in its own right during the clinical knowledge-making process, rather than perpetuate the instrumentalization of women's bodies for childbearing. (Agency for Healthcare Research and Quality, 2010; Wizemann & Pardue, 2001).

Clinicians should also work to mitigate barriers in the enrollment and retention of women in trials, imposed through trial design or trial expectations that can be exceptionally burdensome for women. For example, transportation issues may disproportionately limit the participation of women who travel alone, especially in rural areas due to lack of safe transportation methods or safety concerns (K. A. Liu & Mager, 2016). The logistics of trial participation and frequent follow-ups may interfere with women's work and domestic obligations, and thus clinicians should execute trials with these schedule constraints in mind (Health (U.S.) et al., 2002). Investigators could implement flexible trial and follow up hours, or hold trial locations in personal homes or social places that women frequently visit (i.e. salons, gyms, stores, laundromats, and churches) (K. A. Liu & Mager, 2016). Financial burdens in transportation, parking, or babysitting costs may be burdensome for women of low-income status (K. A. Liu & Mager, 2016), and sponsors should consider setting a budget aside to absorb these costs. Finally, there may be cultural constraints that prevent racial, religious and ethnic minority women from participating, and trial investigators should be respectful of these needs, such as employing an all-female staff during interactions with women (Wilcox et al., 2001). Feedback should also be gathered to inform and improve future efforts of increasing diversity for women.

(3) Increase the diversity of racial and ethnic categories provided to participants, and mitigate systemic barriers linked to race and ethnicity

One of the major barriers discussed in the demographic analysis is the lack of accurate reporting of race and ethnicity data, especially for Hispanic populations. This also applies to Asian/Pacific Islander or Native populations, which can comprise a wide diversity of races and ethnicities that are often generalized into one group. Having broad and overgeneralized racial/ethnic categories seems counterintuitive to the value that recording race-based data has to clinical knowledge, which is to associate health outcomes to specific races or ethnicities for the

sake of improving medical knowledge for that community. Recall that Spencer believed there is a useful classification scheme for race in medical research, namely, by an association of race to medically relevant genetic differences (Q. N. J. Spencer, 2018). For example, risk factors for cancer can be established from distinct genetic polymorphisms which are associated with particular genetic ancestry groups, like the case of breast cancer risk for those with European, African, and Latinx Ancestry (C. Liu et al., 2021). It follows that for the sake of precision and predictive power, race and ethnicity data should have a diversity of categorical options than the current 4 broad socio-continental divisions. In addition, participants should be allowed to declare mixed race or racially/ethnically ambiguous designations to be inclusive of all racial and ethnic subtleties (Q. Spencer, 2014; Q. N. J. Spencer, 2018).

To improve the diversity of underrepresented racial populations, efforts should be made to address any underlying concerns held by these communities. Commonly cited concerns are mistrust or stigma of clinical trials and investigators, lack of information of awareness of the benefits of clinical research, and lack of comfort in clinical trial procedures (Clark et al., 2019). These issues can be mitigated by improving community outreach using culturally-sensitive approaches to recruit participants, for example, emphasizing the benefits of furthering general medical knowledge in Asian communities that value duty to the community (Saraswat et al., 2020). Structural inequalities such as differences in funding between hospitals that serve racial minorities versus White or wealthy populations should also be addressed. As mentioned, as late-stage clinical trials are one of the few ways to give beneficial technologies to communities that would otherwise be inaccessible to them, efforts should be made to prioritize racial minorities and Indigenous communities living in remote locations in these trials. This could mean holding trial locations within the lands of the community itself, or providing transportation and housing costs for these populations (Clark et al., 2019). Another option is to hire language interpreters or conduct trials in the native language where language is a major barrier to participation. Finally, recruiting clinical trial staff reflective of the target minority demographic can improve trust and communication in interactions with participants. The suggestions provided in this paragraph aims to improve the overall trust, transparency and communication between participants, clinicians/investigators and the greater community (Clark et al., 2019).

(4) Decrease age-specific barriers in clinical eligibility and improve age-centric participation, reporting & analysis

Two main barriers to age diversity revealed in this analysis are having strict eligibility criteria and inconsistencies in reporting of age data between trials. Strict eligibility criteria, especially in having long life expectancy ranges, implicitly deter the eligibility of older participants based on biological fitness. As life expectancy is also correlated to cancer progression, long life expectancies can exclude late-stage cancer patients from participating (Aapro et al., 2005; Watts, 2012). It can be argued that life-extending therapies in clinical trials would more greatly benefit the younger populations rather than older populations, while also producing the best therapeutic responses, thus younger participants should be prioritized. However, the ultimate goal of these trials is to ensure effectiveness of the therapy in the real world. As it stands, the average age of real-world cancer patients currently coincide or surpass the top age ranges of the tria's inclusion criteria. Thus, the target population of these therapies is currently being excluded and thus the trial is not a reasonable embodiment of how it would perform in the real-world. Thus, it is recommended that any eligibility criteria with a life expectancy criteria should be justified with sound evidence for the exact threshold chosen, or removed altogether.

For example, as it is known that melanoma progression is correlated to age as an approximate metric of biological fitness. A trial which explores the efficacy of a novel hormonal drug could be potentially justified to restrict certain ages if the researchers explain the correlation between the drug mechanism and biological factors such as differences in drug metabolic activity between pre-menopausal (younger) and post-menopause women (older). Investigators should also reevaluate the use of fitness and performance scores to see if they may exclude older participants from qualifying. Lastly, efforts can be made to improve mobility access at trial centers. A financial allocation could be set aside to address mobility barriers such as ensuring wheelchair access, using transportation shuttles between homes, employing dedicated on-site nurses, or purchasing mobility aids to help with movement. Providing mobility alternatives will

allow investigators to eliminate mobility and transportation constraints in exclusion criteria which disproportionately apply to older populations.

Furthermore, efforts can be made to improve the way age is recorded and analyzed (Lüscher et al., 2020). The recording of age data should be standardized to include both individual count and calculated means, medians and standard deviations for every trial. Standardized benchmarks for age inclusion should also be developed by regulatory bodies to ensure a fair and diverse recruitment of age ranges. For example, investigators could ensure matching of the mean age of participants to prevalence and incidence age data of the same conditions in the real population. Another suggestion could be to ensure participants are uniformly recruited between age brackets (20-29, 30-39, etc), with subgroup analyses run for each age range. As age correlates with differences in biological function, psycho-social behavior, and health outcome, efforts should be made to integrate age data as a determinant of health in any analysis of clinical objectives. Making explicit efforts to integrate age data into the analysis can reveal the subtle ways that age may impact the resulting clinical findings.

(5) Improve the quality, extent and diversity of demographic data collected

Overall, the demographic analysis was limited by the quality, type and range of subpopulations within demographic categories. Typically, only three demographic categories were recorded: race, sex and age, with variable standards of detail between trials. Thus, regulatory bodies should develop uniform standards of reporting demographic data in sufficient detail such that statistically significant analysis can be performed. Efforts should also be made to track additional socioeconomic benchmarks, such as income level, highest attained education, geographical location (urban/rural), or private health insurance status in demographic data. The examples provided have correlations to real-world differences in disease diagnosis, prevalence and prognosis, and thus, clinical research should report these categories for the sake of furthering population health and health equity research, even if they do not conduct the analysis themselves (Lee, 2004). Finally, it is important to increase the diversity of subpopulations within demographic categories, as it can help improve diversity efforts by identifying target minority groups that may have been overlooked by traditional categorizations.

In addition, increasing the diversity of demographic categories can shed light on any additive influences of these categories on health outcomes in epidemiological data analysis. For example, single variate stratification may not reveal strong correlations of health outcome to race alone, but multivariate regression analysis of health data according to race and household income level can shed light on the correlational effect between these factors. Especially in cases where racial and ethnic health disparities are often correlated to other social determinants of health, recording and analyzing race/ethnicity data in clinical trials becomes more useful when these other social factors are studied alongside it. Other suggestions can be to conduct parallel studies that are adequately powered to study social inequalities, including making sound inferences about the causes of inequalities. This may seem more likely to produce actionable knowledge on how to address inequalities, although these trials may be financially and logically more burdensome than ensuring and analyzing for diversity in the primary trials.

Overall, the above recommendations can be summarized as increasing awareness of diversity barriers, widening eligibility criteria, and removing potential barriers to clinical trial access for underrepresented populations. It is noted that the suggestions provided are relevant to all types of trials beyond the context of cancer trials covered in this thesis. However, as the aims, study design and implementation of trials can vary greatly between fields of study and phases of trials, it is important for clinicians to recognize the barriers to diversity are contextual. Establishing equity in clinical trial participation does not simply entail including one participant of each demographic category. Critical reflection is needed for investigators to eliminate systemic barriers in clinical trial diversity.

Conclusion

This thesis has explored a theoretical and analytical framework for diversity in modern clinical trials. Ethical reasons for diversity were to rectify social inequities and ensure justice in risks, benefits, and clinical knowledge for historically exploited populations. Epistemic reasons for diversity were presented as increasing rigor and precision in detecting clinical differences between subpopulations. A cross-sectional study was conducted to provide a comprehensive analysis of demographic data for U.S. Phase III cancer trials from 2017-2021. The results

showed that Black and Indigenous demographic populations, women, and older patients are still underrepresented in most cancer trials. Furthermore, secondary data analysis revealed that external trial factors have a correlational effect on clinical trial participation, such as overrepresentation of females in single-site trials versus an underrepresentation in multi-center trials. Further research is needed to determine the root cause of these correlations. There remain systemic barriers preventing the inclusion of certain demographic populations, such as the exclusion of women of childbearing potential, non-English speakers, and those with addictive or psychiatric disorders. The detailed demographic analysis presented in this thesis, alongside practical recommendations for increasing diversity, was presented with the aim to contribute to existing critical reflections of modern day diversity in cancer clinical trials.

References

Aapro, M. S., Köhne, C.-H., Cohen, H. J., & Extermann, M. (2005). Never Too Old? Age Should Not Be a Barrier to Enrollment in Cancer Clinical Trials. *The Oncologist*, *10*(3), 198–204. https://doi.org/10.1634/theoncologist.10-3-198

Agency for Healthcare Research and Quality,. (2010). *Healthcare Quality and Disparities in Women: Selected Findings From the 2010 National Healthcare Quality and Disparities Reports.* (No. 11-0005-1-EF). https://archive-ahrq-gov.proxy3.library.mcgill.ca/research/findings/nhqrdr/nhqrdr10/women.pdf

Alexander, G. A., Chu, K. C., & Ho, R. C. S. (2000). Representation of Asian Americans in Clinical Cancer Trials. *Annals of Epidemiology*, *10*(8), S61–S67. https://doi.org/10.1016/S1047-2797(00)00198-8

Arpinelli, F., & Bamfi, F. (2006). The FDA guidance for industry on PROs: The point of view of a pharmaceutical company. *Health and Quality of Life Outcomes*, *4*(1), 85. https://doi.org/10.1186/1477-7525-4-85

Artal, R., & Rubenfeld, S. (2017). Ethical issues in research. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *43*, 107–114. https://doi.org/10.1016/j.bpobgyn.2016.12.006

Baquet, C. R., Commiskey, P., Daniel Mullins, C., & Mishra, S. I. (2006). Recruitment and participation in clinical trials: Socio-demographic, rural/urban, and health care access predictors. *Cancer Detection and Prevention*, *30*(1), 24–33. https://doi.org/10.1016/j.cdp.2005.12.001

Barba, V. (2020, February 26). *Female participation up, diversity rates down in 2019 approved*

drugs' trials. Outsourcing-Pharma.Com. https://www.outsourcing-pharma.com/Article/2020/02/26/FDA-release-of-clinical-trials-participation-demographics

Barcelona de Mendoza, V., Huang, Y., Crusto, C. A., Sun, Y. V., & Taylor, J. Y. (2018). Perceived Racial Discrimination and DNA Methylation Among African American Women in the InterGEN Study. *Biological Research for Nursing*, *20*(2), 145–152. https://doi.org/10.1177/1099800417748759

Bartlett, C., Doyal, L., Ebrahim, S., Davey, P., Bachmann, M., Egger, M., & Dieppe, P. (2005). The causes and effects of socio-demographic exclusions from clinical trials. *Health Technology Assessment*, *9*(38). https://doi.org/10.3310/hta9380

Basu, R., Dalla Man, C., Campioni, M., Basu, A., Klee, G., Toffolo, G., Cobelli, C., & Rizza, R. A. (2006). Effects of age and sex on postprandial glucose metabolism: Differences in glucose turnover, insulin secretion, insulin action, and hepatic insulin extraction. *Diabetes*, *55*(7), 2001–2014. https://doi.org/10.2337/db05-1692

Begun, C. (2021, November 18). Mission Possible: Improving Diversity in Cancer Clinical Trials. *PennMedicine News*. https://www.pennmedicine.org/news/news-blog/2021/november/mission-possible-improving-diversity-in-cancer-clinical-trials

Benjamin, R. (2019). Assessing risk, automating racism. *Science*, *366*(6464), 421–422. https://doi.org/10.1126/science.aaz3873

Bierer, B. E., White, S. A., Meloney, L. G., Ahmed, H. R., Strauss, D. H., & Clark, L. T. (2021). *Figure 44: Genotyping for CYP2C19 and clopidogrel from Achieving Diversity,*

Inclusion, and Equity in Clinical Research. Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard (MRCT Center). https://mrctcenter.org/diversity-in-clinical-research/wp-content/uploads/sites/11/2021/09/MRCT-Center-Diversity-Guidance-Document-Version-1.2.pdf

Biliya, S., & Bulla, L. A. (2010). Genomic imprinting: The influence of differential methylation in the two sexes. *Experimental Biology and Medicine (Maywood, N.J.)*, *235*(2), 139–147. https://doi.org/10.1258/ebm.2009.009251

Borno, H. T., Zhang, L., Siegel, A., Chang, E., & Ryan, C. J. (2018). At What Cost to Clinical Trial Enrollment? A Retrospective Study of Patient Travel Burden in Cancer Clinical Trials. *The Oncologist*, *23*(10), 1242–1249. https://doi.org/10.1634/theoncologist.2017-0628

Boyd, C. M., Vollenweider, D., & Puhan, M. A. (2012). Informing Evidence-Based Decision-Making for Patients with Comorbidity: Availability of Necessary Information in Clinical Trials for Chronic Diseases. *PLoS ONE*, *7*(8), e41601. https://doi.org/10.1371/journal.pone.0041601

Braun, L. (2017). Theorizing Race and Racism: Preliminary Reflections on the Medical Curriculum. *American Journal of Law & Medicine*, *43*(2–3), 239–256. https://doi.org/10.1177/0098858817723662

Breton, C. V., Landon, R., Kahn, L. G., Enlow, M. B., Peterson, A. K., Bastain, T., Braun, J., Comstock, S. S., Duarte, C. S., Hipwell, A., Ji, H., LaSalle, J. M., Miller, R. L., Musci, R., Posner, J., Schmidt, R., Suglia, S. F., Tung, I., Weisenberger, D., … Fry, R. (2021). Exploring the evidence for epigenetic regulation of environmental influences on child

health across generations. *Communications Biology*, *4*(1), 769. https://doi.org/10.1038/s42003-021-02316-6

Brody, H., & Hunt, L. M. (2006). BiDil: Assessing a race-based pharmaceutical. *Annals of Family Medicine*, *4*(6), 556–560. https://doi.org/10.1370/afm.582

Brown, D. R., Fouad, M. N., Basen-Engquist, K., & Tortolero-Luna, G. (2000). Recruitment and retention of minority women in cancer screening, prevention, and treatment trials. *Annals of Epidemiology*, *10*(8 Suppl), S13-21. https://doi.org/10.1016/s1047-2797(00)00197-6

Buvinić, M., Medici, A., Fernández, E., & Torres, A. C. (2006). Gender Differentials in Health. In *Disease Control Priorities in Developing Countries* (2nd ed., pp. 195–210). The International Bank for Reconstruction and Development / The World Bank.

Byrd, W. M., & Clayton, L. A. (2001). Race, medicine, and health care in the United States: A historical survey. *Journal of the National Medical Association*, *93*(3 Suppl), 11S-34S.

Caldwell, P. H., Murphy, S. B., Butow, P. N., & Craig, J. C. (2004). Clinical trials in children. *The Lancet*, *364*(9436), 803–811. https://doi.org/10.1016/S0140-6736(04)16942-0

Califf, R. M., & Sugarman, J. (2015). Exploring the ethical and regulatory issues in pragmatic clinical trials. *Clinical Trials*, *12*(5), 436–441. https://doi.org/10.1177/1740774515598334

Canadian Institutes of Health Research. (2015, June 17). *Definitions of Sex and Gender*. https://cihr-irsc.gc.ca/e/47830.html

Cayleff, S. E. (1988). 'Prisoners of their own feebleness': Women, nerves and western medicine—A historical overview. *Social Science & Medicine*, *26*(12), 1199–1208. https://doi.org/10.1016/0277-9536(88)90152-9

Clark, L. T., Watkins, L., Piña, I. L., Elmer, M., Akinboboye, O., Gorham, M., Jamerson, B., McCullough, C., Pierre, C., Polis, A. B., Puckrein, G., & Regnante, J. M. (2019). Increasing Diversity in Clinical Trials: Overcoming Critical Barriers. *Current Problems in Cardiology*, *44*(5), 148–172. https://doi.org/10.1016/j.cpcardiol.2018.11.002

Coffey, C. S. (2010, November 1). *Adaptive designs and small clinical trials*. https://rarediseases.info.nih.gov/files/coffey.pdf

Cooter, R., Society for the Social History of Medicine, & British Paediatric Association (Eds.). (1992). *In the name of the child: Health and welfare, 1880-1940*. Routledge.

Cortes, L. R., Cisternas, C. D., & Forger, N. G. (2019). Does Gender Leave an Epigenetic Imprint on the Brain? *Frontiers in Neuroscience*, *13*, 173. https://doi.org/10.3389/fnins.2019.00173

Cromartie, J. (2018). *Rural America at a Glance, 2018 Edition* (EIB-200). US Department of Agriculture. https://www.ers.usda.gov/webdocs/publications/90556/eib-200.pdf?v=1374.7

Diamond, E. F. (1973). The WillowBrook Experiments. *The Linacre Quarterly*, *40*(2(9)), 133–137.

Đoàn, L. N., Takata, Y., Sakuma, K.-L. K., & Irvin, V. L. (2019). Trends in Clinical Research Including Asian American, Native Hawaiian, and Pacific Islander Participants Funded by the US National Institutes of Health, 1992 to 2018. *JAMA Network Open*, *2*(7), e197432. https://doi.org/10.1001/jamanetworkopen.2019.7432

Duma, N., Vera Aguilera, J., Paludo, J., Haddox, C. L., Gonzalez Velez, M., Wang, Y., Leventakos, K., Hubbard, J. M., Mansfield, A. S., Go, R. S., & Adjei, A. A. (2018).

Representation of Minorities and Women in Oncology Clinical Trials: Review of the Past 14 Years. *Journal of Oncology Practice*, *14*(1), e1–e10. https://doi.org/10.1200/JOP.2017.025288

Duster, T. (2015). A post-genomic surprise. The molecular reinscription of race in science, law and medicine: A post-genomic surprise. *The British Journal of Sociology*, *66*(1), 1–27. https://doi.org/10.1111/1468-4446.12118

Dymanus, K. A., Butaney, M., Magee, D. E., Hird, A. E., Luckenbaugh, A. N., Ma, M. W., Hall, M. E., Huelster, H. L., Laviana, A. A., Davis, N. B., Terris, M. K., Klaassen, Z., & Wallis, C. J. D. (2021). Assessment of gender representation in clinical trials leading to FDA approval for oncology therapeutics between 2014 and 2019: A systematic review‐based cohort study. *Cancer*, *127*(17), 3156–3162. https://doi.org/10.1002/cncr.33533

Ellison, G. T. H., Kaufman, J. S., Head, R. F., Martin, P. A., & Kahn, J. D. (2008). Flaws in the U.S. Food and Drug Administration's Rationale for Supporting the Development and Approval of BiDil as a Treatment for Heart Failure Only in Black Patients. *Journal of Law, Medicine & Ethics*, *36*(3), 449–457. https://doi.org/10.1111/j.1748-720X.2008.290.x

Epstein, S. (2008). *Inclusion: The Politics of Difference in Medical Research.* University of Chicago Press. http://qut.eblib.com.au/patron/FullRecord.aspx?p=408322

Eshera, N., Itana, H., Zhang, L., Soon, G., & Fadiran, E. O. (2015). Demographics of Clinical Trials Participants in Pivotal Clinical Trials for New Molecular Entity Drugs and Biologics Approved by FDA From 2010 to 2012. *American Journal of Therapeutics*, *22*(6), 435–455. https://doi.org/10.1097/MJT.0000000000000177

Fashoyin-Aje, L. A., Fernandes, L. L., Keegan, P., Sridhara, R., & Pazdur, R. (2018). Enrollment of Hispanics in cancer clinical trials: An FDA analysis. *Journal of Clinical Oncology*, *36*(15_suppl), e18670–e18670. https://doi.org/10.1200/JCO.2018.36.15_suppl.e18670

Fashoyin-Aje, L. A., Fernandes, L. L., Lemery, S., Keegan, P., Sridhara, R., Bull, J. C., & Pazdur, R. (2017). Asian representation in clinical trials of new drugs for the treatment of cancer. *Journal of Clinical Oncology*, *35*(15_suppl), 6564–6564. https://doi.org/10.1200/JCO.2017.35.15_suppl.6564

Fazal, M., Malisa, J., Rhee, J.-W., Witteles, R. M., & Rodriguez, F. (2021). Racial and Ethnic Disparities in Cardio-Oncology: A Call to Action. *JACC. CardioOncology*, *3*(2), 201–204. https://doi.org/10.1016/j.jaccao.2021.05.001

Feuerstein, I. M., Jenkins, M. R., Kornstein, S. G., Lauer, M. S., Scott, P. E., Raju, T. N. K., Johnson, T., Devaney, S., Lolic, M., Henderson, M., & Clayton, J. A. (2018). Working Together to Address Women's Health in Research and Drug Development: Summary of the 2017 Women's Health Congress Preconference Symposium. *Journal of Women's Health (2002)*, *27*(10), 1195–1203. https://doi.org/10.1089/jwh.2018.29019.pcss

Fisher, J. A., & Kalbaugh, C. A. (2011). Challenging assumptions about minority participation in US clinical research. *American Journal of Public Health*, *101*(12), 2217–2222. https://doi.org/10.2105/AJPH.2011.300279

Fletcher, J. C. (1993). Women's and Fetal Rights and Interests: Ethical Aspects. *Food and Drug Law Journal*, *48*(2), 213–221. JSTOR.

Fofana, M. O. (2013). The spectre of race in American medicine. *Medical Humanities*, *39*(2), 137–141. https://doi.org/10.1136/medhum-2013-010374

Food and Drug Administration. (2010). *Good Clinical Practice Handbook*. GMP Publications Inc.

Fragou, D., Pakkidi, E., Aschner, M., Samanidou, V., & Kovatsi, L. (2019). Smoking and DNA methylation: Correlation of methylation with smoking behavior and association with diseases and fetus development following prenatal exposure. *Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association*, *129*, 312–327. https://doi.org/10.1016/j.fct.2019.04.059

Franconi, F., Brunelleschi, S., Steardo, L., & Cuomo, V. (2007). Gender differences in drug responses. *Pharmacological Research*, *55*(2), 81–95. https://doi.org/10.1016/j.phrs.2006.11.001

Freestone, E. (2021, July 12). Increasing Diversity in Cancer Clinical Trials. *BYU College of Life Sciences Public Health*. https://ph.byu.edu/increasing-diversity-in-cancer-clinical-trials

Frohard-Dourlent, H., Dobson, S., Clark, B. A., Doull, M., & Saewyc, E. M. (2017). "I would have preferred more options": Accounting for non-binary youth in health research. *Nursing Inquiry*, *24*(1), e12150. https://doi.org/10.1111/nin.12150

Geller, S. E., Koch, A., Pellettieri, B., & Carnes, M. (2011). Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: Have we made progress? *Journal of Women's Health (2002)*, *20*(3), 315–320. https://doi.org/10.1089/jwh.2010.2469

Gopishetty, S., Kota, V., & Guddati, A. K. (2020). Age and race distribution in patients in phase III oncology clinical trials. *American Journal of Translational Research*, *12*(9), 5977–5983.

Gracia, J. N. (2020). COVID-19's Disproportionate Impact on Communities of Color Spotlights

the Nation's Systemic Inequities. *Journal of Public Health Management and Practice*, *26*(6), 518–521. https://doi.org/10.1097/PHH.0000000000001212

Granger, K. (2020, July 14). Why are elderly patients under-represented in clinical trials? *Cambridge Cognition*. https://www.cambridgecognition.com/blog/entry/why-are-elderly-patients-under-represented-in-clinical-trials

Gray, J. (2007). Why Can't a Woman Be More Like a Man? *Clinical Pharmacology & Therapeutics*, *82*(1), 15–17. https://doi.org/10.1038/sj.clpt.6100261

Grulich, A. E., & Kaldor, J. M. (2008). Trends in HIV incidence in homosexual men in developed countries. *Sexual Health*, *5*(2), 113. https://doi.org/10.1071/SH07075

Hackshaw, A. (2008). Small studies: Strengths and limitations. *European Respiratory Journal*, *32*(5), 1141–1143. https://doi.org/10.1183/09031936.00136408

Hahn, R. A., & Stroup, D. F. (1994). Race and ethnicity in public health surveillance: Criteria for the scientific use of social categories. *Public Health Reports (Washington, D.C.: 1974)*, *109*(1), 7–15.

Hamel, L. M., Penner, L. A., Albrecht, T. L., Heath, E., Gwede, C. K., & Eggly, S. (2016). Barriers to Clinical Trial Enrollment in Racial and Ethnic Minority Patients with Cancer. *Cancer Control*, *23*(4), 327–337. https://doi.org/10.1177/107327481602300404

Hao, T., Chen, B., & Qu, Y. (2016). An Automated Method for Gender Information Identification from Clinical Trial Texts. In X. Yin, J. Geller, Y. Li, R. Zhou, H. Wang, & Y. Zhang (Eds.), *Health Information Science* (Vol. 10038, pp. 109–118). Springer International Publishing. https://doi.org/10.1007/978-3-319-48335-1_12

Haslam, S. A., O'Brien, A., Jetten, J., Vormedal, K., & Penna, S. (2005). Taking the strain: Social identity, social support, and the experience of stress. *British Journal of Social Psychology*, *44*(3), 355–370. https://doi.org/10.1348/014466605X37468

Health (U.S.), N. I. of, Research, N. I. of H. (U S.) O. of E., Committee, N. O. N., & Committee, N. T. (2002). *Outreach Notebook for the Inclusion, Recruitment and Retention of Women and Minority Subjects in Clinical Research: Principal Investigators' Notebook*. National Institutes of Health, Office of Research on Women's Health, Office of Extramural Research, NIH Outreach Notebook Committee, NIH Tracking and Inclusion Committee. https://books.google.ca/books?id=-TxdY1pRVTUC

Humphreys, M. (2013). *Marrow of tragedy: The health crisis of the American Civil War*. The Johns Hopkins University Press.

Institute of Medicine (US). (1994). Justice in Clinical Studies: Guiding Principles. In A. Mastroianni, R. Faden, & D. Federman (Eds.), *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies: Volume I* (pp. 75–83). National Academies Press (US).

Ivanov, M., Kacevska, M., & Ingelman-Sundberg, M. (2012). Epigenomics and interindividual differences in drug response. *Clinical Pharmacology and Therapeutics*, *92*(6), 727–736. https://doi.org/10.1038/clpt.2012.152

Jayakrishnan, T., Aulakh, S., Baksh, M., Nguyen, K., Ailawadhi, M., Samreen, A., Parrondo, R., Sher, T., Roy, V., Manochakian, R., Paulus, A., Chanan-Khan, A., & Ailawadhi, S. (2021). Landmark Cancer Clinical Trials and Real-World Patient Populations: Examining Race and Age Reporting. *Cancers*, *13*(22), 5770.

https://doi.org/10.3390/cancers13225770

Jedwab, R., Khan, A. M., Russ, J., & Zaveri, E. D. (2021). Epidemics, pandemics, and social conflict: Lessons from the past and possible scenarios for COVID-19. *World Development*, *147*, 105629. https://doi.org/10.1016/j.worlddev.2021.105629

Joly, Y., Huerne, K., Arych, M., Bombard, Y., De Paor, A., Dove, E. S., Granados Moreno, P., Ho, C. W. L., Ho, C.-H., Van Hoyweghen, I., Kim, H., Lebret, A., Minssen, T., Ó. Cathaoir, K., Prince, A. E. R., Nair, A. P. S., Otlowski, M., Pepper, M. S., Sladek, R., … Dalpé, G. (2021). The Genetic Discrimination Observatory: Confronting novel issues in genetic discrimination. *Trends in Genetics*, *37*(11), 951–954. https://doi.org/10.1016/j.tig.2021.08.004

Joseph, G., & Dohan, D. (2009). Recruiting minorities where they receive care: Institutional barriers to cancer clinical trials recruitment in a safety-net hospital. *Contemporary Clinical Trials*, *30*(6), 552–559. https://doi.org/10.1016/j.cct.2009.06.009

Kamin Mukaz, D., Zakai, N. A., Cruz-Flores, S., McCullough, L. D., & Cushman, M. (2020). Identifying Genetic and Biological Determinants of Race-Ethnic Disparities in Stroke in the United States. *Stroke*, *51*(11), 3417–3424. https://doi.org/10.1161/STROKEAHA.120.030425

Kaminsky, Z., Wang, S.-C., & Petronis, A. (2006). Complex disease, gender and epigenetics. *Annals of Medicine*, *38*(8), 530–544. https://doi.org/10.1080/07853890600989211

Kaptchuk, T. J. (2001). The double-blind, randomized, placebo-controlled trial. *Journal of Clinical Epidemiology*, *54*(6), 541–549. https://doi.org/10.1016/S0895-4356(00)00347-4

Kaufman, J. S. (1999). How Inconsistencies in Racial Classification Demystify the Race

Construct in Public Health Statistics. *Epidemiology*, *10*(2), 101–103.

Keating, P., & Cambrosio, A. (2002). From Screening to Clinical Research: The Cure of Leukemia and the Early Development of the Cooperative Oncology Groups, 1955–1966. *Bulletin of the History of Medicine*, *76*(2), 299–334.

Kennedy, B. R., Mathis, C. C., & Woods, A. K. (2007). African Americans and their distrust of the health care system: Healthcare for diverse populations. *Journal of Cultural Diversity*, *14*(2), 56–60.

Kim, E. J., Kim, T., Conigliaro, J., Liebschutz, J. M., Paasche-Orlow, M. K., & Hanchate, A. D. (2018). Racial and Ethnic Disparities in Diagnosis of Chronic Medical Conditions in the USA. *Journal of General Internal Medicine*, *33*(7), 1116–1123. https://doi.org/10.1007/s11606-018-4471-1

Kim, J. H., & Scialli, A. R. (2011). Thalidomide: The tragedy of birth defects and the effective treatment of disease. *Toxicological Sciences: An Official Journal of the Society of Toxicology*, *122*(1), 1–6. https://doi.org/10.1093/toxsci/kfr088

Kraybill, E. N. (2004). The Challenge of Informed Consent in Neonatal Research. *Journal of Perinatology*, *24*(7), 407–408. https://doi.org/10.1038/sj.jp.7211149

Lacal, I., & Ventura, R. (2018). Epigenetic Inheritance: Concepts, Mechanisms and Perspectives. *Frontiers in Molecular Neuroscience*, *11*, 292. https://doi.org/10.3389/fnmol.2018.00292

Laventhal, N., Tarini, B. A., & Lantos, J. (2012). Ethical issues in neonatal and pediatric clinical trials. *Pediatric Clinics of North America*, *59*(5), 1205–1220. https://doi.org/10.1016/j.pcl.2012.07.007

Lee, S. J. (2004). A need for greater reporting of socioeconomic status and race in clinical trials.

Annals of the Rheumatic Diseases, *63*(12), 1700–1701. https://doi.org/10.1136/ard.2003.019588

Li, J., Djenaba, J. A., Soman, A., Rim, S. H., & Master, V. A. (2012). Recent Trends in Prostate Cancer Incidence by Age, Cancer Stage, and Grade, the United States, 2001–2007. *Prostate Cancer*, *2012*, 1–8. https://doi.org/10.1155/2012/691380

Liu, C., Zeinomar, N., Chung, W. K., Kiryluk, K., Gharavi, A. G., Hripcsak, G., Crew, K. D., Shang, N., Khan, A., Fasel, D., Manolio, T. A., Jarvik, G. P., Rowley, R., Justice, A. E., Rahm, A. K., Fullerton, S. M., Smoller, J. W., Larson, E. B., Crane, P. K., … Weng, C. (2021). Generalizability of Polygenic Risk Scores for Breast Cancer Among Women With European, African, and Latinx Ancestry. *JAMA Network Open*, *4*(8), e2119084. https://doi.org/10.1001/jamanetworkopen.2021.19084

Liu, K. A., & Mager, N. A. D. (2016). Women's involvement in clinical trials: Historical perspective and future implications. *Pharmacy Practice*, *14*(1), 708. https://doi.org/10.18549/PharmPract.2016.01.708

Ludmir, E. B., Fuller, C. D., Moningi, S., Mainwaring, W., Lin, T. A., Miller, A. B., Jethanandani, A., Espinoza, A. F., Verma, V., Smith, B. D., Smith, G. L., VanderWalde, N. A., Holliday, E. B., Guadagnolo, B. A., Stinchcombe, T. E., Jagsi, R., Gomez, D. R., Minsky, B. D., Rödel, C., & Fokas, E. (2020). Sex-Based Disparities Among Cancer Clinical Trial Participants. *Journal of the National Cancer Institute*, *112*(2), 211–213. https://doi.org/10.1093/jnci/djz154

Lüscher, T. F., Miller, V. M., Bairey Merz, C. N., & Crea, F. (2020). Diversity is richness: Why data reporting according to sex, age, and ethnicity matters. *European Heart Journal*,

41(33), 3117–3121. https://doi.org/10.1093/eurheartj/ehaa277

Macklin, R. (2010). Enrolling pregnant women in biomedical research. *The Lancet*, *375*(9715), 632–633. https://doi.org/10.1016/S0140-6736(10)60257-7

Mahipal, A., & Nguyen, D. (2014). Risks and Benefits of Phase 1 Clinical Trial Participation. *Cancer Control*, *21*(3), 193–199. https://doi.org/10.1177/107327481402100303

Mahmud, A., Zalay, O., Springer, A., Arts, K., & Eisenhauer, E. (2018). Barriers to Participation in Clinical Trials: A Physician Survey. *Current Oncology*, *25*(2), 119–125. https://doi.org/10.3747/co.25.3857

Martinson, J. J., Chapman, N. H., Rees, D. C., Liu, Y.-T., & Clegg, J. B. (1997). Global distribution of the CCR5 gene 32-basepair deletion. *Nature Genetics*, *16*(1), 100–103. https://doi.org/10.1038/ng0597-100

Matsuno, E., & Budge, S. L. (2017). Non-binary/Genderqueer Identities: A Critical Review of the Literature. *Current Sexual Health Reports*, *9*(3), 116–120. https://doi.org/10.1007/s11930-017-0111-8

Maugh II, T. H. (1996, August 9). Scientists Offer Explanation of HIV Resistance. *Los Angeles Times*. https://www.latimes.com/archives/la-xpm-1996-08-09-mn-32746-story.html

Mendis, S., Anand, S., Karasinska, J. M., Dasari, A., Unger, J. M., Gothwal, A., Ellis, L. M., Varadhachary, G., Kopetz, S., Overman, M. J., Raghav, K., & Loree, J. M. (2021). Sex Representation in Clinical Trials Associated with FDA Cancer Drug Approvals Differs Between Solid and Hematologic Malignancies. *The Oncologist*, *26*(2), 107–114. https://doi.org/10.1002/onco.13534

Mentz, R. J., Hernandez, A. F., Berdan, L. G., Rorick, T., O'Brien, E. C., Ibarra, J. C., Curtis, L.

H., & Peterson, E. D. (2016). Good Clinical Practice Guidance and Pragmatic Clinical Trials: Balancing the Best of Both Worlds. *Circulation*, *133*(9), 872–880. https://doi.org/10.1161/CIRCULATIONAHA.115.019902

Mogull, R. G. (2011). Hispanic-American Poverty. *Journal of Applied Business Research (JABR)*, *21*(3). https://doi.org/10.19030/jabr.v21i3.1472

Mone, A., & Mehl, V. (2020, November 2). Age is a primary determinant of melanoma treatment resistance, two studies find. *John Hopkins Medicine Newsroom*. https://www.hopkinsmedicine.org/news/newsroom/news-releases/age-is-a-primary-determinant-of-melanoma-treatment-resistance-study-finds

Mudaranthakam, D. P., Gajewski, B., Krebill, H., Coulter, J., Springer, M., Calhoun, E., Hughes, D., Mayo, M., & Doolittle, G. (2022). Barriers to Clinical Trial Participation: Comparative Study Between Rural and Urban Participants. *JMIR Cancer*, *8*(2), e33240. https://doi.org/10.2196/33240

Murthy, V. H., Krumholz, H. M., & Gross, C. P. (2004). Participation in Cancer Clinical Trials: Race-, Sex-, and Age-Based Disparities. *JAMA*, *291*(22), 2720. https://doi.org/10.1001/jama.291.22.2720

National Institute on Aging. (2017, May 17). *Clinical Trials: Benefits, Risks, and Safety*. National Institute of Health. https://www.nia.nih.gov/health/clinical-trials-benefits-risks-and-safety

Nazha, B., Mishra, M., Pentz, R., & Owonikoko, T. K. (2019). Enrollment of Racial Minorities in Clinical Trials: Old Problem Assumes New Urgency in the Age of Immunotherapy. *American Society of Clinical Oncology Educational Book. American Society of Clinical*

Oncology. Annual Meeting, *39*, 3–10. https://doi.org/10.1200/EDBK_100021

Nellhaus, E. M., & Davies, T. H. (2017). Evolution of Clinical Trials throughout History. *Marshall Journal of Medicine*, *3*(1). https://doi.org/10.18590/mjm.2017.vol3.iss1.9

Nguyen, H.-A. T., Zheng, A., Gugel, A., & Kistin, C. J. (2021). Asians and Asian Subgroups are Underrepresented in Medical Research Studies Published in High-Impact Generalist Journals. *Journal of Immigrant and Minority Health*, *23*(3), 646–649. https://doi.org/10.1007/s10903-021-01142-6

Nie, J.-B. (2002). Japanese doctors' experimentation in wartime China. *The Lancet*, *360*, s5–s6. https://doi.org/10.1016/S0140-6736(02)11797-1

NIH National Cancer Institute. (2021, April 22). *Common Cancer Types*. https://www.cancer.gov/types/common-cancers

NIH Office of Research on Women Health. (2003). *Science Meets Reality: Recruitment and Retention of Women in Clinical Studies, and the Critical Role of Relevance*. NIH Office of Research on Women's Health.

Noah, B. A. (2003). The Participation of Underrepresented Minorities in Clinical Research. *American Journal of Law & Medicine*, *29*(2–3), 221–245. https://doi.org/10.1017/S0098858800002823

O'Connell, H. A. (2012). The Impact of Slavery on Racial Inequality in Poverty in the Contemporary U.S. South. *Social Forces*, *90*(3), 713–734. https://doi.org/10.1093/sf/sor021

Office for Human Research Protections. (2016, March 18). *Special Protections for Children as Research Subjects*. Hhs.Gov.

https://www.hhs.gov/ohrp/regulations-and-policy/guidance/special-protections-for-childr en/index.html

Office of the Secretary. (1979). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Department of Health, Education, and Welfare. https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf

Office on Women's Health. (2019, April 1). *Policy of inclusion of women in clinical trials*. OASH. https://www.womenshealth.gov/30-achievements/04

Osborne, N. G., & Feit, M. D. (1992). The Use of Race in Medical Research. *JAMA: The Journal of the American Medical Association*, *267*(2), 275. https://doi.org/10.1001/jama.1992.03480020085037

Parra, A., Karnad, A. B., & Thompson, I. M. (2014). Hispanic accrual on randomized cancer clinical trials: A call to arms. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *32*(18), 1871–1873. https://doi.org/10.1200/JCO.2013.51.7946

Perry, D. J., Wasserfall, C. H., Oram, R. A., Williams, M. D., Posgai, A., Muir, A. B., Haller, M. J., Schatz, D. A., Wallet, M. A., Mathews, C. E., Atkinson, M. A., & Brusko, T. M. (2018). Application of a Genetic Risk Score to Racially Diverse Type 1 Diabetes Populations Demonstrates the Need for Diversity in Risk-Modeling. *Scientific Reports*, *8*(1), 4529. https://doi.org/10.1038/s41598-018-22574-5

Phillips, S. P. (2008). Measuring the health effects of gender. *Journal of Epidemiology & Community Health*, *62*(4), 368–371. https://doi.org/10.1136/jech.2007.062158

Piferrer, F. (2021). Epigenetic mechanisms in sex determination and in the evolutionary transitions between sexual systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *376*(1832), 20200110. https://doi.org/10.1098/rstb.2020.0110

Poon, R., Khanijow, K., Umarjee, S., Fadiran, E., Yu, M., Zhang, L., & Parekh, A. (2013). Participation of Women and Sex Analyses in Late-Phase Clinical Trials of New Molecular Entity Drugs and Biologics Approved by the FDA in 2007–2009. *Journal of Women's Health*, *22*(7), 604–616. https://doi.org/10.1089/jwh.2012.3753

Population Reference Bureau. (2004). *Asian-American Children Are Members of a Diverse and Urban Population*. https://www.prb.org/resources/asian-american-children-are-members-of-a-diverse-and-urban-population/

Pruitt, S. L., Laccetti, A. L., Xuan, L., Halm, E. A., & Gerber, D. E. (2017). Revisiting a longstanding clinical trial exclusion criterion: Impact of prior cancer in early-stage lung cancer. *British Journal of Cancer*, *116*(6), 717–725. https://doi.org/10.1038/bjc.2017.27

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramamoorthy, A., Pacanowski, M., Bull, J., & Zhang, L. (2015). Racial/ethnic differences in drug disposition and response: Review of recently approved drugs. *Clinical Pharmacology & Therapeutics*, *97*(3), 263–273. https://doi.org/10.1002/cpt.61

Reverby, S. (Ed.). (2000). *Tuskegee's truths: Rethinking the Tuskegee syphilis study*. University of North Carolina Press.

Ridda, I., Lindley, R., & MacIntyre, R. C. (2008). Research: The challenges of clinical trials in the exclusion zone: The case of the frail elderly: Exclusion zone: The case of the frail elderly. *Australasian Journal on Ageing*, *27*(2), 61–66. https://doi.org/10.1111/j.1741-6612.2008.00288.x

Rivara, F. P., & Finberg, L. (2001). Use of the Terms Race and Ethnicity. *Archives of Pediatrics & Adolescent Medicine*, *155*(2), 119. https://doi.org/10.1001/archpedi.155.2.119

Rivers, D., August, E. M., Sehovic, I., Lee Green, B., & Quinn, G. P. (2013). A systematic review of the factors influencing African Americans' participation in cancer clinical trials. *Contemporary Clinical Trials*, *35*(2), 13–32. https://doi.org/10.1016/j.cct.2013.03.007

Robinson, W. M., & Unruh, B. T. (2011). The Hepatitis Experiment at the Willowbrook State School. In *The Oxford Textbook of Clinical Research Ethics* (pp. 80–85). Oxford University Press.

Russell, L. (2010, December 16). Fact Sheet: Health Disparities by Race and Ethnicity. *Center for American Progress*. https://www.americanprogress.org/article/fact-sheet-health-disparities-by-race-and-ethnicity/

Saey, T. H. (2021, March 4). DNA databases are too white, so genetics doesn't help everyone. How do we fix that? *ScienceNews*. https://www.sciencenews.org/article/genetics-race-dna-databases-reference-genome-too-white

Saraswat, N., Pillay, R., Everett, B., & George, A. (2020). Knowledge, attitudes and practices of

South Asian immigrants in developed countries regarding oral cancer: An integrative review. *BMC Cancer*, *20*(1), 477. https://doi.org/10.1186/s12885-020-06944-9

Savitt, T. L. (1982). The Use of Blacks for Medical Experimentation and Demonstration in the Old South. *The Journal of Southern History*, *48*(3), 331. https://doi.org/10.2307/2207450

Schwartz, J. B. (2007). The Current State of Knowledge on Age, Sex, and Their Interactions on Clinical Pharmacology. *Clinical Pharmacology & Therapeutics*, *82*(1), 87–96. https://doi.org/10.1038/sj.clpt.6100226

Seidenfeld, J., Horstmann, E., Emanuel, E. J., & Grady, C. (2008). Participants in phase 1 oncology research trials: Are they vulnerable? *Archives of Internal Medicine*, *168*(1), 16–20. https://doi.org/10.1001/archinternmed.2007.6

Shah, N. S., & Kandula, N. R. (2020). Addressing Asian American Misrepresentation and Underrepresentation in Research. *Ethnicity & Disease*, *30*(3), 513–516. https://doi.org/10.18865/ed.30.3.513

Sharath, S. E., Kougias, P., Pisimisis, G., & Barshes, N. R. (2016). The impact of clinical, psychological, behavioral, social, and environmental factors on self-perceived symptom severity in a male cohort with intermittent claudication. *Journal of Vascular Surgery*, *63*(5), 1296-1304.e4. https://doi.org/10.1016/j.jvs.2015.11.040

Sharma, S., Kelly, T. K., & Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis*, *31*(1), 27–36. https://doi.org/10.1093/carcin/bgp220

Shenoy, P., & Harugeri, A. (2015). Elderly patients' participation in clinical trials. *Perspectives in Clinical Research*, *6*(4), 184–189. https://doi.org/10.4103/2229-3485.167099

Simon, M. A., de la Riva, E. E., Bergan, R., Norbeck, C., McKoy, J. M., Kulesza, P., Dong, X.,

Schink, J., & Fleisher, L. (2014). Improving Diversity in Cancer Research Trials: The Story of the Cancer Disparities Research Network. *Journal of Cancer Education*, *29*(2), 366–374. https://doi.org/10.1007/s13187-014-0617-y

Spencer, Q. (2014). A Radical Solution to the Race Problem. *Philosophy of Science*, *81*(5), 1025–1038. https://doi.org/10.1086/677694

Spencer, Q. (2015a). Philosophy of race meets population genetics. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *52*, 46–55. https://doi.org/10.1016/j.shpsc.2015.04.003

Spencer, Q. (2015b). Philosophy of race meets population genetics. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *52*, 46–55. https://doi.org/10.1016/j.shpsc.2015.04.003

Spencer, Q. N. J. (2018). A racial classification for medical genetics. *Philosophical Studies*, *175*(5), 1013–1037. https://doi.org/10.1007/s11098-018-1072-0

Srikanth, A. (2020, October 3). Why Black and Indigenous Americans are skeptical of a vaccine. *Changing America*. https://thehill.com/changing-america/respect/diversity-inclusion/520787-why-black-and-indigenous-americans-are-skeptical

Stanford Medicine Cancer Institute. (2022). *Diversity in Cancer Clinical Trials*. Cancer Institute. https://med.stanford.edu/cancer/community/trials.html

Sternberg, A. (2020). Ongoing Minority Underrepresentation in Clinical Research Leads to Efforts to Bridge the Divide. *Targeted Therapies in Oncology*, *9*, 91.

Strauss, D., de la Salle, S., Sloshower, J., & Williams, M. T. (2021). Research abuses against

people of colour and other vulnerable groups in early psychedelic research. *Journal of Medical Ethics*, medethics-2021-107262. https://doi.org/10.1136/medethics-2021-107262

Sullivan, L. S. (2020). Trust, Risk, and Race in American Medicine. *Hastings Center Report*, *50*(1), 18–26. https://doi.org/10.1002/hast.1080

Taylor, J. S. W., & Ellis, G. R. (2002). Racial differences in responses to drug treatment: Implications for pharmacotherapy of heart failure. *American Journal of Cardiovascular Drugs: Drugs, Devices, and Other Interventions*, *2*(6), 389–399. https://doi.org/10.2165/00129784-200202060-00004

Townsley, C. A., Selby, R., & Siu, L. L. (2005). Systematic Review of Barriers to the Recruitment of Older Patients With Cancer Onto Clinical Trials. *Journal of Clinical Oncology*, *23*(13), 3112–3124. https://doi.org/10.1200/JCO.2005.00.141

Uhl, K., Parekh, A., & Kweder, S. (2007). Females in Clinical Studies: Where are We Going? *Clinical Pharmacology & Therapeutics*. https://doi.org/10.1038/sj.clpt.6100112

Unger, J. M., Hershman, D. L., Albain, K. S., Moinpour, C. M., Petersen, J. A., Burg, K., & Crowley, J. J. (2013). Patient Income Level and Cancer Clinical Trial Participation. *Journal of Clinical Oncology*, *31*(5), 536–542. https://doi.org/10.1200/JCO.2012.45.4553

Unger, J. M., Vaidya, R., Hershman, D. L., Minasian, L. M., & Fleury, M. E. (2019). Systematic Review and Meta-Analysis of the Magnitude of Structural, Clinical, and Physician and Patient Barriers to Cancer Clinical Trial Participation. *Journal of the National Cancer Institute*, *111*(3), 245–255. https://doi.org/10.1093/jnci/djy221

U.S. Cancer Statistics Working Group. (2021, June). *U.S. Cancer Statistics Data Visualizations Tool, based on 2020 submission data (1999-2018): U.S. Department of Health and*

Human Services, Centers for Disease Control and Prevention and National Cancer Institute. www.cdc.gov/cancer/dataviz

U.S. Department of Health and Human Services, Food and Drug Administration, Office of the Commissioner, Office of Minority Health, Office of Women's Health, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, & Center for Devices and Radiologic Health. (2016). *Collection of Race and Ethnicity Data in Clinical Trials: Guidance for Industry and Food and Drug Administration Staff*. Food and Drug Administration. https://www.fda.gov/media/75453/download

U.S. Department of Health and Human Services, Food and Drug Administration, Oncology Center of Excellence, Center for Drug Evaluation and Research, & Center for Biologics Evaluation and Research. (2022). *Inclusion of Older Adults in Cancer Clinical Trials: Guidance for Industry*. Food and Drug Administration. https://www.fda.gov/media/156616/download

Valdes-Stauber, J., Vietz, E., & Kilian, R. (2013). The impact of clinical conditions and social factors on the psychological distress of cancer patients: An explorative study at a consultation and liaison service in a rural general hospital. *BMC Psychiatry*, *13*(1), 226. https://doi.org/10.1186/1471-244X-13-226

Varma, T., Wallach, J. D., Miller, J. E., Schnabel, D., Skydel, J. J., Zhang, A. D., Dinan, M. A., Ross, J. S., & Gross, C. P. (2021). Reporting of Study Participant Demographic Characteristics and Demographic Representation in Premarketing and Postmarketing Studies of Novel Cancer Therapeutics. *JAMA Network Open*, *4*(4), e217063. https://doi.org/10.1001/jamanetworkopen.2021.7063

Viby-Mogensen, J., Engbaek, J., Eriksson, L. I., Gramstad, L., Jensen, E., Jensen, F. S., Koscielniak-Nielsen, Z., Skovgaard, L. T., & Østergaard, D. (1996). Good Clinical Research Practice (GCRP) in pharmacodynamic studies of neuromuscular blocking agents. *Acta Anaesthesiologica Scandinavica*, *40*(1), 59–74. https://doi.org/10.1111/j.1399-6576.1996.tb04389.x

Wailoo, K. (2006). Stigma, race, and disease in 20th century America. *The Lancet*, *367*(9509), 531–533. https://doi.org/10.1016/S0140-6736(06)68186-5

Wald, P. (2000). Future Perfect: Grammar, Genes, and Geography,. *New Literary History*, *31*(4), 681–708.

Wall, L. L. (2006). The medical ethics of Dr J Marion Sims: A fresh look at the historical record. *Journal of Medical Ethics*, *32*(6), 346–350. https://doi.org/10.1136/jme.2005.012559

Watts, G. (2012). Why the exclusion of older people from clinical research must stop. *BMJ*, *344*(may21 1), e3445–e3445. https://doi.org/10.1136/bmj.e3445

Welch, M. J., Lally, R., Miller, J. E., Pittman, S., Brodsky, L., Caplan, A. L., Uhlenbrauck, G., Louzao, D. M., Fischer, J. H., & Wilfond, B. (2015). The ethics and regulatory landscape of including vulnerable populations in pragmatic clinical trials. *Clinical Trials*, *12*(5), 503–510. https://doi.org/10.1177/1740774515597701

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed. 2016). Springer International Publishing : Imprint: Springer. https://doi.org/10.1007/978-3-319-24277-4

Wilcox, S., Shumaker, S. A., Bowen, D. J., Naughton, M. J., Rosal, M. C., Ludlam, S. E., Dugan, E., Hunt, J. R., & Stevens, S. (2001). Promoting adherence and retention to clinical trials in special populations: A women's health initiative workshop. *Controlled Clinical Trials*,

22(3), 279–289. https://doi.org/10.1016/s0197-2456(00)00130-6

Williams, D. R. (1997). Race and health: Basic questions, emerging directions. *Annals of Epidemiology*, *7*(5), 322–333. https://doi.org/10.1016/S1047-2797(97)00051-3

Williams, D. R., & Mohammed, S. A. (2009). Discrimination and racial disparities in health: Evidence and needed research. *Journal of Behavioral Medicine*, *32*(1), 20–47. https://doi.org/10.1007/s10865-008-9185-0

Wizemann, T. M., & Pardue, M. L. (Eds.). (2001). *Exploring the biological contributions to human health: Does sex matter?* National Academy Press.

Yanow, D. (2015). *Constructing "race" and "ethnicity" in America: Category-making in public policy and administration*. http://site.ebrary.com/id/11025586

Young, R. C. (2010). Cancer clinical trials—A chronic but curable crisis. *The New England Journal of Medicine*, *363*(4), 306–309. https://doi.org/10.1056/NEJMp1005843

Yuan, N. P., Bartgis, J., & Demers, D. (2014). Promoting ethical research with American Indian and Alaska Native people living in urban areas. *American Journal of Public Health*, *104*(11), 2085–2091. https://doi.org/10.2105/AJPH.2014.302027

Zhang, Y., & Sirard, M.-A. (2021). Epigenetic inheritance of acquired traits through DNA methylation. *Animal Frontiers*, *11*(6), 19–27. https://doi.org/10.1093/af/vfab052

Zucker, I., & Prendergast, B. J. (2020). Sex differences in pharmacokinetics predict adverse drug reactions in women. *Biology of Sex Differences*, *11*(1), 32. https://doi.org/10.1186/s13293-020-00308-5

Copyright

Appendix A: Search strategy and results on clinicaltrials.gov

Search Strategy:

Condition: **[Cancer Type]** chosen between breast cancer, lung cancer, prostate cancer, colorectal cancer, melanoma, lymphoma, kidney cancer, leukemia, pancreatic cancer, or liver cancer
Study type: All Studies
Study Results: Studies with results
Status: completed
Country: **[Location]** chosen between United States, Canada, or Mexico
Phase: 3
Results First Posted from: 01-01-2017 until (no end date)

Results:

If a country is missing for a particular cancer type, it is because the search returned no results (which was often Mexico).

Search #1: 31 results Condition: **Lung Cancer** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States** Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	URL: https://clinicaltrials.gov/ct2/results?cond=Lung+Cancer&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst
Search #2: 9 results Condition: **Melanoma** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States** Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	URL: https://clinicaltrials.gov/ct2/results?cond=Melanoma&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst
Search #3: 41 results Condition: **Lymphoma** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States** Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	URL: https://clinicaltrials.gov/ct2/results?cond=Lymphoma&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst

Search #4: 9 results Condition: **Kidney Cancer** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States** Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	URL: https://clinicaltrials.gov/ct2/results?cond=Kidney+Cancer&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst
Search #5: 33 results Condition: **Leukemia** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States** Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	URL: https://clinicaltrials.gov/ct2/results?cond=Leukemia&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst
Search #6: 9 results Condition: **Pancreatic Cancer** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States** Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	URL: https://clinicaltrials.gov/ct2/results?cond=Pancreatic+Cancer&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst
Search #7: 14 results Condition: **Liver Cancer** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States** Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	URL: https://clinicaltrials.gov/ct2/results?cond=Liver+Cancer&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst
Search #8: 53 results Condition: **Breast Cancer** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States**	URL: https://clinicaltrials.gov/ct2/results?cond=Breast+Cancer&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_

Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst
Search #9: 33 results Condition: **Prostate Cancer** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States** Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	URL: https://clinicaltrials.gov/ct2/results?cond=Prostate+Cancer&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst
Search #10: 14 results Condition: **Colorectal Cancer** Study type: All Studies Study Results: Studies with results Status: completed Country: **United States** Phase: 3 Results First Posted from: 01-01-2017 until (no end date)	URL: https://clinicaltrials.gov/ct2/results?cond=Colo-rectal+Cancer&term=&type=&rslt=With&recrs=e&age_v=&gndr=&intr=&titles=&outc=&spons=&lead=&id=&cntry=US&state=&city=&dist=&locn=&phase=2&rsub=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s=&sfpd_e=&rfpd_s=01%2F01%2F2017&rfpd_e=&lupd_s=&lupd_e=&sort=nwst

Appendix B: Python Script for Data Extraction

```
## Importing libraries ##
import xml.etree.ElementTree as ET #Gives us an XML tree object
import os # To navigate directories and files

class extract_data:
      """Easily extract data from XML files"""

      ## Constructor
      #     Params
      #     dir: list of files from which data must be manipulated
      def __init__(self, dir=os.getcwd()):
            # Make sure dir is in the right format
            self.dir = dir if os.path.exists(dir) else os.getcwd()
            self.files = os.listdir(dir)
            # Make a directory to store the results of our extraction
            if 'parsed_data' not in self.files:
                  os.mkdir('parsed_data')
```

```
      ## Function to extract the relevant data from the XML files
      def parse(self):
            for f in self.files:
                  if ".xml" in f: # Only parse XML files
                        # Get the XML root
                        tree = ET.parse(f)
                        root = tree.getroot()
                        # New XML tree to hold the extracted data
                        newRoot = ET.Element('info')

                        # Get sponsorship data
                        sponsors = root.find('sponsors')
                        newRoot.append(sponsors)

                        # Get location count
                        locations = root.findall('location')
                        loc_count = ET.Element('location_count')
                        loc_count.text = str(len(locations))
                        newRoot.append(loc_count)

                        # Extract baseline characteristics
                        baseline = root.find('clinical_results/baseline')
                        newRoot.append(baseline)

                        # Save our new tree
                        newTree = ET.ElementTree(newRoot)
                        newTree.write('parsed_data/info_' + f)

extractor = extract_data()
extractor.parse()
```

Appendix C: Summary of data extracted for primary outcome analysis

Trial Identifier	Cancer Type	Total # of Participants	Sex Data	Race Data	Ethnicity Data	Age Data	Eligibility Criteria Pertaining to Age
NCT00003830	Breast	5611	Male: 0 Female: 5611	Not reported	Not reported	Mean: 56 Standard Deviation: 11.1	18 Years and above
NCT00009945	Breast	3323	Male: 0 Female: 3323	Not reported	Not reported	Mean: 54 Standard Deviation: 10.4	18 Years and above
NCT00014222	Breast	2103	Male: 0 Female: 2103	American Indian or Alaska Native: 20 Asian: 74 Native Hawaiian or Other Pacific Islander: 2 Black or African American: 98 White: 1866 More than one race: 0 Unknown or Not Reported: 43	Not reported	Median: 47.7 Full Range: 22.7-63.8	60 Years and under
NCT00041119	Breast	3871	Male: 0 Female: 3871	Not reported	Not reported	<18: 0 18-65: 3399 >65: 472 Mean: 53.4 Standard Deviation: 9.6	18 Years and above
NCT00053898	Breast	3104	Male: 0 Female: 3104	Not reported	Not reported	Mean: 61.0 Standard Deviation: 7.8	Any
NCT00075764	Breast	694	Male: 0 Female: 694	Not reported	Not reported	Median: 65 Full Range: 27-92	18 Years and above
NCT00093795	Breast	4867	Male: 0 Female: 4867	Not reported	Not reported	Mean: 51 Standard Deviation: 9.5	18 Years and above
NCT00195013	Breast	30	Male: 0 Female: 30	Not reported	Not reported	<18: 0 18-65: 25 >65: 5 Median: 58 Full Range: 37-74	18 Years and above
NCT00265759	Breast	610	Male: 0 Female: 610	Not reported	Not reported	Median: 65 Full Range: 43-90	Any
NCT00296036	Breast	127	Male: 26 Female: 101	Not reported	Not reported	<50: 20 50-60: 36 >60: 71	18 Years and above

NCT00376597	Breast	554	Male: 0 Female: 554	Not reported	Not reported	Mean: 57.7 Standard Deviation: 11.3	18 Years and above
NCT00382018	Breast	564	Male: 0 Female: 564	White: 469 Black: 95	Not reported	<55: 230 55+: 334	18 Years and above
NCT00775645	Breast	409	Male: 0 Female: 409	American Indian or Alaska Native: 1 Asian: 20 Native Hawaiian or Other Pacific Islander: 6 Black or African American: 37 White: 325 More than one race: 3 Unknown or Not Reported: 17	Hispanic or Latino: 36 Not Hispanic or Latino: 358 Unknown or Not Reported: 15	Median: 52 Full Range: 26-80	18 Years - 120 Years
NCT00789581	Breast	614	Male: 0 Female: 614	American Indian or Alaska Native: 2 Asian: 4 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 113 White: 488 More than one race: 0 Unknown or Not Reported: 7	Not reported	<50: 217 50+: 397	18 Years and above
NCT00956813	Breast	210	Male: 0 Female: 210	Not reported	Not reported	<50: 49 50+: 161	18 Years and above
NCT01224678	Breast	300	Male: 0 Female: 300	American Indian or Alaska Native: 1 Asian: 14 Native Hawaiian or Other Pacific Islander: 2 Black or African American: 35 White: 238 More than one race: 5 Unknown or Not Reported: 5	Hispanic or Latino: 42 Not Hispanic or Latino: 253 Unknown or Not Reported: 5	Median: 43 Full Range: 22.7-59.4	55 Years and under
NCT01376349	Breast	443	Male: 0 Female: 443	Not reported	Not reported	Mean: 57.3 Standard Deviation: 7.4	18 Years and above
NCT01385137	Breast	249	Male: 0 Female: 249	White: 217 Black: 20 Asian: 4 Native American: 1	Hispanic or Latino: 16 Not Hispanic or Latino: 225	Median: 59.2 Full Range: 40-84	18 Years - 120 Years

				Multiracial: 2 Unknown: 5	Unknown or Not Reported: 8		
NCT01573442	Breast	208	Male: 0 Female: 208	American Indian or Alaska Native: 0 Asian: 5 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 6 White: 194 More than one race: 0 Unknown or Not Reported: 3	Not reported	Mean: 60.0 Standard Deviation: 9.2	18 Years and above
NCT01591746	Breast	131	Male: 0 Female: 131	Not reported	Not reported	Mean: 49.1 Standard Deviation: 11.3	18 Years and above
NCT01598298	Breast	289	Male: 0 Female: 289	American Indian or Alaska Native: 2 Asian: 9 Native Hawaiian or Other Pacific Islander: 1 Black or African American: 27 White: 248 More than one race: 1 Unknown or Not Reported: 1	Hispanic or Latino: 11 Not Hispanic or Latino: 277 Unknown or Not Reported: 1	Median: 60 Full Range: 27-83	120 Years and under
NCT01856543	Breast	143	Male: 0 Female: 143	American Indian or Alaska Native: 0 Asian: 13 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 25 White: 95 More than one race: 0 Unknown or Not Reported: 10	Not reported	Mean: 49 Full Range: 26-80	18 Years and above
NCT01945775	Breast	431	Male: 7 Female: 424	American Indian or Alaska Native: 0 Asian: 47 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 12 White: 298 More than one race: 0 Unknown or Not Reported: 74	Hispanic or Latino: 46 Not Hispanic or Latino: 318 Unknown or Not Reported: 67	Mean: 48.1 Standard Deviation: 11.80	18 Years and above

NCT02574455	Breast	529	Male: 2 Female: 527	Asian: 22 Black: 62 White: 418 Other: 27	Hispanic or Latino: 45 Not Hispanic or Latino: 460 Unknown or Not Reported: 24	Mean: 54.0 Standard Deviation: 11.50	18 Years and above
NCT02961790	Breast	113	Male: 0 Female: 113	American Indian or Alaska Native: 0 Asian: 2 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 3 White: 108 More than one race: 0 Unknown or Not Reported: 0	Not reported	Mean: 57.1 Standard Deviation: 8.2	18 Years and above
NCT00265850	Colorectal	1137	Male: 697 Female: 440	Not reported	Not reported	Median: 59.1 Full Range: 20.8-89.5	18 Years and above
NCT01099449	Colorectal	353	Male: 169 Female: 184	Not reported	Not reported	Median: 56 Full Range: 50-65	18 Years and above
NCT01931150	Colorectal	11	Male: 9 Female: 2	Not reported	Not reported	<18: 0 18-65: 10 >65: 1	18 Years and above
NCT02254486	Colorectal	621	Male: 327 Female: 294	Not reported	Not reported	<18: 1 18-65: 481 >65: 139 Mean: 57.5 Standard Deviation: 10.44	18 Years - 85 Years
NCT02776683	Colorectal	123	Male: 68 Female: 55	American Indian or Alaska Native: 0 Asian: 33 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 6 White: 82 More than one race: 0 Unknown or Not Reported: 2	Hispanic or Latino: 15 Not Hispanic or Latino: 106 Unknown or Not Reported: 2	Mean: 58 Standard Deviation: 11.87	18 Years and above
NCT00087022	Kidney	864	Male: 574 Female: 290	Not reported	Not reported	Mean: 58.0 Standard Deviation: 9.98	18 Years - 120 Years
NCT01606787	Kidney	199	Male: 126 Female: 73	American Indian or Alaska Native: 0	Not reported	Median: 58 Full Range: 48-66	18 Years and above

				Asian: 10 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 15 White: 161 More than one race: 0 Unknown or Not Reported: 13			
NCT00469144	Leukemia	225	Male: 114 Female: 111	Not reported	Not reported	Median: 50 Full Range: 14-66	65 Years and under
NCT00887068	Leukemia	181	Male: 108 Female: 73	American Indian or Alaska Native: 0 Asian: 0 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 2 White: 171 More than one race: 0 Unknown or Not Reported: 8	Not reported	<18: 0 18-65: 144 >65: 37	18 Years - 75 Years
NCT01802333	Leukemia	738	Male: 378 Female: 360	American Indian or Alaska Native: 6 Asian: 17 Native Hawaiian or Other Pacific Islander: 3 Black or African American: 55 White: 613 More than one race: 2 Unknown or Not Reported: 42	Hispanic or Latino: 62 Not Hispanic or Latino: 628 Unknown or Not Reported: 48	Median: 49.8 Full Range: 18.8-61.0 <40: 188 >40: 550	18 Years - 60 Years
NCT02801578	Leukemia	11	Male: 5 Female: 6	American Indian or Alaska Native: 0 Asian: 0 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 0 White: 11 More than one race: 0 Unknown or Not Reported: 0	Not reported	<18: 0 18-65: 8 >65: 3 Median: 68 Full Range: 52-79	18 Years and above
NCT00788697	Liver	240	Male: 123 Female: 117	White: 161 Black: 32 Asian: 15 Other: 32	Not reported	<18: 0 18-65: 190 >65: 50	18 Years and above
NCT00829413	Liver	259	Male: 136 Female: 123	White: 206 Black: 22	Not reported	<18: 0 18-65: 185	18 Years and above

				Asian: 12 Other: 19		>65: 74 Mean: 56.9 Standard Deviation: 13.4	
NCT01596283	Liver	135	Male: 60 Female: 75	Not reported	Not reported	Mean: 57 Standard Deviation: 13	18 Years and above
NCT00003901	Lung	1047	Male: 538 Female: 509	White: 959 Black/African American: 60 Asian: 14 American Indian/Alaska Native: 1 Other: 2	Hispanic/Latino: 11	Median: 67.2 Full Range: 33.6-89.5	18 Years and above
NCT00153803	Lung	245	Male: 145 Female: 100	American Indian or Alaska Native: 0 Asian: 3 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 13 White: 215 More than one race: 0 Unknown or Not Reported: 14	Not reported	Median: 67.5 Full Range: 38-89	18 Years and above
NCT00693992	Lung	210	Male: 117 Female: 93	American Indian or Alaska Native: 0 Asian: 3 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 31 White: 172 More than one race: 0 Unknown or Not Reported: 4	Not reported	Median: 66 Full Range: 25-89	18 Years and above
NCT01355497	Lung	320	Male: 227 Female: 93	Not reported	Not reported	Median: 61 Full Range: 40-81	30 Years and above
NCT02027428	Lung	202	Male: 130 Female: 72	American Indian or Alaska Native: 1 Asian: 1 Black or African American: 11 White: 187 Other: 2	Hispanic or Latino: 8 Not Hispanic or Latino: 193 Unknown or Not Reported: 1	Mean: 67.0 Standard Deviation: 8.70 <65: 75 >65: 127 <70: 120 >70: 82 <75: 165 >75: 37	18 Years and above
NCT02785939	Lung	41	Male: 27 Female: 14	Asian: 1 Black: 3	Hispanic: 0	Median: 64.9 Full Range: 46.9-80.7	18 Years and above

				Native American: 1 White: 36			
NCT02965378	Lung	34	Male: 24 Female: 10	Black: 4 White: 30	Hispanic: 2	Median: 66.2 Full Range: 49.0-88.0	25 Years and above
NCT00118209	Lymphoma	491	Male: 265 Female: 225	American Indian or Alaska Native: 4 Asian: 17 Native Hawaiian or Other Pacific Islander: 1 Black or African American: 60 White: 385 More than one race: 5 Unknown or Not Reported: 19	Hispanic or Latino: 31 Not Hispanic or Latino: 434 Unknown or Not Reported: 26	Median: 58 Full Range: 18-86	18 Years and above
NCT00566228	Lymphoma	121	Male: 85 Female: 36	Not reported	Not reported	Median: 58 Full Range: 50-64	18 Years - 120 Years
NCT00577993	Lymphoma	193	Male: 90 Female: 103	American Indian or Alaska Native: 0 Asian: 2 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 11 White: 158 More than one race: 0 Unknown or Not Reported: 22	Hispanic or Latino: 13 Not Hispanic or Latino: 158 Unknown or Not Reported: 22	Mean: 52 Full Range: 19-76	76 Years and under
NCT01146834	Lymphoma	47	Male: 30 Female: 17	American Indian or Alaska Native: 0 Asian: 0 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 15 White: 27 More than one race: 0 Unknown or Not Reported: 5	Hispanic or Latino: 2 Not Hispanic or Latino: 44 Unknown or Not Reported: 1	<18: 0 18-65: 30 >65: 17	18 Years and above
NCT01728805	Lymphoma	372	Male: 216 Female: 156	White: 260 Other: 63 Not Reported: 49	Not reported	<18: 0 18-65: 188 >65: 184 Mean: 63 Full Range: 25-101	18 Years and above
NCT00019682	Melanoma	185	Male: 120 Female: 65	American Indian or Alaska Native: 0 Asian: 0	Not reported	Mean: 48.6 Full Range: 18-65	18 Years and above

				Native Hawaiian or Other Pacific Islander: 0 Black or African American: 0 White: 184 More than one race: 0 Unknown or Not Reported: 1			
NCT01936467	Pancreas	121	Male: 70 Female: 51	Not reported	Not reported	Mean: 64.15 Standard Deviation: 13.8	18 Years - 90 Years
NCT02340728	Pancreas	24	Male: 14 Female: 10	American Indian or Alaska Native: 0 Asian: 0 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 2 White: 20 More than one race: 0 Unknown or Not Reported: 2	Hispanic or Latino: 0 Not Hispanic or Latino: 24 Unknown or Not Reported: 0	Mean: 60.4 Full Range: 42-84	18 Years and above
NCT00002597	Prostate	1979	Male: 1979 Female: 0	Not reported	Not reported	Median: 71 Full Range: 47-91	18 Years and above
NCT00004124	Prostate	961	Male: 961 Female: 0	White: 810 Black: 116 Asian: 17 Other: 18	Hispanic: 55	Median: 60 Full Range: 40-86	18 Years - 120 Years
NCT00116142	Prostate	350	Male: 350 Female: 0	White: 271 Black: 9 Asian: 4 Other: 66	Not reported	Median: 66 Full Range: 43-86	30 Years and above
NCT00132301	Prostate	297	Male: 297 Female: 0	American Indian or Alaska Native: 0 Asian: 0 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 74 White: 207 More than one race: 0 Unknown or Not Reported: 16	Hispanic or Latino: 28 Not Hispanic or Latino: 269 Unknown or Not Reported: 0	<18: 0 18-65: 221 >65: 76 Mean: 62.27 Standard Deviation: 5.6	Any
NCT00134056	Prostate	994	Male: 994 Female: 0	White: 806 Black: 137 Asian: 20 Unknown: 24	Hispanic or Latino: 41 Not Hispanic or Latino: 953 Unknown or	Median: 69 Full Range: 40-92	18 Years and above

					Not Reported: 0		
NCT00142506	Prostate	290	Male: 290 Female: 0	Not reported	Not reported	<18: 0 18-65: 172 >65: 118	Any
NCT00329797	Prostate	96	Male: 96 Female: 0	Not reported	Not reported	Median: 71 Full Range: 51-87	18 Years and above
NCT00779402	Prostate	176	Male: 176 Female: 0	American Indian or Alaska Native: 0 Asian: 0 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 12 White: 162 More than one race: 0 Unknown or Not Reported: 2	Not reported	Mean: 64.7 Standard Deviation: 7.18	18 Years - 80 Years
NCT01238172	Prostate	443	Male: 443 Female: 0	American Indian or Alaska Native: 1 Asian: 15 Black or African American: 50 More than one race: 2 Native Hawaiian or Pacific Islander: 1 White: 357 Not Reported: 1	Hispanic or Latino: 16	Mean: 63.6 Standard Deviation: 6.5	50 Years - 80 Years
NCT01415960	Prostate	161	Male: 161 Female: 0	Not reported	Not reported	<18: 0 18-65: 42 >65: 119 Mean: 71 Standard Deviation: 9.02	18 Years and above
NCT01538628	Prostate	222	Male: 222 Female: 0	American Indian or Alaska Native: 0 Asian: 2 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 24 White: 188 More than one race: 2 Unknown or Not Reported: 6	Not reported	Mean: 66.39 Standard Deviation: 7.39	18 Years and above
NCT02260817	Prostate	109	Male: 109	American Indian or Alaska Native: 0	Hispanic or Latino: 0	<18: 0 18-65: 63	18 Years and above

			Female: 0	Asian: 2 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 8 White: 99 More than one race: 0 Unknown or Not Reported: 0	Not Hispanic or Latino: 109 Unknown or Not Reported: 0	>65: 46	
NCT02680041	Prostate	213	Male: 213 Female: 0	American Indian or Alaska Native: 1 Asian: 3 Native Hawaiian or Other Pacific Islander: 2 Black or African American: 17 White: 188 More than one race: 0 Unknown or Not Reported: 2	Hispanic or Latino: 9 Not Hispanic or Latino: 202 Unknown or Not Reported: 2	Median: 67 Full Range: 46-90	18 Years and above
NCT02712320	Prostate	30	Male: 30 Female: 0	American Indian or Alaska Native: 0 Asian: 1 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 3 White: 25 More than one race: 0 Unknown or Not Reported: 1	Not reported	Mean: 75 Standard Deviation: 7.86	18 Years and above
NCT02918357	Prostate	385	Male: 385 Female: 0	White: 318 Black or African American: 8 Native American or Alaska Native: 1 Asian: 15 Native Hawaiian: 0 Other Pacific Islander: 0 Other: 14 Unknown: 29	Not reported	Median: 70 Full Range: 45-95	18 Years and above
NCT02919111	Prostate	299	Male: 299 Female: 0	American Indian or Alaska Native: 3 Asian: 19 Native Hawaiian or Other Pacific Islander: 3 Black or African American: 8 White: 218 More than one race: 1	Hispanic or Latino: 11 Not Hispanic or Latino: 262 Unknown or Not Reported: 26	40-49: 3 50-59: 25 60-69: 126 70-79: 124 80-89: 19 90-99: 2	18 Years and above

				Unknown or Not Reported: 47			
NCT02981368	Prostate	385	Male: 385 Female: 0	White: 334 Black: 29 Asian: 11 Other: 5 Unknown or Not Reported: 6	Not reported	Median: 66 Full Range: 45-86	18 Years and above
NCT03353740	Prostate	346	Male: 346 Female: 0	American Indian or Alaska Native: 0 Asian: 20 Native Hawaiian or Other Pacific Islander: 2 Black or African American: 7 White: 276 More than one race: 1 Unknown or Not Reported: 40	Hispanic or Latino: 12 Not Hispanic or Latino: 323 Unknown or Not Reported: 11	40-49: 2 50-59: 22 60-69: 125 70-79: 157 80-89: 40	18 Years and above
NCT03404648	Prostate	19	Male: 19 Female: 0	Not reported	Not reported	Mean: 63.9 Standard Deviation: 6.0	18 Years and above
NCT03739684	Prostate	208	Male: 208 Female: 0	Asian: 3 Black or African American: 15 White: 188 Other, including not reported: 2	Not reported	Median: 68 Full Range: 43-91	18 Years and above
NCT03803475	Prostate	485	Male: 485 Female: 0	American Indian or Alaska Native: 0 Asian: 26 Native Hawaiian or Other Pacific Islander: 3 Black or African American: 8 White: 408 More than one race: 2 Unknown or Not Reported: 38	Hispanic or Latino: 16 Not Hispanic or Latino: 441 Unknown or Not Reported: 28	Mean: 70.13 Full Range: 48-92	18 Years and above
NCT00869206	Breast/Prostate	1822	Male: 980 Female: 842	Not reported	Not reported	Mean: 65.3 Standard Deviation: 11.9	18 Years and above
NCT02195232	Colorectal/Lung /Pancreas	64	Male: 34 Female: 23	American Indian or Alaska Native: 0 Asian: 2 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 9 White: 42 More than one race: 0 Unknown or Not Reported: 4	Not reported	<18: 0 18-65: 32 >65: 25	18 Years and above

NCT00003816	Leukemia/Lymphoma	361	Male: 222 Female: 139	White: 344 African American: 12 Asian: 3 Native American: 2	Hispanic: 3 Non-Hispanic: 358	Median: 44 Full Range: 4-68	4 Years - 70 Years
NCT00452439	Leukemia/Lymphoma	72	Male: 45 Female: 27	American Indian or Alaska Native: 0 Asian: 2 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 0 White: 70 More than one race: 0 Unknown or Not Reported: 0	Hispanic or Latino: 35 Not Hispanic or Latino: 37 Unknown or Not Reported: 0	Median: 36 Full Range: 23-53	18 Years and above
NCT00516503	Leukemia/Lymphoma	203	Male: 77 Female: 126	Not reported	Not reported	Median: 61 Full Range: 31-86	18 Years and above
NCT01231412	Leukemia/Lymphoma	174	Male: 117 Female: 57	Not reported	Not reported	<18: 0 18-65: 106 >65: 68 Median: 62.655 Full Range: 36.47-67.83	Any
NCT01295710	Leukemia/Lymphoma	254	Male: 139 Female: 115	American Indian or Alaska Native: 1 Asian: 3 Native Hawaiian or Other Pacific Islander: 0 Black or African American: 4 White: 234 More than one race: 1 Unknown or Not Reported: 11	Hispanic or Latino: 11 Not Hispanic or Latino: 236 Unknown or Not Reported: 7	18-40: 85 40-65: 169	18 Years - 65 Years
NCT02349412	Liver/Lung/Pancreas	391	Male: 221 Female: 170	American Indian or Alaska Native: 4 Asian: 15 Native Hawaiian or Other Pacific Islander: 2 Black or African American: 46 White: 303 More than one race: 0 Unknown or Not Reported: 21	Not reported	Mean: 65.2 Full Range: 34-97	18 Years and above
NCT01438476	Liver/Pancreas	140	Male: 78 Female: 62	American Indian or Alaska Native: 0 Asian: 2	Hispanic or Latino: 10 Not Hispanic	<18: 0 18-65: 112 >65: 28	18 Years and above

				Native Hawaiian or Other Pacific Islander: 0 Black or African American: 14 White: 119 More than one race: 5 Unknown or Not Reported: 0	or Latino: 130 Unknown or Not Reported: 0	Median: 56.8 Full Range: 48.9-63.8	
NCT00377156	Lung/Breast/Prostate	213	Male: 109 Female: 104	Not reported	Not reported	Mean: 60.6 Standard Deviation: 10.5 18-59: 97 >60: 116	18 Years and above

Appendix D: Real-world incidence data for U.S. cancers of each cancer type

Type of Cancer	Sex Data	Total Sex Count	Race/Ethnicity Data	Total Race or Ethnicity Count	Age Data (mean)	Total Age Count
Female Breast Cancer	Female: 254744	254744	White: 207932 Black: 29661 American Indian and Alaska Native: 1366 Asian and Pacific Islander: 11861 Hispanic: 22560	250820	62.88	254743
Colon and Rectum Cancer	Male: 74564 Female: 66510	141074	White: 114092 Black: 17473 American Indian and Alaska Native: 984 Asian and Pacific Islander: 5825 Hispanic: 13707	138374	66.03	141071
Kidney and Renal Pelvis Cancer	Male: 41721 Female: 24038	65759	White: 54511 Black: 7917 American Indian and Alaska Native: 660 Asian and Pacific Islander: 1648 Hispanic: 7399	64736	63.59	65759
Leukemias	Male: 29388 Female: 20786	50174	White: 42480 Black: 4252 American Indian and Alaska Native: 292 Asian and Pacific Islander: 1596 Hispanic: 5068	48620	62.37	50174
Liver and Intrahepatic Bile Duct Cancer	Male: 24539 Female: 10099	34638	White: 26795 Black: 4687 American Indian and Alaska Native: 394 Asian and Pacific Islander: 2298 Hispanic: 5501	34174	66.52	34638
Lung and Bronchus	Male: 111009	218520	White: 185643	217003	69.95	218493

Cancer	Female: 107511		Black: 23417 American Indian and Alaska Native: 1353 Asian and Pacific Islander: 6590 Hispanic: 10183			
Lymphoma	Male: 43550 Female: 35840	79390	White: 67253 Black: 6959 American Indian and Alaska Native: 418 Asian and Pacific Islander: 2901 Hispanic: 8259	77531	63.39	79379
Melanomas of the Skin	Male: 49547 Female: 34449	83996	White: 78430 Black: 354 American Indian and Alaska Native: 201 Asian and Pacific Islander:260 Hispanic: 1819	79245	64.18	83982
Pancreas Cancer	Male: 27285 Female: 25261	52546	White: 43006 Black: 6836 American Indian and Alaska Native: 293 Asian and Pacific Islander: 1933 Hispanic: 4544	52068	69.56	52529
Prostate Cancer	Male: 211893	211893	White: 163503 Black: 33589 American Indian and Alaska Native: 853 Asian and Pacific Islander: 5188 Hispanic: 14801	203133	67.21	211865

The following table lists the sex proportions and representation scores for each clinical trial:

Appendix E: Sex analysis between clinical trial and real-world

Cancer Type	Trial Identifier	Proportion (trial)	Proportion (world)	Representation Score	Mean (95% CI)	p-values from t-test
Colorectal Cancer	NCT00265850	Male: 0.613 Female: 0.387	Male: 0.529 Female: 0.471	Male: 1.16 Female: 0.82	Male: 1.11 (0.87 - 1.35) Female: 0.84 (0.57 - 1.10)	Male: 0.29 Female: 0.18
	NCT01099449	Male: 0.479 Female: 0.521		Male: 0.91 Female: 1.07		
	NCT01931150	Male: 0.818 Female: 0.182		Male: 1.55 Female: 0.39		
	NCT02195232	Male: 0.531 Female: 0.359		Male: 1.01 Female: 0.76		
	NCT02254486	Male: 0.527		Male: 1.00		

		Female: 0.473		Female: 1.01		
	NCT02776683	Male: 0.553 Female:0.447		Male: 1.05 Female: 0.95		
Kidney Cancer	NCT00087022	Male: 0.664 Female: 0.336	Male: 0.634 Female: 0.366	Male: 1.05 Female: 0.92	Male: 1.02 (0.71 - 1.33) Female: 0.96 (0.42 - 1.50)	Not enough trials for statistical analysis
	NCT01606787	Male: 0.633 Female: 0.367		Male: 1.00 Female: 1.00		
Leukemia	NCT00003816	Male: 0.615 Female: 0.385	Male: 0.586 Female: 0.414	Male: 1.05 Female: 0.93	Male: 0.93 (0.81 - 1.05) Female: 1.10 (0.93 - 1.27)	Male: 0.23 Female: 0.23
	NCT00452439	Male: 0.625 Female: 0.375		Male: 1.07 Female: 0.91		
	NCT00469144	Male: 0.507 Female: 0.493		Male: 0.87 Female: 1.19		
	NCT00516503	Male: 0.379 Female: 0.621		Male: 0.65 Female: 1.50		
	NCT0887068	Male: 0.597 Female: 0.403		Male: 1.02 Female: 0.97		
	NCT01231412	Male: 0.672 Female: 0.328		Male: 1.15 Female: 0.79		
	NCT01295710	Male: 0.547 Female: 0.453		Male: 0.93 Female: 1.09		
	NCT01802333	Male: 0.512 Female: 0.488		Male: 0.87 Female: 1.18		
	NCT02801578	Male: 0.455 Female: 0.545		Male: 0.78 Female: 1.32		
Liver Cancer	NCT00788697	Male: 0.513 Female: 0.487	Male: 0.708 Female: 0.292	Male: 0.72 Female: 1.67	Male: 0.74 (0.65 - 0.82) Female: 1.64 (1.44 - 1.85)	Male: 0.0009 Female: 0.0009
	NCT00829413	Male: 0.525 Female: 0.475		Male: 0.74 Female: 1.63		
	NCT01438476	Male: 0.557 Female: 0.443		Male: 0.79 Female: 1.52		
	NCT01596283	Male: 0.444 Female: 0.556		Male: 0.63 Female: 1.91		
	NCT02349412	Male: 0.565 Female: 0.435		Male: 0.80 Female: 1.49		
Lung Cancer	NCT00377156	Male: 0.512 Female: 0.488	Male: 0.508 Female: 0.492	Male: 1.01 Female: 0.99	Male: 1.18 (1.07 - 1.29) Female: 0.79 (0.69 -	Male: 0.004 Female: 0.002

	NCT02195232	Male: 0.531 Female: 0.359		Male: 1.05 Female: 0.73	0.90)	
	NCT02349412	Male: 0.565 Female: 0.435		Male: 1.11 Female: 0.88		
	NCT00003901	Male: 0.514 Female: 0.486		Male: 1.01 Female: 0.99		
	NCT00153803	Male: 0.592 Female: 0.408		Male: 1.17 Female: 0.83		
	NCT00693992	Male: 0.557 Female: 0.443		Male: 1.10 Female: 0.90		
	NCT01355497	Male: 0.709 Female: 0.291		Male: 1.40 Female: 0.59		
	NCT02027428	Male: 0.644 Female: 0.356		Male: 1.27 Female: 0.72		
	NCT02785939	Male: 0.659 Female: 0.341		Male: 1.30 Female: 0.69		
	NCT02965378	Male: 0.706 Female: 0.294		Male: 1.39 Female: 0.60		
Lymphoma	NCT00003816	Male: 0.615 Female: 0.385	Male: 0.549 Female: 0.451	Male: 1.12 Female: 0.85	Male: 1.05 (0.92 - 1.18) Female: 0.94 (0.78 - 1.09)	Male: 0.39 Female: 0.38
	NCT00452439	Male: 0.625 Female: 0.375		Male: 1.14 Female: 0.83		
	NCT00516503	Male: 0.379 Female: 0.621		Male: 0.69 Female: 1.37		
	NCT01231412	Male: 0.672 Female: 0.328		Male: 1.23 Female: 0.73		
	NCT01295710	Male: 0.547 Female: 0.453		Male: 1.00 Female: 1.00		
	NCT00118209	Male: 0.540 Female: 0.458		Male: 0.98 Female: 1.02		
	NCT00566228	Male: 0.702 Female: 0.298		Male: 1.28 Female: 0.65		
	NCT00577993	Male: 0.466 Female: 0.534		Male: 0.85 Female: 1.18		
	NCT01146834	Male: 0.638 Female: 0.362		Male: 1.16 Female: 0.80		
	NCT01728805	Male: 0.581		Male: 1.06		

		Female: 0.419		Female: 0.93		
Melanoma	NCT00019682	Male: 0.649 Female: 0.351	Male: 0.590 Female: 0.410	Male: 1.10 Female: 0.86	Male: 1.10 (1.10 - 1.10) Female: 0.86 (0.86 - 0.86)	Not enough trials to run statistical analysis
Pancreas Cancer	NCT02195232	Male: 0.531 Female: 0.359	Male: 0.519 Female: 0.481	Male: 1.02 Female: 0.75	Male: 1.08 (1.04 - 1.13) Female: 0.86 (0.78 - 0.95)	Male: 0.009 Female: 0.01
	NCT01438476	Male: 0.557 Female: 0.443		Male: 1.07 Female: 0.92		
	NCT02349412	Male: 0.565 Female: 0.435		Male: 1.09 Female: 0.90		
	NCT01936467	Male: 0.579 Female: 0.421		Male: 1.11 Female: 0.88		
	NCT02340728	Male: 0.583 Female: 0.417		Male: 1.12 Female: 0.87		
All Cancers	**n/a**	**n/a**	**n/a**	**Male: 1.05** **Female: 0.98**	Male: 1.04 (0.98 - 1.11) Female: 0.98 (0.88 - 1.08)	**Male: 0.16** **Female: 0.65**

The following table lists the race/ethnicity proportions and representation scores for each clinical trial:

Appendix F: Race and ethnicity analysis between clinical trial and real-world

Cancer Type	Trial Identifier	Proportion (trial)	Proportion (world)	Representation Score	p-values from t-test
Breast Cancer	NCT00014222	White: 0.887 Black: 0.047 American Indian and Alaska Native: 0.010 Asian and Pacific Islander: 0.036 Hispanic: N/A	White: 0.829 Black: 0.118 American Indian and Alaska Native: 0.005 Asian and Pacific Islander: 0.047 Hispanic: 0.090	White: 1.07 Black: 0.39 American Indian and Alaska Native: 1.75 Asian and Pacific Islander: 0.76 Hispanic: N/A	White: 0.79 Black: 0.21 American Indian and Alaska Native: 0.03 Asian and Pacific Islander: 0.53 Hispanic: 0.85
	NCT00382018	White: 0.832 Black: 0.168 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: N/A		White: 1.00 Black: 1.42 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: N/A	
	NCT00775645	White: 0.795 Black: 0.090 American Indian and Alaska Native: 0.002		White: 0.96 Black: 0.76 American Indian and Alaska	

		Asian and Pacific Islander: 0.064 Hispanic: 0.088		Native: 0.45 Asian and Pacific Islander: 1.34 Hispanic: 0.98	
	NCT00789581	White: 0.795 Black: 0.184 American Indian and Alaska Native: 0.003 Asian and Pacific Islander: 0.007 Hispanic: N/A		White: 0.96 Black: 1.56 American Indian and Alaska Native: 0.60 Asian and Pacific Islander: 0.14 Hispanic: N/A	
	NCT01224678	White: 0.793 Black: 0.117 American Indian and Alaska Native: 0.003 Asian and Pacific Islander: 0.053 Hispanic: 0.140		White: 0.96 Black: 0.99 American Indian and Alaska Native: 0.61 Asian and Pacific Islander: 1.13 Hispanic: 1.56	
	NCT01385137	White: 0.871 Black: 0.080 American Indian and Alaska Native: 0.004 Asian and Pacific Islander: 0.016 Hispanic: 0.064		White: 1.05 Black: 0.68 American Indian and Alaska Native: 0.74 Asian and Pacific Islander: 0.34 Hispanic: 0.71	
	NCT01573442	White: 0.933 Black: 0.029 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.024 Hispanic: N/A		White: 1.13 Black: 0.24 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.51 Hispanic: N/A	
	NCT01598298	White: 0.858 Black: 0.093 American Indian and Alaska Native: 0.007 Asian and Pacific Islander: 0.035 Hispanic: 0.038		White: 1.04 Black: 0.79 American Indian and Alaska Native: 1.27 Asian and Pacific Islander: 0.73 Hispanic: 0.42	
	NCT01856543	White: 0.664 Black: 0.175 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.091 Hispanic: N/A		White: 0.80 Black: 1.48 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 1.92 Hispanic: N/A	
	NCT01945775	White: 0.691 Black: 0.028		White: 0.83 Black: 0.24	

		American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.109 Hispanic: 0.107		American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 2.31 Hispanic: 1.19	
	NCT02574455	White: 0.790 Black: 0.117 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.042 Hispanic: 0.085		White: 0.95 Black: 0.99 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.88 Hispanic: 0.95	
	NCT02961790	White: 0.956 Black: 0.027 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.018 Hispanic: N/A		White: 1.15 Black: 0.22 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.37 Hispanic: N/A	
Colorectal Cancer	NCT02195232	White: 0.656 Black: 0.141 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.031 Hispanic: N/A	White: 0.829 Black: 0.126 American Indian and Alaska Native: 0.007 Asian and Pacific Islander: 0.042 Hispanic: 0.099	White: 0.80 Black: 1.11 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.74 Hispanic: N/A	Not enough trials to perform statistical analysis
	NCT02776683	White: 0.667 Black: 0.049 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.268 Hispanic: 0.122		White: 0.81 Black: 0.39 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 6.37 Hispanic: 1.23	
Kidney Cancer	NCT01606787	White: 0.809 Black: 0.075 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.050 Hispanic: N/A	White: 0.842 Black: 0.122 American Indian and Alaska Native: 0.010 Asian and Pacific Islander: 0.025 Hispanic: 0.114	White: 0.96 Black: 0.62 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 1.97 Hispanic: N/A	Not enough trials to perform statistical analysis
Leukemia	NCT00003816	White: 0.953 Black: 0.033 American Indian and Alaska Native: 0.006 Asian and Pacific Islander: 0.008 Hispanic: 0.008	White: 0.874 Black: 0.087 American Indian and Alaska Native: 0.006 Asian and Pacific Islander: 0.033 Hispanic: 0.104	White: 1.09 Black: 0.38 American Indian and Alaska Native: 0.92 Asian and Pacific Islander: 0.25 Hispanic: 0.08	White: 0.045 Black: 0.002 American Indian and Alaska Native: 0.08 Asian and Pacific Islander: 0.01 Hispanic: 0.9

	NCT00452439	White: 0.972 Black: 0.000 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.028 Hispanic: 0.486		White: 1.11 Black: 0.00 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.85 Hispanic: 4.66	
	NCT00887068	White: 0.945 Black: 0.011 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: N/A		White: 1.08 Black: 0.13 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: N/A	
	NCT01295710	White: 0.921 Black: 0.016 American Indian and Alaska Native: 0.004 Asian and Pacific Islander: 0.012 Hispanic: 0.043		White: 1.05 Black: 0.18 American Indian and Alaska Native: 0.66 Asian and Pacific Islander: 0.36 Hispanic: 0.42	
	NCT01802333	White: 0.831 Black: 0.075 American Indian and Alaska Native: 0.008 Asian and Pacific Islander: 0.027 Hispanic: 0.084		White: 0.95 Black: 0.85 American Indian and Alaska Native: 1.35 Asian and Pacific Islander: 0.83 Hispanic: 0.81	
	NCT02801578	White: 1.000 Black: 0.000 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.000 Hispanic: N/A		White: 1.14 Black: 0.00 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: N/A	
Liver Cancer	NCT00788697	White: 0.671 Black: 0.133 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.063 Hispanic: N/A	White: 0.784 Black: 0.137 American Indian and Alaska Native: 0.012 Asian and Pacific Islander: 0.067 Hispanic: 0.161	White: 0.86 Black: 0.97 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.93 Hispanic: N/A	White: 0.8 Black: 0.07 American Indian and Alaska Native: N/A* Asian and Pacific Islander: 0.08 Hispanic: N/A* *: There were not enough studies reporting these measures to perform statistical analysis on these
	NCT00829413	White: 0.795 Black: 0.085 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.046 Hispanic: N/A		White: 1.01 Black: 0.62 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.69 Hispanic: N/A	

	NCT01438476	White: 0.850 Black: 0.100 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.014 Hispanic: 0.071		White: 1.08 Black: 0.73 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.21 Hispanic: 0.44	demographics
	NCT02349412	White: 0.775 Black: 0.118 American Indian and Alaska Native: 0.010 Asian and Pacific Islander: 0.043 Hispanic: N/A		White: 0.99 Black: 0.86 American Indian and Alaska Native: 0.83 Asian and Pacific Islander: 0.64 Hispanic: N/A	
Lung Cancer	NCT02195232	White: 0.656 Black: 0.141 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.031 Hispanic: N/A	White: 0.855 Black: 0.108 American Indian and Alaska Native: 0.006 Asian and Pacific Islander: 0.030 Hispanic: 0.047	White: 0.77 Black: 1.30 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 1.03 Hispanic: N/A	White: 0.7 Black: 0.4 American Indian and Alaska Native: 0.7 Asian and Pacific Islander: 0.04 Hispanic: 0.2
	NCT02349412	White: 0.775 Black: 0.118 American Indian and Alaska Native: 0.010 Asian and Pacific Islander: 0.043 Hispanic: N/A		White: 0.91 Black: 1.09 American Indian and Alaska Native: 1.67 Asian and Pacific Islander: 1.43 Hispanic: N/A	
	NCT00003901	White: 0.916 Black: 0.057 American Indian and Alaska Native: 0.001 Asian and Pacific Islander: 0.013 Hispanic: 0.011		White: 1.07 Black: 0.53 American Indian and Alaska Native: 0.15 Asian and Pacific Islander: 0.44 Hispanic: 0.22	
	NCT00153803	White: 0.878 Black: 0.053 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.012 Hispanic: N/A		White: 1.03 Black: 0.49 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.40 Hispanic: N/A	
	NCT00693992	White: 0.819 Black: 0.148 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.014 Hispanic: N/A		White: 0.96 Black: 1.37 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.47 Hispanic: N/A	

	NCT02027428	White: 0.926 Black: 0.054 American Indian and Alaska Native: 0.005 Asian and Pacific Islander: 0.005 Hispanic: 0.040		White: 1.08 Black: 0.50 American Indian and Alaska Native: 0.79 Asian and Pacific Islander: 0.16 Hispanic: 0.84	
	NCT02785939	White: 0.878 Black: 0.073 American Indian and Alaska Native: 0.024 Asian and Pacific Islander: 0.024 Hispanic: 0.000		White: 1.03 Black: 0.68 American Indian and Alaska Native: 3.91 Asian and Pacific Islander: 0.80 Hispanic: 0.00	
	NCT02965378	White: 0.882 Black: 0.118 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.000 Hispanic: 0.059		White: 1.03 Black: 1.09 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: 1.25	
Lymphoma	NCT00003816	White: 0.953 Black: 0.033 American Indian and Alaska Native: 0.006 Asian and Pacific Islander: 0.008 Hispanic: 0.008	White: 0.867 Black: 0.090 American Indian and Alaska Native: 0.005 Asian and Pacific Islander: 0.037 Hispanic: 0.107	White: 1.10 Black: 0.37 American Indian and Alaska Native: 1.03 Asian and Pacific Islander: 0.22 Hispanic: 0.08	White: 0.4 Black: 0.6 American Indian and Alaska Native: 0.1 Asian and Pacific Islander: 0.01 Hispanic: 0.4
	NCT00452439	White: 0.972 Black: 0.000 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.028 Hispanic: 0.486		White: 1.12 Black: 0.00 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.74 Hispanic: 4.56	
	NCT01295710	White: 0.921 Black: 0.016 American Indian and Alaska Native: 0.004 Asian and Pacific Islander: 0.012 Hispanic: 0.043		White: 1.06 Black: 0.18 American Indian and Alaska Native: 0.73 Asian and Pacific Islander: 0.32 Hispanic: 0.41	
	NCT00118209	White: 0.784 Black: 0.122 American Indian and Alaska Native: 0.008 Asian and Pacific Islander: 0.037 Hispanic: 0.063		White: 0.90 Black: 1.36 American Indian and Alaska Native: 1.51 Asian and Pacific Islander: 0.98 Hispanic: 0.59	

	NCT00577993	White: 0.819 Black: 0.057 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.010 Hispanic: 0.067		White: 0.94 Black: 0.63 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.28 Hispanic: 0.63	
	NCT01146834	White: 0.574 Black: 0.319 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.000 Hispanic: 0.043		White: 0.66 Black: 3.56 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: 0.40	
	NCT01728805	White: 0.699 Black: N/A American Indian and Alaska Native: N/A Asian and Pacific Islander: N/A Hispanic: N/A		White: 0.81 Black: N/A American Indian and Alaska Native: N/A Asian and Pacific Islander: N/A Hispanic: N/A	
Melanoma	NCT00019682	White: 0.995 Black: 0.000 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.000 Hispanic: N/A	White: 0.990 Black: 0.004 American Indian and Alaska Native: 0.003 Asian and Pacific Islander: 0.003 Hispanic: 0.023	White: 1.00 Black: 0.00 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: N/A	Not enough trials for statistical analysis
Pancreas Cancer	NCT02195232	White: 0.656 Black: 0.141 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.031 Hispanic: N/A	White: 0.826 Black: 0.131 American Indian and Alaska Native: 0.006 Asian and Pacific Islander: 0.037 Hispanic: 0.087	White: 0.795 Black: 1.07 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.84 Hispanic: N/A	White: 0.4 Black: 0.2 American Indian and Alaska Native: 0.3 Asian and Pacific Islander: 0.2 Hispanic: N/A
	NCT01438476	White: 0.850 Black: 0.100 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.014 Hispanic: 0.071		White: 1.03 Black: 0.76 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.38 Hispanic: 0.82	
	NCT02349412	White: 0.775 Black: 0.118 American Indian and Alaska Native: 0.010 Asian and Pacific Islander: 0.043 Hispanic: N/A		White: 0.94 Black: 0.90 American Indian and Alaska Native: 1.67 Asian and Pacific Islander: 1.16	

				Hispanic: N/A	
	NCT02340728	White: 0.833 Black: 0.083 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.000 Hispanic: 0.000		White: 1.01 Black: 0.63 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: 0.00	
Prostate Cancer	NCT00004124	White: 0.843 Black: 0.121 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.018 Hispanic: 0.057	White: 0.805 Black: 0.165 American Indian and Alaska Native: 0.004 Asian and Pacific Islander: 0.026 Hispanic: 0.073	White: 1.05 Black: 0.73 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.69 Hispanic: 0.79	White: 0.1 Black: 0.0004 American Indian and Alaska Native: 0.06 Asian and Pacific Islander: 0.7 Hispanic: 0.006
	NCT00116142	White: 0.774 Black: 0.026 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.011 Hispanic: N/A		White: 0.96 Black: 0.16 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.45 Hispanic: N/A	
	NCT00132301	White: 0.697 Black: 0.249 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.000 Hispanic: 0.094		White: 0.87 Black: 1.51 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: 1.29	
	NCT00134056	White: 0.811 Black: 0.138 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.020 Hispanic: 0.041		White: 1.01 Black: 0.83 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.79 Hispanic: 0.57	
	NCT00779402	White: 0.920 Black: 0.068 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.000 Hispanic: N/A		White: 1.14 Black: 0.41 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.00 Hispanic: N/A	
	NCT01238172	White: 0.806 Black: 0.113 American Indian and Alaska Native: 0.002 Asian and Pacific Islander: 0.036		White: 1.00 Black: 0.68 American Indian and Alaska Native: 0.54	

		Hispanic: 0.036		Asian and Pacific Islander: 1.41 Hispanic: 0.50	
	NCT01538628	White: 0.847 Black: 0.108 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.009 Hispanic: N/A		White: 1.05 Black: 0.65 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.35 Hispanic: N/A	
	NCT02260817	White: 0.908 Black: 0.073 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.018 Hispanic: 0.000		White: 1.13 Black: 0.44 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 0.72 Hispanic: 0.00	
	NCT02680041	White: 0.883 Black: 0.080 American Indian and Alaska Native: 0.005 Asian and Pacific Islander: 0.023 Hispanic: 0.042		White: 1.10 Black: 0.48 American Indian and Alaska Native: 1.12 Asian and Pacific Islander: 0.92 Hispanic: 0.58	
	NCT02712320	White: 0.833 Black: 0.100 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.033 Hispanic: N/A		White: 1.04 Black: 0.60 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 1.31 Hispanic: N/A	
	NCT02918357	White: 0.826 Black: 0.021 American Indian and Alaska Native: 0.003 Asian and Pacific Islander: 0.039 Hispanic: N/A		White: 1.03 Black: 0.13 American Indian and Alaska Native: 0.62 Asian and Pacific Islander: 1.53 Hispanic: N/A	
	NCT02919111	White: 0.729 Black: 0.027 American Indian and Alaska Native: 0.010 Asian and Pacific Islander: 0.074 Hispanic: 0.037		White: 0.91 Black: 0.16 American Indian and Alaska Native: 2.39 Asian and Pacific Islander: 2.88 Hispanic: 0.50	
	NCT02981368	White: 0.867 Black: 0.075 American Indian and Alaska Native: N/A		White: 1.08 Black: 0.46 American Indian and Alaska	

		Asian and Pacific Islander: 0.029 Hispanic: N/A		Native: N/A Asian and Pacific Islander: 1.12 Hispanic: N/A	
	NCT03353740	White: 0.798 Black: 0.020 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.064 Hispanic: 0.035		White: 0.99 Black: 0.12 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 2.49 Hispanic: 0.48	
	NCT03739684	White: 0.904 Black: 0.072 American Indian and Alaska Native: N/A Asian and Pacific Islander: 0.014 Hispanic: N/A		White: 1.12 Black: 0.44 American Indian and Alaska Native: N/A Asian and Pacific Islander: 2.49 Hispanic: 0.48	
	NCT03803475	White: 0.841 Black: 0.016 American Indian and Alaska Native: 0.000 Asian and Pacific Islander: 0.060 Hispanic: 0.033		White: 1.05 Black: 0.10 American Indian and Alaska Native: 0.00 Asian and Pacific Islander: 2.34 Hispanic: 0.45	
All Cancers	**n/a**	**n/a**	**n/a**	White: 0.99 Black: 0.67 American Indian and Alaska Native: 0.45 Asian and Pacific Islander: 0.87 Hispanic: 0.78	White: 0.6 Black: 0.0001 American Indian and Alaska Native: 3e-05 Asian and Pacific Islander: 0.4 Hispanic: 0.2

The following table lists the mean age (world), mean age (trial), and mean age difference for each clinical trial:

Appendix G: Age analysis between clinical trial and real-world

Cancer Type	Trial Identifier	Mean Age (trial)	Mean Age (world)	Mean Age Difference	p-values from t-test
Breast Cancer	NCT00003830	56	62.88	-6.88	6e-05
	NCT00009945	54		-8.88	
	NCT00041119	53.4		-9.48	
	NCT00053898	61		-1.88	

	NCT00093795	51		-11.88	
	NCT00376597	57.7		-5.18	
	NCT00377156	60.6		-2.28	
	NCT00869206	65.3		2.42	
	NCT01376349	57.3		-5.58	
	NCT01573442	60		-5.88	
	NCT01591746	49.1		-13.78	
	NCT01856543	49		-13.88	
	NCT01945775	48.1		-14.78	
	NCT02574455	54		-8.88	
	NCT02961790	57.1		-5.78	
Colorectal Cancer	NCT02254486	57.5	66.03	-8.53	Not enough trials for statistical analysis
	NCT02776683	58		-8.03	
Kidney Cancer	NCT00087022	58	63.59	-5.59	Not enough trials for statistical analysis
Liver Cancer	NCT00829413	56.9	66.52	-9.62	Not enough trials for statistical analysis
	NCT01596283	57		-9.52	
Lung Cancer	NCT00377156	60.6	69.95	-9.35	Not enough trials for statistical analysis
	NCT02027428	67		-2.95	
Lymphoma	NCT00577993	52	63.39	-11.39	Not enough trials for statistical analysis
	NCT01728805	63		-0.39	
Melanoma	NCT00019682	48.6	64.18	-15.58	Not enough trials for statistical analysis
Pancreas Cancer	NCT01936467	64.15	69.56	-5.41	Not enough trials for statistical analysis
	NCT02340728	60.40		-9.16	
Prostate Cancer	NCT00377156	60.60	67.21	-6.61	0.5
	NCT00869206	65.30		-1.91	
	NCT00132301	62.27		-4.94	
	NCT00779402	64.70		-2.51	
	NCT01238172	63.60		-3.61	
	NCT01415960	71.00		3.79	
	NCT01538628	66.39		-0.82	

	NCT02712320	75.00		7.79	
	NCT03404648	63.90		-3.31	
	NCT03803475	70.13		2.92	
All Cancers	**n/a**	**n/a**	**n/a**	**-5.9**	**3e-07**

Appendix H: Summary of data extracted for secondary outcome analysis

Trial Identifier	Cancer Type, Primary Location	Representation Score (Sex)	Representation Score (Race/Ethnicity)	Mean Age Difference	Funding Status	Location Type	Number of Participants
NCT00003830	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-6.88	Other	Multi-centre	5611
NCT00009945	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-8.88	Other	Multi-centre	3323
NCT00014222	Breast	Male: N/A Female: N/A	White: 1.07 Black: 0.39 American Indian or Alaska Native: 1.75 Asian or Pacific Islander: 0.76 Hispanic: N/A	N/A	Other	Multi-centre	2103
NCT00041119	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-9.48	Other	Multi-centre	3871
NCT00053898	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-1.88	Other	Multi-centre	3104
NCT00075764	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-centre	694
NCT00093795	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-11.88	Other	Multi-centre	4867
NCT00195013	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A	N/A	Other	Single-site	30

			Asian or Pacific Islander: N/A Hispanic: N/A				
NCT00265759	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-centre	610
NCT00296036	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-centre	127
NCT00376597	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-5.18	Other	Multi-centre	554
NCT00382018	Breast	Male: N/A Female: N/A	White: 1.00 Black: 1.42 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.00 Hispanic: N/A	N/A	Other	Multi-centre	564
NCT00775645	Breast	Male: N/A Female: N/A	White: 0.96 Black: 0.76 American Indian or Alaska Native: 0.45 Asian or Pacific Islander: 1.34 Hispanic: 0.98	N/A	Other	Multi-centre	409
NCT00789581	Breast	Male: N/A Female: N/A	White: 0.96 Black: 1.56 American Indian or Alaska Native: 0.60 Asian or Pacific Islander: 0.14 Hispanic: N/A	N/A	Other	Multi-centre	614
NCT00956813	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-centre	210
NCT01224678	Breast	Male: N/A Female: N/A	White: 0.96 Black: 0.99 American Indian or Alaska Native: 0.61 Asian or Pacific Islander: 1.13	N/A	Other	Multi-centre	300

			Hispanic: 1.56				
NCT01376349	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-5.58	Other	Multi-centre	443
NCT01385137	Breast	Male: N/A Female: N/A	White: 1.05 Black: 0.68 American Indian or Alaska Native: 0.74 Asian or Pacific Islander: 0.34 Hispanic: 0.71	N/A	Other	Multi-centre	249
NCT01573442	Breast	Male: N/A Female: N/A	White: 1.13 Black: 0.24 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.51 Hispanic: N/A	-2.88	Other	Multi-centre	208
NCT01591746	Breast	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-13.78	Other	Single-site	131
NCT01598298	Breast	Male: N/A Female: N/A	White: 1.04 Black: 0.79 American Indian or Alaska Native: 1.27 Asian or Pacific Islander: 0.73 Hispanic: 0.42	N/A	Other	Multi-centre	289
NCT01856543	Breast	Male: N/A Female: N/A	White: 0.80 Black: 1.48 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 1.92 Hispanic: N/A	-13.88	Other	Multi-centre	143
NCT01945775	Breast	Male: N/A Female: N/A	White: 0.83 Black: 0.24 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 2.31 Hispanic: 1.19	-14.78	Industry	Multi-centre	431
NCT02574455	Breast	Male: N/A Female: N/A	White: 0.95 Black: 0.99 American Indian or Alaska Native: N/A Asian or Pacific Islander: 0.88 Hispanic: 0.95	-8.88	Industry	Multi-centre	529

NCT02961790	Breast	Male: N/A Female: N/A	White: 1.15 Black: 0.22 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.37 Hispanic: N/A	-5.78	Other	Multi-centre	113
NCT00265850	Colorectal	Male: 1.16 Female: 0.82	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-centre	1137
NCT01099449	Colorectal	Male: 0.91 Female: 1.11	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-centre	353
NCT01931150	Colorectal	Male: 1.55 Female: 0.39	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Single-site	11
NCT02254486	Colorectal	Male: 1.00 Female: 1.00	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-8.53	Industry	Multi-centre	621
NCT02776683	Colorectal	Male: 1.05 Female: 0.95	White: 0.81 Black: 0.39 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 6.37 Hispanic: 1.23	-8.03	Industry	Multi-centre	123
NCT00087022	Kidney	Male: 1.05 Female: 0.92	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-5.59	Industry	Multi-centre	864
NCT01606787	Kidney	Male: 1.00 Female: 1.00	White: 0.96 Black: 0.62 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 1.97 Hispanic: N/A	N/A	Other	Single-site	199
NCT00469144	Leukemia	Male: 0.87	White: N/A	N/A	Other	Single-site	225

		Female: 1.19	Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A				
NCT00887068	Leukemia	Male: 1.02 Female: 0.97	White: 1.08 Black: 0.13 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.00 Hispanic: N/A	N/A	Other	Single-site	181
NCT01802333	Leukemia	Male: 0.87 Female: 1.18	White: 0.95 Black: 0.18 American Indian or Alaska Native: 1.35 Asian or Pacific Islander: 0.83 Hispanic: 0.81	N/A	NIH	Multi-center	738
NCT02801578	Leukemia	Male: 0.78 Female: 1.32	White: 1.14 Black: 0.00 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.00 Hispanic: N/A	N/A	Other	Single-site	11
NCT00788697	Liver	Male: 0.72 Female: 1.67	White: 0.86 Black: 0.97 American Indian or Alaska Native: N/A Asian or Pacific Islander: 0.93 Hispanic: N/A	N/A	Industry	Single-site	240
NCT00829413	Liver	Male: 0.74 Female: 1.63	White: 1.01 Black: 0.62 American Indian or Alaska Native: N/A Asian or Pacific Islander: 0.69 Hispanic: N/A	-9.62	Industry	Single-site	259
NCT01596283	Liver	Male: 0.63 Female: 1.91	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-9.52	Other	Single-site	135
NCT00003901	Lung	Male: 1.01 Female: 0.99	White: 1.07 Black: 0.53 American Indian or Alaska Native: 0.15 Asian or Pacific Islander: 0.44 Hispanic: 0.22	N/A	Other	Multi-center	1047
NCT00153803	Lung	Male: 1.17 Female: 0.83	White: 1.03 Black: 0.49	N/A	Other	Multi-center	245

			American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.40 Hispanic: N/A				
NCT00693992	Lung	Male: 1.10 Female: 0.90	White: 0.96 Black: 1.37 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.47 Hispanic: N/A	N/A	NIH	Multi-center	210
NCT01355497	Lung	Male: 1.40 Female: 0.59	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Industry	Multi-center	320
NCT02027428	Lung	Male: 1.27 Female: 0.72	White: 1.08 Black: 0.50 American Indian or Alaska Native: 0.79 Asian or Pacific Islander: 0.16 Hispanic: 0.84	-2.95	Industry	Multi-center	202
NCT02785939	Lung	Male: 1.30 Female: 0.69	White: 1.03 Black: 0.68 American Indian or Alaska Native: 3.91 Asian or Pacific Islander: 0.80 Hispanic: 0.00	N/A	Other	Multi-center	41
NCT02965378	Lung	Male: 1.39 Female: 0.60	White: 1.03 Black: 1.09 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.00 Hispanic: 1.25	N/A	Other	Multi-center	34
NCT00118209	Lymphoma	Male: 0.98 Female: 1.02	White: 0.90 Black: 1.36 American Indian or Alaska Native: 1.51 Asian or Pacific Islander: 0.98 Hispanic: 0.59	N/A	Other	Multi-center	491
NCT00566228	Lymphoma	Male: 1.28 Female: 0.66	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Single-site	121
NCT00577993	Lymphoma	Male: 0.85 Female: 1.18	White: 0.94 Black: 0.63 American Indian or Alaska Native: 0.00	-11.39	Other	Single-site	193

			Asian or Pacific Islander: 0.28 Hispanic: 0.63				
NCT01146834	Lymphoma	Male: 1.16 Female: 0.80	White: 0.66 Black: 3.56 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.00 Hispanic: 0.40	N/A	Other	Multi-center	47
NCT01728805	Lymphoma	Male: 1.06 Female: 0.93	White: 0.81 Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-0.39	Industry	Multi-center	372
NCT00019682	Melanoma	Male: 1.10 Female: 0.86	White: 1.00 Black: 0.00 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.00 Hispanic: N/A	-15.58	NIH	Multi-center	185
NCT01936467	Pancreas	Male: 1.11 Female: 0.88	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-5.41	Other	Multi-center	121
NCT02340728	Pancreas	Male: 1.12 Female: 0.87	White: 1.01 Black: 0.63 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.00 Hispanic: 0.00	-9.16	Other	Single-site	24
NCT00002597	Prostate	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-center	1979
NCT00004124	Prostate	Male: N/A Female: N/A	White: 1.05 Black: 0.73 American Indian or Alaska Native: N/A Asian or Pacific Islander: 0.69 Hispanic: 0.79	N/A	Other	Multi-center	961
NCT00116142	Prostate	Male: N/A Female: N/A	White: 0.96 Black: 0.16 American Indian or Alaska Native: N/A Asian or Pacific Islander: 0.45	N/A	Other	Single-site	350

			Hispanic: N/A				
NCT00132301	Prostate	Male: N/A Female: N/A	White: 0.87 Black: 1.51 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.00 Hispanic: 1.29	-4.94	U.S. Fed	Multi-center	297
NCT00134056	Prostate	Male: N/A Female: N/A	White: 1.01 Black: 0.83 American Indian or Alaska Native: N/A Asian or Pacific Islander: 0.79 Hispanic: 0.57	N/A	Other	Multi-center	994
NCT00142506	Prostate	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-center	290
NCT00329797	Prostate	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-center	96
NCT00779402	Prostate	Male: N/A Female: N/A	White: 1.14 Black: 0.41 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.00 Hispanic: N/A	-2.51	Industry	Multi-center	176
NCT01238172	Prostate	Male: N/A Female: N/A	White: 1.00 Black: 0.68 American Indian or Alaska Native: 0.54 Asian or Pacific Islander: 1.41 Hispanic: 0.50	-3.61	Other	Multi-center	443
NCT01415960	Prostate	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	3.79	Industry	Multi-center	161
NCT01538628	Prostate	Male: N/A Female: N/A	White: 1.05 Black: 0.65 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.35 Hispanic: N/A	-0.82	Industry	Multi-center	222

NCT02260817	Prostate	Male: N/A Female: N/A	White: 1.13 Black: 0.44 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.72 Hispanic: 0.00	N/A	Industry	Single-site	109
NCT02680041	Prostate	Male: N/A Female: N/A	White: 1.10 Black: 0.48 American Indian or Alaska Native: 1.12 Asian or Pacific Islander: 0.92 Hispanic: 0.58	N/A	Industry	Multi-center	213
NCT02712320	Prostate	Male: N/A Female: N/A	White: 1.04 Black: 0.60 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 1.31 Hispanic: N/A	7.79	Industry	Multi-center	30
NCT02918357	Prostate	Male: N/A Female: N/A	White: 1.03 Black: 0.13 American Indian or Alaska Native: 0.62 Asian or Pacific Islander: 1.53 Hispanic: N/A	N/A	Other	Single-site	385
NCT02919111	Prostate	Male: N/A Female: N/A	White: 0.91 Black: 0.16 American Indian or Alaska Native: 2.39 Asian or Pacific Islander: 2.88 Hispanic: 0.50	N/A	Other	Single-site	299
NCT02981368	Prostate	Male: N/A Female: N/A	White: 1.08 Black: 0.46 American Indian or Alaska Native: N/A Asian or Pacific Islander: 1.12 Hispanic: N/A	N/A	Industry	Multi-center	385
NCT03353740	Prostate	Male: N/A Female: N/A	White: 0.99 Black: 0.12 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 2.49 Hispanic: 0.48	N/A	Other	Single-site	346
NCT03404648	Prostate	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-3.31	Other	Single-site	19
NCT03739684	Prostate	Male: N/A	White: 1.12	N/A	Industry	Multi-center	208

		Female: N/A	Black: 0.43 American Indian or Alaska Native: N/A Asian or Pacific Islander: 0.56 Hispanic: N/A				
NCT03803475	Prostate	Male: N/A Female: N/A	White: 1.05 Black: 1.00 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 2.34 Hispanic: 0.45	2.92	Other	Single-site	485
NCT00869206	Breast / Prostate	Male: N/A Female: N/A	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-1.91	Other	Multi-center	1822
NCT02195232	Colorectal / Lung / Pancreas	Male: 1.01 Female: 0.76	White: 0.80 Black: 1.11 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.74 Hispanic: N/A	N/A	Other	Multi-center	64
NCT00003816	Leukemia / Lymphoma	Male: 1.05 Female: 0.93	White: 1.09 Black: 0.39 American Indian or Alaska Native: 0.92 Asian or Pacific Islander: 0.25 Hispanic: 0.08	N/A	Other	Single-site	361
NCT00452439	Leukemia / Lymphoma	Male: 1.07 Female: 0.91	White: 1.11 Black: 0.00 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.85 Hispanic: 4.66	N/A	Other	Single-site	72
NCT00516503	Leukemia / Lymphoma	Male: 0.69 Female: 1.37	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-center	203
NCT01231412	Leukemia / Lymphoma	Male: 1.15 Female: 0.79	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	N/A	Other	Multi-center	174
NCT01295710	Leukemia /	Male: 1.00 Female: 1.00	White: 1.05 Black: 0.18	N/A	Industry	Multi-center	254

	Lymphoma		American Indian or Alaska Native: 0.66 Asian or Pacific Islander: 0.36 Hispanic: 0.42				
NCT02349412	Liver / Lung / Pancreas	Male: 1.09 Female: 0.90	White: 0.70 Black: 0.09 American Indian or Alaska Native: 0.41 Asian or Pacific Islander: 0.25 Hispanic: 0.60	N/A	Other	Multi-center	391
NCT01438476	Liver / Pancreas	Male: 1.07 Female: 0.92	White: 1.03 Black: 0.76 American Indian or Alaska Native: 0.00 Asian or Pacific Islander: 0.38 Hispanic: 0.82	N/A	Other	Single-site	140
NCT00377156	Lung / Breast / Prostate	Male: 1.01 Female: 0.99	White: N/A Black: N/A American Indian or Alaska Native: N/A Asian or Pacific Islander: N/A Hispanic: N/A	-2.28	Other	Multi-center	213

Appendix I: Lead sponsorship analysis for its impact on sex race, ethnicity and age

Lead Sponsor	Mean Sex Representation Scores	P-values for Sex Representation	Mean Race/Ethnicity Representation Scores	P-values for Race/Ethnicity Representation	Mean Age Difference	P-values for Mean Age Difference
Industry	Male: 1.03 Female: 1.05	Male: 0.9 Female: 0.7	White: 1.00 Black: 0.53 American Indian or Alaska Native: 0.29 Asian or Pacific Islander: 1.19 Hispanic: 0.74	White: 0.7 Black: 0.4 American Indian or Alaska Native: 0.8 Asian or Pacific Islander: 0.5 Hispanic: 0.9	-4.21	0.2
Other	Male: 1.05 Female: 0.95		White: 0.99 Black: 0.72 American Indian or Alaska Native: 0.51 Asian or Pacific Islander: 0.80 Hispanic: 0.77		-6.49	
NIH	Male: 1.02 Female: 0.98		White: 0.97 Black: 0.52 American Indian or Alaska Native: 0.45 Asian or Pacific Islander: 0.43 Hispanic: N/A		-15.58	

Appendix J: Trial location analysis for its impact on sex race, ethnicity and age

Location Type	Mean Sex Representation Scores	P-values for Sex Representation	Mean Race/Ethnicity Representation Scores	P-values for Race/Ethnicity Representation	Mean Age Difference	P-values for Mean Age Difference
Single-site	Male: 0.98 Female: 1.11	Male: 0.2 Female: 0.09	White: 1.02 Black: 0.42 American Indian or Alaska Native: 0.30 Asian or Pacific Islander: 0.99 Hispanic: 0.85	White: 0.2 Black: 0.01 American Indian or Alaska Native: 0.4 Asian or Pacific Islander: 0.6 Hispanic: 0.9	-7.69	0.4
Multi-center	Male: 1.08 Female: 0.90		White: 0.98 Black: 0.78 American Indian or Alaska Native: 0.51 Asian or Pacific Islander: 0.82 Hispanic: 0.76		-5.44	

Appendix K: Number of total participants per trial analysis for its impact on sex race, ethnicity and age

Location Type	Mean Sex Representation Scores	P-values for Sex Representation	Mean Race/Ethnicity Representation Scores	P-values for Race/Ethnicity Representation	Mean Age Difference	P-values for Mean Age Difference
N = 0-99	Male: 1.17 Female: 0.79	Male: 0.1 Female: 0.1	White: 0.98 Black: 0.96 American Indian or Alaska Native: 0.49 Asian or Pacific Islander: 0.46 Hispanic: 1.26	White: 0.6 Black: 0.3 American Indian or Alaska Native: 0.6 Asian or Pacific Islander: 0.5 Hispanic: 0.3	-1.56	0.3
N = 100-999	Male: 1.01 Female: 1.04		White: 0.99 Black: 0.62 American Indian or Alaska Native: 0.41 Asian or Pacific Islander: 0.96 Hispanic: 0.71		-6.21	
N = 1000+	Male: 1.09 Female: 0.91		White: 1.07 Black: 0.46 American Indian or Alaska Native: 0.95 Asian or Pacific Islander: 0.60 Hispanic: 0.22		-6.82	

Appendix L: Other demographic data grouped by keywords

Exclusion Criteria	Pregnancy / Breastfeeding	Childbearing potential	Other Malignancies	Mental / Psychiatric / Addictive Disorder	Concurrent investigational drug use	Presence of Comorbidities	HIV negative	Performance Score	Language Barrier	Health Insurance Status	Life Expectancy	Assisted Mobility
Number of	45	9	47	20	37	71	9	58	18	1	18	1

Trials												
Total Number of Trials Considered for Each Category	63	63	85	85	85	85	85	85	85	85	85	85