9.20.2021

WHI Race and Ethnicity Language and Data Application and Interpretation Guide

created by: The **WHI Race and Ethnicity Task Force** *(see end of document)*

Intended purpose: This document provides guidance for WHI Publications & Presentations (P&P) reviewers and presenters and authors of WHI manuscripts and presentations on the **language, analytic approach and interpretation of WHI Race and Ethnicity data**. The WHI Race and Ethnicity Task Force recognizes that the concepts, terms and ideas in this Guide will continue to evolve and have recommended that the P&P committee review and update these guidelines periodically, to reflect contemporary thinking.

- *This revision replaces the Draft Guide posted on 3/23/2021, which replaced the 1/27/20 WHI P&P Working Group's report: "Diversity in WHI: Guidance for Authors and Investigators on How to Address in Manuscripts and Proposals".*

The Guide is organized into the following sections; *(Appendices appear, in order, at the end of the document):*

A. Overarching Principles for Conceptualizing Race and Ethnicity

B. WHI Race and Ethnicity Variables: Baseline Form 2 (1993-1998) and Form 41(2003)
   - *Appendix A1. WHI Race and Ethnicity Coding Diagram (F41 mapped dataset)*
   - *Appendix A2. Frequency of race and ethnic categories before and after mapping*
   - *Appendix A3. NIH Enrollment Table based on Form 41 mapped data*

C. Current Definitions and Terminology for Race and Ethnicity

D. WHI Cohort Representativeness of U.S. Women Aged 50-79 by Race and Ethnicity at Enrollment (1993-1998) and by 2019 (the active WHI Extension Study cohort)
   - *Appendix B1: WHI participant race and ethnicity (Form 2) by age groups at Baseline (1993-1998) compared with the U.S. Census 1995 population estimates for women.*
   - *Appendix B2: WHI Extension Study participant race and ethnicity as of September 2019 compared with the U.S. 2019 population estimates for women age 70 and over.*

E. Consideration of Geographic Distribution of Race and Ethnicity participants by WHI Clinical Centers (CC) and Regions (Northeast, South, Midwest, West)
   - *Appendix C1: Map of US with 40 WHI CCs (including 10 "minority recruitment centers")*
   - *Appendix C2: Distribution of WHI participants who identified as Hispanic/Latina, by CC*
   - *Appendix C3: Distribution of WHI participants who identified as a signal race, by CC*
   - *Appendix C4: Distribution of WHI participants enrolled as a given race (C4 a-e) or Hispanic/Latina ethnicity (C4f), ordered by CC with the highest to lowest percent.*

F. Specific Considerations regarding Race and Ethnicity in WHI Analyses and Interpretation (Discussion) of WHI Race and Ethnicity Data

Introduction. This Guide is based on published material from several sources, including:

- A February 2021 Journal of the American Medical Association (JAMA) editorial on reporting of race and ethnicity in medical and science journals (AMA Manual of Style subsection on Race/Ethnicity, Section 11.12.3)[1] and updated AMA guidance (August 2021) which provided suggestions to "encourage fairness, equity, consistency, and clarity in use and reporting of race and ethnicity in medical and science journals."[2];
- The November 2020 American Heart Association (AHA) Structural Racism and Health Equity Language Guide[3];
- The November 2020 AHA Presidential Advisory on structural racism[4];
- A July 2020 Health Affairs blog on a new standard for publishing on racism.[5]

A. Overarching principles:

1. **Race** is a socio-political construct that often serves as a proxy for both historical and ongoing disadvantage in social, economic, environmental, and structural factors, arising from racism. It is **not** rooted in **biology**; in fact, a "biological" basis for race has been debunked in the scientific literature.[5-14] *It is important to think carefully about why race variables are being used, i.e. what is represented by "race",[6] and how it is contributing to the scientific question being addressed*.
2. **Ethnicity** is another socio-political construct **not** rooted in **biology,** that refers to the social characteristics people may have in common, such as language, religion, regional background, traditions and culture. Ethnicity is not equivalent to race and should not be presented as a race category nor should "race/ethnicity" be listed as a descriptor. *Our ability to analyze "ethnicity" in WHI is restricted to "non-Hispanic" or "Hispanic," who can be of any race, as appears on **Form 41** (per Census 2000).*
3. **Ancestry**, a person's country or region of origin or an individual's lineage of descent, and **genetic admixture**, which refers to genetic exchange among people from different ancestries, are appropriate terms for genetic analyses and are **not** interchangeable with "race" or "ethnicity", neither of which can be determined by the geographic origin of a person's ancestry.
4. When describing **race or ethnicity**, it is best to be specific and refer to people by terms and categorization as they self-identify *(if known),* but it is understood that individuals have been asked to choose among limited categories *(influenced by current socio-political biases)* when reporting their race or ethnicity, *as is true for race and ethnicity data collected using the WHI forms, as well as U.S. Census reporting.*
5. The **dynamic and contextual nature of race and ethnicity** requires periodic reconceptualization.
    - *The TF recommends routine review and, if appropriate, revision of this Guide, recognizing that these factors are grounded in conceptual models and structural racism and are likely to change as we become more educated on these issues.*

    Note: To enhance the rigor of research across each WHI race and ethnic category, greater detail would be required, including racial or ethnic subpopulation, religion, immigration status and acculturation measures, as well as other social determinants of health, including everyday discrimination.

B. WHI Race and Ethnicity Variables: Baseline Form 2 [link] and Form 41 [link]

1. **Baseline (1993-1998) WHI Form 2, Question #15** asked participants to "describe your **race or ethnic group**." ["If of mixed blood, which group do identify with most?"]. 6 categories: (1) American Indian or Alaska Native; (2) Asian or Pacific Islander *(ancestry is Chinese, Indo- Chinese, Korean, Japanese, Pacific Islander, Vietnamese)*; (3) Black or African-American *(not of Hispanic origin);* (4) Hispanic/Latino *(ancestry is Mexican, Cuban, Puerto Rican, Central American, or South American);* (5) White *(not of Hispanic origin);* and, ("8") Other *(Specify).* [Notes: "Other" was checked by 1849 participants; 413 participants left the question blank; therefore, the data are "missing".]
2. **Form 41** was developed by the WHI Special Populations Advisory Board *(chaired by Electra Paskett, PhD)* in **2003** to **collect 2000 Census race and ethnicity data** in all WHI participants who were active in WHI at the time.

    Q 1. (Ethnicity) asked: **Are you Spanish/Hispanic/Latino?** Mark "No" box if not

Spanish/Hispanic/ Latino. (0) No, not Spanish/Hispanic/Latino; (1) Yes, Puerto Rican; (2) Yes, Mexican, Mexican American, or Chicano; (3) Yes, Cuban; (4) Yes, other Spanish/Hispanic/Latina. Please specify what group.

Q 2. (Race) asked: **"What is your race?** Mark **one or more races** to indicate what you consider yourself to be: (1) White; (2) Black, African-American, or Negro; (3) American Indian or Alaska Native *(Please specify enrolled or principal tribe);* (4) Asian Indian; (5) Chinese; (6) Filipino; (7) Japanese; (8) Korean; (9) Vietnamese; (10) Other Asian (Please specify race); (11) Native Hawaiian; (12) Guamanian or Chamorro; (13) Samoan; (14) Other Pacific Islander *(Please specify race)*; (15) Some other race *(Please specify race).*

3. **Notes regarding WHI Race and Ethnicity Variables** *(and Language)*

Census 2000 instructions and categories (and, thus WHI Form 41) presented several subgroups for Asians (#4-10 above) and Pacific Islanders (#11-14 above) as individual races, even though these subgroups are generally grouped as "Asian" or "Pacific Islander" race, respectively, or combined as Asian/Pacific Islander *(as in WHI Form 2)*. Form 41 data enable WHI to separate Asian Americans from Pacific Islanders *(see Appendices A1 and A2 and text below.)*

Similarly, "American Indian" is often combined with "or Alaska Native", as one race category (as in Form 2 and Form 41), with no means to separate them. Neither of the WHI forms (nor did the 2000 Census) defined either term, but Form 41 encouraged a write-in entry for: "*Please specify enrolled or principal tribe"*. [*Note: the write-in answers that identify tribal affiliation were not entered in the WHI database.]*.

4. **WHI Race and Ethnicity Mapping of Form 2 onto Form 41.**

In Fall 2020-Winter 2021, the WHI-CCC developed an algorithm to map WHI Form 2 combined race/ethnicity variables onto WHI Form 41 race and ethnicity variables, as shown in:

- **Appendix A1**. WHI Race and Ethnicity Coding Diagram.

The **F41-mapped dataset** (**labeled 'f41_imputed_ctos_inv.dat'**) is posted on the WHI website and **should be used in all future WHI analyses and papers**.

The **Other** descriptor: The "other" category should only be used when a participant selected "other" as her identity. It should never be used to combine race (or subgroup) groups identified by participants for any purpose, but rather, each race (or, if relevant, subgroup) should be described, such as in a Baseline Characteristics table. If a woman did not answer a question, the data were "missing" and race and/or ethnicity were "unknown".

The frequency of race and ethnic categories before and after application of the mapping algorithm is presented in *Appendix A2*, with the numbers of women by self-identified race and ethnicity "as collected on Form 2 or Form 41", including the Form 41 race and ethnicity subgroups, juxtaposed with the numbers after the mapping algorithm was applied. Note that women who identified as more than one race are presented as a single race category; therefore, there is no overlap and the numbers that add up to 161,808 women.

- **Appendix A2**. Frequency of race and ethnic categories before and after application of the mapping algorithm

One of the goals of this mapping process was to generate an **NIH Enrollment Table** with both Race and Ethnicity, based on the WHI-mapped Form 41 data:

- **Appendix A3**. NIH Enrollment Table based on WHI-mapped Form 41 data

C. Race and Ethnicity Definitions and Terminology (Language)

A few basic language and writing guidelines and race and ethnicity definitions are presented here. See sources cited in the Introduction, particularly the updated AMA guidance,[2] for more detailed definitions, including for definitions for structural racism.

1. **Race.** The U.S. Census Bureau defines race as a person's self-identification with one or more social groups: The names of races should be capitalized. Racial and ethnic terms should not be used in noun form; the adjectival form is preferred, e.g. Black participants, White participants (-or- participants who are Black, White).
    - **African American or Black**: African American is acceptable for Black Americans of African descent; if preference is unknown, use Black. Also use Black in racial, ethnic, cultural differences outside the U.S. "Black" is only acceptable as an adjective, e.g. Black women - not Blacks; capitalize when referring to people, communities, populations or culture. [Do not use dated or potentially offensive terms, e.g. "Negro" and "colored," *unless part of formal name of an organization or a quotation. (Note: "Negro" was dropped from Census 2020.)*]
    - **Asian**: The term "Asian American" is acceptable when describing individuals who identify with Asian descent among the U.S. population. When possible, refer to a person's self-identified county of origin. Form 41 collected data on six specific racial or ethnic subpopulations and "other Asian *(please specify)"; however, WHI forms with the write-in data were discarded before the data were entered.* [Do not use "Oriental" or "Asiatic."]
    - **Native Hawaiian** and **"Other Pacific Islander"**: Form 41 distinguishes Guamanian or Chamorro and Samoan from "Other Pacific Islander *(Please specify race)".*
        - Many individuals identify as "Asian American Pacific Islander (AAPI"); however, this was not a category a WHI participant could choose to (self-) identify her race. The term should not be used to describe a WHI participant.
    - **American Indian** or **Alaska Native**: Although current AMA guidelines suggest that AIAN persons should be identified by their self-identified tribal affiliation, WHI did not distinguish between these two groups in aggregate nor at the tribal level.
        - *Many individuals identify as "Native American" or "Indigenous People"; however, WHI participants were not given these as options to (self-) identify race; therefore, these terms should not be used to describe WHI participants.*
        - Note: the 2020 Census defined "American Indian" as all individuals who identify with any of the original peoples of North and South America (including Central America) and who maintain tribal affiliation or community attachment. This definition was not presented in WHI Forms 2 and 41 and it should not be assumed that it applies to a WHI participant.
    - **White**: Use uppercase *White* to describe people and populations in professional copy, e.g. AMA; use lowercase *white* in consumer copy, e.g. AP Stylebook *(per AHA).* [Do not use "Caucasian".]

2. **Ethnicity.** The U.S. **O**ffice of **M**anagement and **B**udget **defines ethnicity** as either **"Hispanic or Latino"** or **"Not Hispanic or Latino."** OMB defines "Hispanic or Latino" as a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of **race**.
    - Note: WHI Form 2 did not make a clear distinction between race and ethnicity and described "Hispanic/Latino" as *"ancestry is Mexican, Cuban, Puerto Rican, Central American, or South American*, which is inconsistent with the 2020 definition of "American Indian" described above.
    - Note: There is considerable controversy about applying this ethnic term to indigenous people or people of another European origin, such as Brazilians with Portuguese roots, or other world regions, such as from African or Asian countries.
    - There is also controversy regarding inclusion of "Spanish" in the Hispanic/Latino category, as was done on Form 41 (but not Form 2), as many consider this "of European origin", rather than an ethnic group distinct from "non-Hispanic" people.
    - One can use Non-Hispanic *(an adjective, not a noun)* for professional audiences, to distinguish from Hispanic White women, Hispanic Black women, etc.
        - **Latina** (women) is an acceptable term for WHI participants who checked "Latino". The AHA guide considers "**Latinx**" also to be acceptable, it should not be used to describe WHI participants as they did not have this option.

3. **Specific Language Considerations**

    The use of terms to group race and ethnicities should be well conceptualized and descriptive of the grouping. For example, grouping may be useful when describing specific experiences with systemic racism that similarly affect multiple groups. Ideally, individuals should be referred to by their self-identified race and ethnicity.
    - Terms that re-enforce the White race versus an "others" dichotomy should not be used, such as "non-white", "minority", or other commonplace terms for racialized groups.
    - There should be clear rationale for using terms such as *"underserved groups", "vulnerable", "underrepresented populations', "historically marginalized populations or racial and ethnic groups"* and other terms, which may be suitable in certain contexts.
    - WHI participants were not asked how they would like to be referred, beyond the race and ethnic categories offered in Forms 2 and 41, and terms such as "woman of color", "People of Color"**,** Black, Indigenous and other People of Color (BIPOC)" or "brown" may not resonate with them, WHI investigators should use the terms participants self-identified when discussing WHI participants or data.
    - A key principle is that "inclusive language supports diversity and conveys respect. Language that imparts bias toward or against persons or groups on characteristics or demographics must be avoided." [1.]
    - The terms "multiracial, mixed race, and multiethnic" are acceptable in reports of studies if the specific categories the terms comprise are defined. If the criteria for data quality and confidentiality are met, at a minimum, the number of individuals identifying with more than one race should be reported.

- The nonspecific group label "other" should be avoided unless it was a prespecified form category; in such cases, the categories included in "other" groups should be defined and reported if possible. Authors are advised to be as specific as possible when reporting on racial and ethnic categories (even if these categories contain small percentages).
- Language is changing rapidly as more diverse perspectives and opinions are heard and some of what the AMA[1,2] and/or AHA[3] sources cited in this Guide suggest regarding terminology will likely be revised. The key point for WHI is that authors should use the specific terms participants selected on Forms 2 and 41, as mapped, to refer to WHI participants' self-identified race and ethnicity, despite changes in terminology since WHI baseline and 2003 data collection. Different terminology might be appropriate when referring to other cohorts or persons.

D. WHI Cohort Representativeness of U.S. Women by Age and Race and Ethnicity

Authors are encouraged to address how representative of the U.S. population the WHI cohort is, in the context of interpreting the generalizability of the analytical results. When evaluating the context of results from WHI that includes a range of race and ethnic, socioeconomic, and/or educational subgroups, it is important to consider generalizability along with relevant confounders and mediators, noting that *comparisons of WHI participants to the U.S. population should be both age- race and -ethnicity specific, i.e. based on the proportion of older women within each race and ethnic group*. Discussion sections should address implications for analyses examining racial and/or ethnic inequities, which may be underestimated compared to those observed in the general U.S. (older) female population.

- **Appendix B1**: WHI participant race and ethnicity (Form 2) by age groups at Baseline (1993-1998) compared with the U.S. Census 1995 population estimates for women.
- **Appendix B2:** WHI Extension Study participant race and ethnicity (using Form 2 for comparability) as of September 2019 compared with the U.S. 2019 population estimates for women age 70 and over.

E. Geographic Origin and U.S. Regionalization Considerations

1. Awareness of the relevance of geographic origin and regionalization associated with racial and ethnic designations may be important, depending on the research question.
   - For example, sickle cell anemia, should be thought of connected not to race but to geographic ancestry, as it arose where malaria is or was present, including sub-Saharan Africa, the Mediterranean, and the Indian subcontinent.
2. **U.S. geographic diversity** is particularly rich in the WHI and may be relevant when focusing on race and ethnicity in WHI analyses. Recruitment strategies varied across the 40 clinical centers (CCs), which have been grouped as four regional centers: Northeast, South, Midwest, and West. For example, to enhance racial and ethnic diversity of the WHI cohort, 10 CCs were designated as "minority recruitment centers" and were expected to enroll 60% of their participants from four* specified race and ethnicity groups: African American; Asian/Pacific Islander *(*as a combined category)*; Native American/Alaska Native; and Hispanic.
   - **Appendix C1**: Map of US with location of 40 WHI CCs (with distinction of 10 "minority recruitment centers")

   A consequence of this strategy, which should be recognized in the development of

WHI manuscripts and presentations, is a potentially confounding influence of geographic and regional sociocultural factors on racial and ethnic comparisons. This approach may also have influenced the distribution of representation by race or ethnicity (e.g. Tucson recruited a higher number of Mexican American Hispanics while Miami recruited more Cuban Hispanics). Thus, comparisons among women who identified as Cuban, Puerto Rican, Mexican American, or "other Spanish/ Hispanic/Latina" should acknowledge differences that may have nothing to do with or may intersect with ethnicity related to residing in Miami, New York, San Antonio, La Jolla, Tucson or other US sites.

Further, any comparisons between "Asian" and "Pacific Islander" participants is possibly confounded by the fact that 54% of the "Asian" and 69% of the "Pacific Islander" WHI participants were enrolled at the Honolulu (Hawaii) site, with most residing on the island of Oahu, while most of the mainland Asian participants were enrolled by California sites. Graphs were developed for this Guide to inform writing groups on this point. *(Note that CCs were coded, with only "minority CC" status and WHI regional center shown, so as not to deidentify participants enrolled at CCs in very small numbers of a given group.)*

- **Appendix C2:** Distribution of WHI enrollment of participants who identified as (mapped) Hispanic/Latina (N=7312), by Form 41 subgroups, with CCs orderd by highest to lowest total Hispanic/Latina recruitment. (* indicates "minority recruitment center").
  - Note: 153,034 (of the total 161,808) WHI participants marked "No, Not Spanish/Hispanic/Latino" and 1462 are "unknown".
- **Appendix C3:** Distribution of WHI enrollment of participants who identified as a (mapped) given (one) race (N= 156,657), with 10 "minority recruitment CCs" as first 10 CCs, followed by the 30 other CCs, ordered by total percent of enrollment of targeted race and ethnicity groups.
  - Note: 157,582 WHI participants reported one race (as presented in C3), 1880 reported more than one race, and 2346 are "unknown".
- **Appendix C4: a-f:** Distribution of WHI participants who identified as a (mapped) given (one) race (C4a-e), *ordered by CC with the highest to lowest percent of each race* [a. White (N=137,628); b. Black/African American (N=14,327); c. Asian (N=4025); d. American Indian/Alaska Native (N=540); e. Pacific Islander (N=137)] and ethnicity [f. Hispanic/Latina ethnicity (N=7312).

  The large numbers and more even distribution of Black, White and Hispanic participants enrolled across the U.S. clinical centers provides the opportunity to study the role of geographic region on health, taking into account differences by age and social determinants of health)

F. <u>Specific Considerations for Including Race and Ethnicity in WHI Analyses:</u>

- <u>Develop Questions and Methodological Strategies Informed by Conceptual Frameworks</u>[15]

  In the study design and data interpretation stages of race- and ethnicity-focused research, identify conceptual models to target interpretation of the structural factors and racism underlying race and ethnic disparities. For example:

  - Public Health Critical Race Methodology (PHCR)[6,7] offers conceptual guidance

for distinguishing racism and health inequities from race as a risk factor.

- "Scientists can consider using frameworks such as the National Institute of Minority Health and Health Disparities Research Framework[17] to develop study questions that consider domains of influence (e.g., behavioral, sociocultural/environmental) with levels of influence (e.g., individual, interpersonal, societal)."[15]

Note: It is recommended that WHI writing groups seek advice from the Race, Ethnicity and Health Disparities Scientific Interest Group [link] and/or consider inviting a member of this SIG to join the writing group wherein the evaluation of race or ethnicity related research questions are a focus (for overall expertise and working knowledge of this guidance).

- Reporting of demographic data on race and ethnicity
  - Manuscripts should include an explanation of who identified participant race and ethnicity and the source of the classifications used (e.g. in WHI, by self-report)
  - Rationale for use of race as a key variable within WHI: For papers and ancillary studies where race is the **primary exposure** of interest or where analyses are **stratified by race** and/or **ethnicity**, authors should provide a clear, written definition and rationale for why race is being used (e.g., what it is serving as a proxy for).

- Characterization of racial and ethnic identity is not fixed; available options using national surveys (e.g., census) have changed over time. WHI participants self-identified their race and ethnicity per Census 2000 categories in 2003 (Form 41). We acknowledge WHI limitations due to F2/F41 categories (Census 2000), but Form 41 is not available on all participants from Baseline. (*Current Census 2020 classifications are not available in WHI*.

- Analyses by Race and/or Ethnicity
  - Comparisons between race or ethnic groups should be informed by research questions.
  - Comparisons of Non-Hispanic White participants to other race or ethnic groups is not required of studies within WHI. Clear justification of the informative value of such an analysis is required. If there is scientific rationale to make such comparisons, one should not group "all other race or ethnic" categories to be compared with White.
  - Within group analyses serve to highlight the heterogeneity and resilience available within racial and ethnic groups. It is important that research questions examine associations within select historically marginalized race and ethnic groups; however, the decision to make comparisons between race or ethnic groups should be informed by the research questions

- **Data Interpretation & Reporting**

  We have an ethical responsibility to present data on all race and ethnic sub-groups, but appropriate interpretation is important.

  - **Statistical power for race and ethnicity subgroup analyses**: If conducted, racial or ethnic subpopulation analyses, *as is the case for all subgroup analyses*, should be sufficiently powered to detect differences by that group.

Results from analyses with insufficient power based on smaller sample size should be reported with caution. When describing WHI results across race and ethnicity groups, it is **essential that authors provide a clear context for interpretation and for applicability to any subgroups**. The discussion should clearly acknowledge that sample selection limits interpretation of findings to the overall U.S. population or racial or ethnic subpopulations identified in the manuscript.

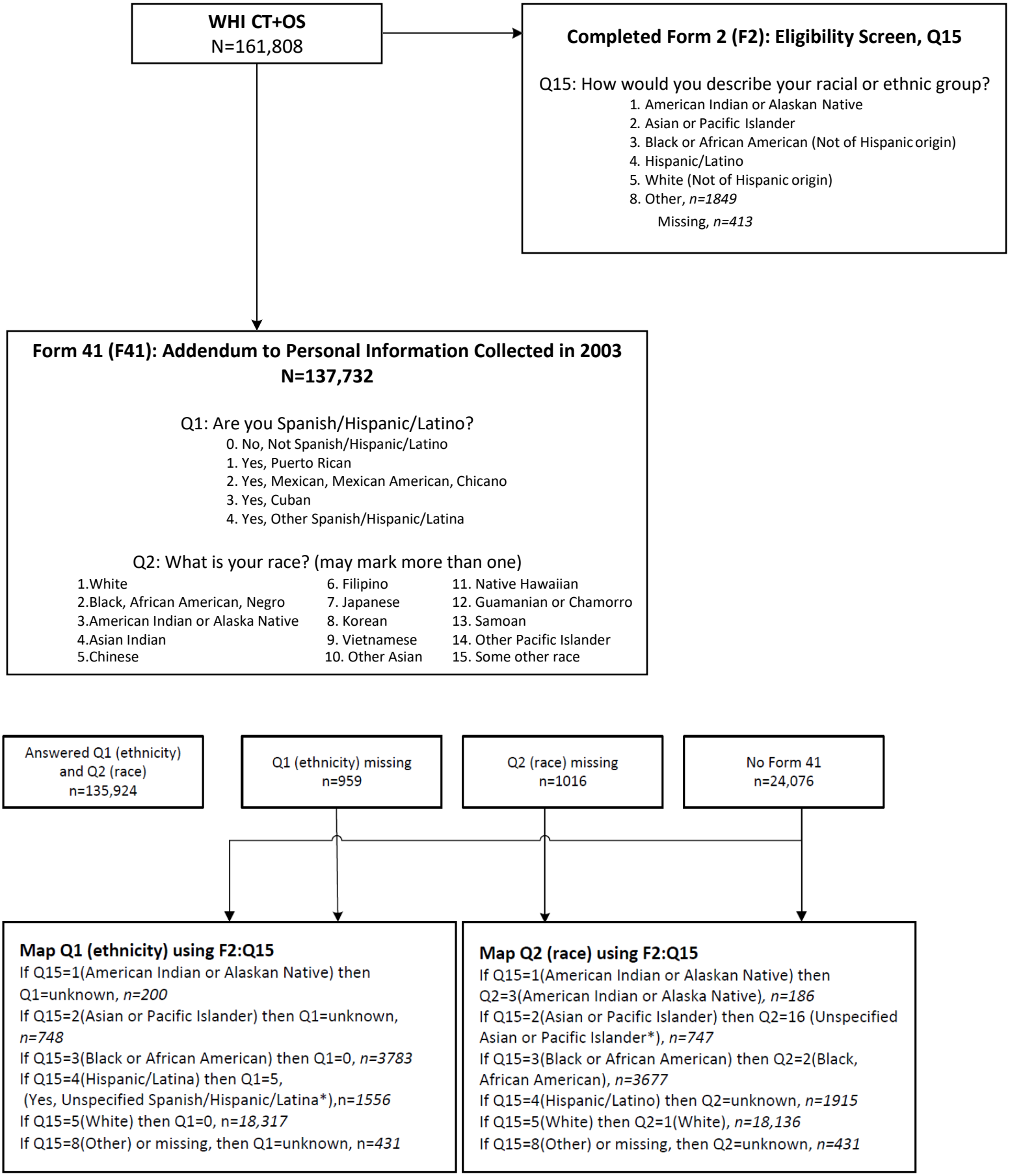
- Over time, the **WHI sample composition has been influenced by selective drop-out** that can be investigated through the use of inverse probability weighting and other methods. Although there is inequality across all variables, the WHI participants have similarities compared to other women in their age range. [See **Appendix** B1 (Baseline) and **Appendix** B2 (September 2019)]
- **Limitation:** Race and Ethnicity are defined and interpreted within a socio-political framework as a proxy for both historical and ongoing differences in advantages arising from racism in social determinants of health, such as education, income, resilience and stressful life events (see WHI psychosocial constructs [link]). Other structural factors may be important for the data interpretation of racially disparate outcomes, such as discrimination, racial capitalism, adverse childhood experiences, and inheritance of wealth, but are not available in WHI.

WHI Race and Ethnicity Task Force:

Marcia L. Stefanick, PhD (TF Chair), Aaron Aragaki, MS, Khadijah Breathett, MD, Crystal Cené, MD, MPH (Race, Ethnicity and Health Equity SIG Chair), Peggye Dilworth-Anderson, PhD, Shawna Follis, PhD, Lorena Garcia, PhD, Monik C. Jiménez, ScD, Charles Kooperberg, PhD, Kamal Masaki, MD, Electra D. Paskett, PhD, Mary Pettinger, MS, Cynthia A. Thomson, PhD, RD.

References/Resources

1. Flanagin A, Frey T, Christiansen SL, Bauchner H. The Reporting of Race and Ethnicity in Medical and Science Journals. JAMA. 2021 Feb 22. [link]
2. Flanagin A, Frey T, Christiansen SL, AMA Manual of Style Committee. Updated Guidance on the Reporting of Race and Ethnicity in Medical and Science Journals. JAMA. 2021 08 17;326(7):621-627.[link]
3. AHA Structural Racism and Health Equity Language Guide. [link]
4. Churchwell K, Elkind MSV, Benjamin RM, et al. American Heart Association. Call to Action: Structural Racism as a Fundamental Driver of Health Disparities: A Presidential Advisory From the American Heart Association. Circulation. 2020 Nov 10:CIR0000000000000936. [link]
5. Boyd RW, Lindo EG, Weeks LD, McLemore MR. On Racism: A New Standard For Publishing On Racial Health Inequities. Health Affairs July 2, 2020 [link]
6. Ford CL, Airhihenbuwa CO. Critical race theory, Race Equity, and Public Health:Toward Antiracism Praxis. American Journal of Public Health, 100(Suppl 1), S30e35. [link]
7. Ford CL, Airhihenbuwa CO. The public health critical race methodology: Praxis for antiracism research. Soc Sci Med. 2010;71(8):1390-1398. [link]
8. New AMA policies recognize race as a social, not biological, construct | American Medical Association (ama-assn.org), Nov 16, 2020 AMA Press release. [link]
9. Race Is a Social Construct, Scientists Argue. Scientific American. [link]
10. Race as a Social Construction. Psychology Today. [link]
11. Kwabi-Addo B. Race: a biological or social concept. Health Outcomes in a Foreign Land 2017. [link]
12. Allen TW. The invention of the white race. Vol 2. Verso, 1994. [link]
13. Social Constructions of Race. Cultural Anthropology. [link]
14. Ioannidis JPA, Powe NR, Yancy C. Recalibrating the Use of Race in Medical Research. *JAMA.* 2021;325(7):623-624. [link]
15. Breathett K, Spatz ES, Kramer DB, et al. The Groundwater of Racial and Ethnic Disparities Research: A Statement From Circulation: Cardiovascular Quality and Outcomes. *Circ Cardiovasc Qual Outcomes.* 2021;14(2):e007868. [link]
16. U.S. National Census 2020 [link]
17. NIMHD Minority Health and Health Disparities Research Framework. 2018. [link]
18. Fouad MN, Corbie-Smith G, Curb D, et al. Special populations recruitment for the Women's Health Initiative: successes and limitations. Control Clin Trials. 2004;25(4):335-352.[link]



* new category

Appendix A2. Frequency of race and ethnicity categories before and after application of mapping algorithm

N=161,808	As collected on Form 41 or Form 2 N	Mapped value after algorithm application N
Ethnicity: Spanish/Hispanic/Latino		
No, Not Spanish/Hispanic/Latino	131,017	153,034
Did not complete 2003 WHI Form/White or Black on baseline WHI Form	22017	
Yes, Puerto Rican	779	779
Yes, Mexican, Mexican American, Chicano	2693	2693
Yes, Cuban	396	396
Yes, Other Spanish/Hispanic/Latina	1888	1888
Yes, Unspecified Spanish/Hispanic/Latina (Did not complete 2003 WHI Form/Hispanic on baseline WHI Form)	1556	1556
Unknown (Did not complete 2003 WHI Form/Not White/Black/Hispanic on baseline Form)	1379	1379
Unknown (Did not complete2003 or baseline Forms)	83	83
Total	**161,808**	**161,808**
Race:		
One reported race	134,836	157,582
White	119,492	137,628
Black, African American, or Negro	10,650	14,327
American Indian or Alaska Native	354	540
Asian (combining #4-10 from 2003 Form)	3278	4025
Asian Indian	*83*	*83*
Chinese	*747*	*747*
Filipino	*321*	*321*
Japanese	*1962*	*1962*
Korean	*91*	*91*
Vietnamese	*10*	*10*
Other Asian	*64*	*64*
Unspecified Asian (Did not complete 2003 Form/Asian or Pacific Islander, baseline form)	*747*	*747*
Pacific Islander (#11-15 from 2003 Form)	137	137
Native Hawaiian	*97*	*97*
Guamanian or Chamorro	*10*	*10*
Samoan	*2*	*2*
Other Pacific Islander	*28*	*28*
Some other race	925	925
More than one race	1880	1880
Unknown (Did not complete 2003 /White, Black, Asian/ PI, American Indian/Alaskan Native on baseline form)	24,661	In one of above categories
Unknown (Did not complete 2003 Form/ Hispanic or Other on baseline form)	2264	2346
Unknown (Did not complete2003 or baseline Forms)	82	
Total	**161,808**	**161.808**

Appendix A3. NIH Enrollment Table based on WHI mapped Form 41 data

Racial Categories	Ethnic Categories									Total
	Not Hispanic or Latino			Hispanic or Latino			Unknown/Not Reported Ethnicity			
	Female	Male	Unknown/ Not Reported	Female	Male	Unknown/ Not Reported	Female	Male	Unknown/ Not Reported	
American Indian/ Alaskan Native	292	0	0	53	0	0	195	0	0	540
Asian	3216	0	0	60	0	0	749	0	0	4025
Native Hawaiian or Other Pacific Islander	119	0	0	18	0	0	0	0	0	137
Black or African American	14166	0	0	160	0	0	1	0	0	14327
White	133321	0	0	4300	0	0	7	0	0	137628
More than one Race	1662	0	0	211	0	0	7	0	0	1880
Unknown or Not reported	341	0	0	2510	0	0	420	0	0	3271
Total	153117	0	0	7312	0	0	1379	0	0	161808

Coding instructions

1) Columncategories:useForm41imputedQuestion1(Ethnicity);combine'Yes,PuertoRican','Yes,Mexican,MexicanAmerican,orChicano','Yes,Cuban'and 'Yes, other Spanish/Hispanic/Latino' into 'Hispanic/Latino'.
   Row categories: count the number of race categories marked in Question 2 (Race); If number of race categories is greater than one, category = 'More than one race'; else if number of race categories equals one, use categories for American Indian/Alaskan Native, White, Black or African American as is, and create aggregated categories for Asian = Asian Indian or Chinese or Filipino or Japanese or Korean or Vietnamese or Other Asian, and Native Hawaiian or Pacific Islander = Native Hawaiian or Guamanian/Chamorro or Samoan or Other Pacific Islander

Appendix B1. Race and ethnicity (Form 41 imputed) by age groups of WHI Participants at Baseline (1993-1998) compared with the US Census 1995 population estimates for women.

US 1995[4]	***Ethnicity***	***Race***						
	Spanish/ Hispanic/ Latino	*Black/African American*	*American Indian/Alaska Native*	*Asian/ Pacific Islander*		*White*		
Total, %	5.9%	9.9%	0.6%	2.7%		86.8%		
Age, %								
50 to 54 years	7.1%	10.9%	0.7%	3.3%		85.2%		
55 to 59 year	7.0%	11.1%	0.7%	3.1%		85.2%		
60 to 64 years	6.4%	10.6%	0.6%	2.9%		85.9%		
65 to 69 years	5.6%	9.7%	0.5%	2.6%		87.3%		
70 to 74 years	4.6%	8.3%	0.4%	2.1%		89.2%		
75 to 79 year	3.9%	8.0%	0.4%	1.6%		90.0%		
WHI Baseline	***Ethnicity[1]***	***Race***						
N = 161,808	*Spanish/ Hispanic/ Latino*	*Black/African American*	*American Indian/Alaska Native*	*Asian[2]*	*Pacific Islander[3]*	*White*	*Unknown*	*Some Other Race*	*Two or more races*
Total, N (%)	7312 (4.5%)	14,327 (8.9%)	540 (0.3%)	4,025 (2.5%)	137 (0.1%)	137,628 (85.1%)	2,346 (1.4%)	925 (0.6%)	1,880 (1.2%)
Age, %									
50 to 54 years	7.8%	12.4%	0.6%	2.9%	0.2%	78.9%	2.4%	1.1%	1.6%
55 to 59 year	5.9%	10.3%	0.4%	2.4%	0.1%	83.1%	1.7%	0.8%	1.3%
60 to 64 years	4.5%	9.7%	0.3%	2.3%	0.1%	84.4%	1.5%	0.5%	1.2%
65 to 69 years	3.4%	7.1%	0.3%	2.4%	0.1%	87.7%	1.0%	0.4%	1.0%
70 to 74 years	2.6%	6.2%	0.3%	2.6%	0.0%	88.6%	1.1%	0.3%	0.9%
75 to 79 year	2.2%	6.4%	0.2%	2.8%	0.0%	88.3%	1.1%	0.3%	0.9%

1. Includes Puerto Rican, Mexican, Mexican American, or Chicano, Cuban and other Spanish/Hispanic/Latino
2. Includes Asian Indian or Chinese or Filipino or Japanese or Korean or Vietnamese or Other Asian
3. Includes Native Hawaiian or Guamanian/Chamorro or Samoan or Other Pacific Islander
4. Source: Day, Jennifer Cheeseman, Population Projections of the United States by Age, Sex, Race, and Hispanic Origin: 1995 to 2050, U.S. Bureau of the Census, Current Population Reports, P25- 1130, U.S. Government Printing Office, Washington, DC, 1996.

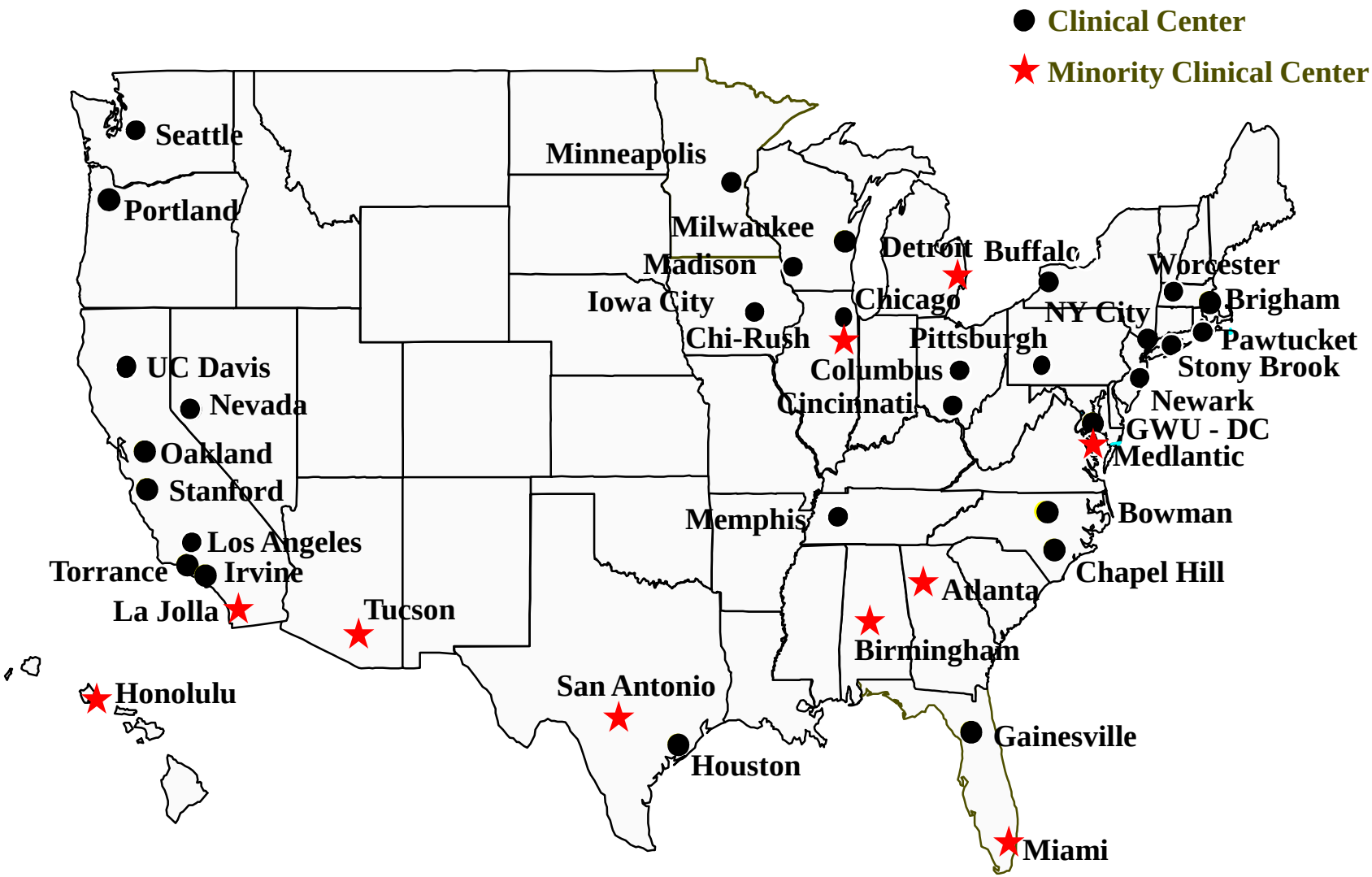
Appendix B2**.** Race and ethnicity (Form 41 imputed) of WHI Extension Study Participants in 2019 compared with the US Census 2019 population estimates for women.

US 2019[4]	***Ethnicity***	***Race***						
	Spanish/ Hispanic/ Latino	*Black/African American*	*American Indian/Alaska Native*	*Asian*	*Pacific Islander*	*White*	*Two or more races*	
Total, %	8.4%	9.7%	0.7%	4.6%	0.1%	84.0%	0.9%	
Age, %								
70 to 74 years	8.6%	10.2%	0.8%	4.9%	0.1%	83.0%	1.0%	
75 to 79 year	8.4%	9.7%	0.7%	4.5%	0.1%	84.1%	0.9%	
80 to 84 years	8.6%	9.6%	0.6%	4.5%	0.1%	84.3%	0.8%	
85 and over	7.7%	8.7%	0.5%	4.4%	0.1%	85.6%	0.7%	

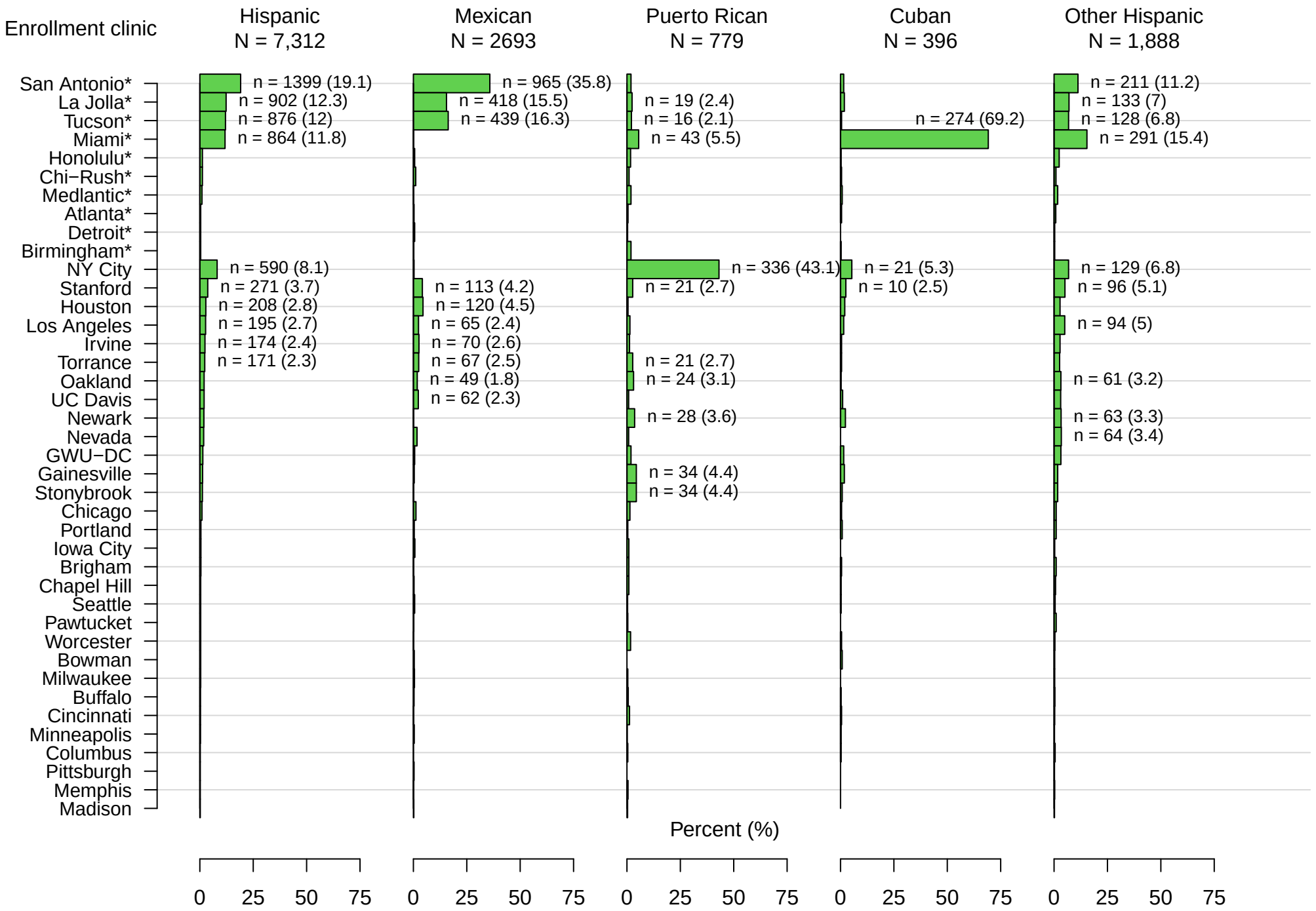
WHI 2019	***Ethnicity***[1]	***Race***							
N = 67,140	*Spanish/ Hispanic/ Latino*	*Black/African American*	*American Indian/Alaska Native*	*Asian*[2]	*Pacific Islander*[3]	*White*	*Unknown*	*Some Other Race*	*Two or more races*
Total, N (%)	2,302 (3.4%)	4,247 (6.3%)	150 (0.2%)	1,394 (2.1%)	49 (0.1%)	59,819 (89.1%)	284 (0.4%)	362 (0.5%)	835 (1.2%)
Age, %									
70 to 74 years	244 (6.6%)	396 (10.6%)	18 (0.5%)	130 (3.5%)	9 (0.2%)	3,028 (81.3%)	31 (0.8%)	50 (1.3%)	64 (1.7%)
75 to 79 year	733 (4.3%)	1,278 (7.5%)	55 (0.3%)	376 (2.2%)	15 (0.1%)	14,929 (87.2%)	97 (0.6%)	115 (0.7%)	253 (1.5%)
80 to 84 years	660 (3.3%)	1,303 (6.5%)	42 (0.2%)	375 (1.9%)	14 (0.1%)	17,914 (89.2%)	82 (0.4%)	99 (0.5%)	259 (1.3%)
85 to 89 years	426 (2.8%)	808 (5.3%)	21 (0.1%)	310 (2.0%)	8 (0.1%)	13,852 (90.6%)	55 (0.4%)	60 (0.4%)	168 (1.1%)
90 to 94 years	198 (2.3%)	375 (4.4%)	10 (0.1%)	158 (1.8%)	3 (0.0%)	7,936 (92.2%)	16 (0.2%)	39 (0.3%)	77 (0.9%)
Over 95 years	41 (1.8%)	87 (3.8%)	4 (0.2%)	45 (1.9%)	0	2,160 (93.0%)	3 (0.1%)	9 (0.4%)	14 (0.6%)

1. Includes Puerto Rican, Mexican, Mexican American, or Chicano, Cuban and other Spanish/Hispanic/Latino
2. Includes Asian Indian or Chinese or Filipino or Japanese or Korean or Vietnamese or Other Asian
3. Includes Native Hawaiian or Guamanian/Chamorro or Samoan or Other Pacific Islander
4. Source: US Census Bureau, Population Division. Annual Estimates of the Resident Population by Sex, Age, Race, and Hispanic Origin for the United States: April 1, 2010 to July 1, 2019 (NC- EST2019-ASR6H

WHI Clinical Centers

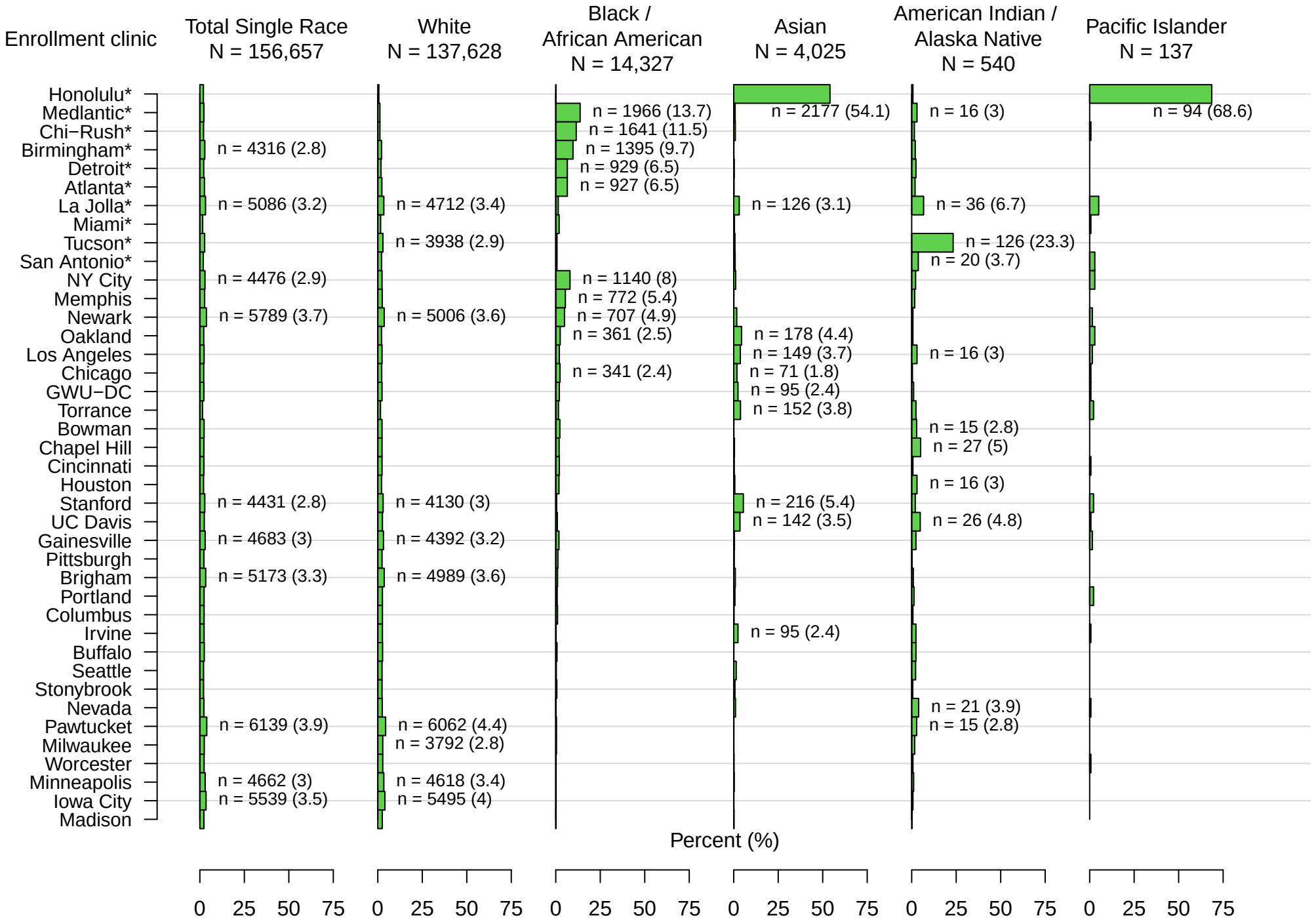


Appendix C2. Distribution of WHI participants who identified as Hispanic/Latina by subgroups (N=7312).

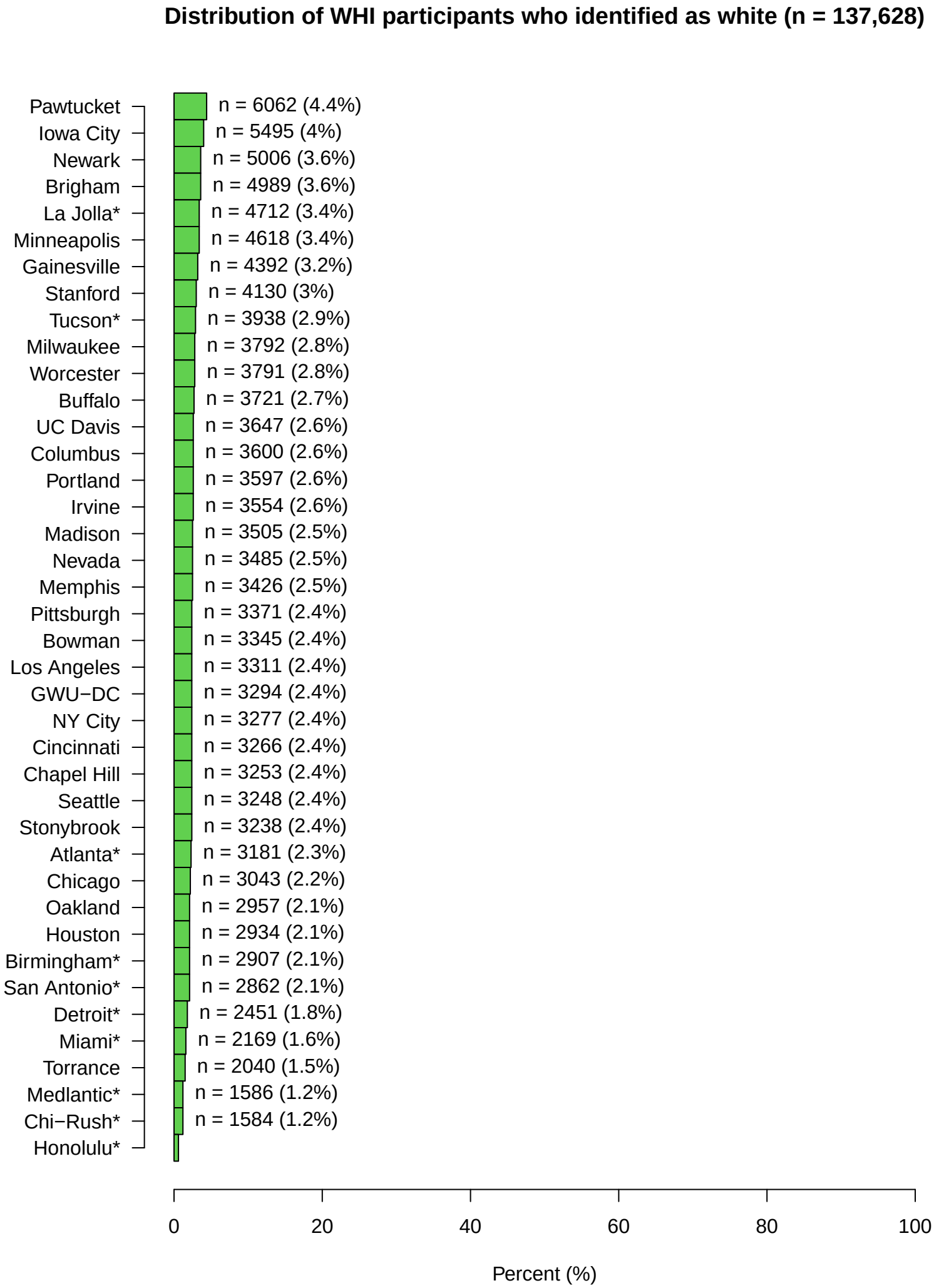


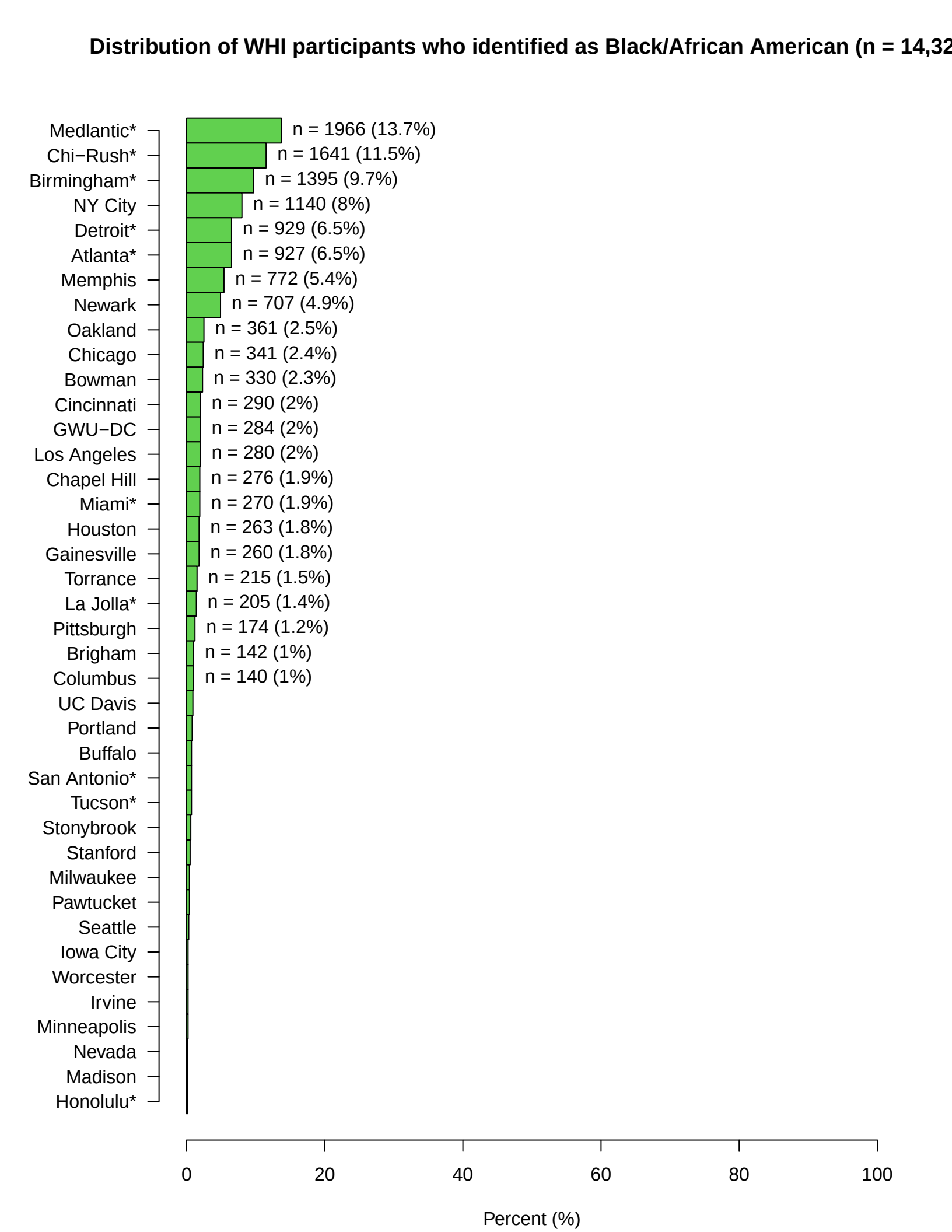
Clinical centers were ordered by highest to lowest total Hispanic/Latina recruitment. * Indicates 'minority recruitment center'.
'No, Not Spanish/Hispanic/Latino' reported by n = 153,117 participants; ethnicity unknown for n = 1379 participants.

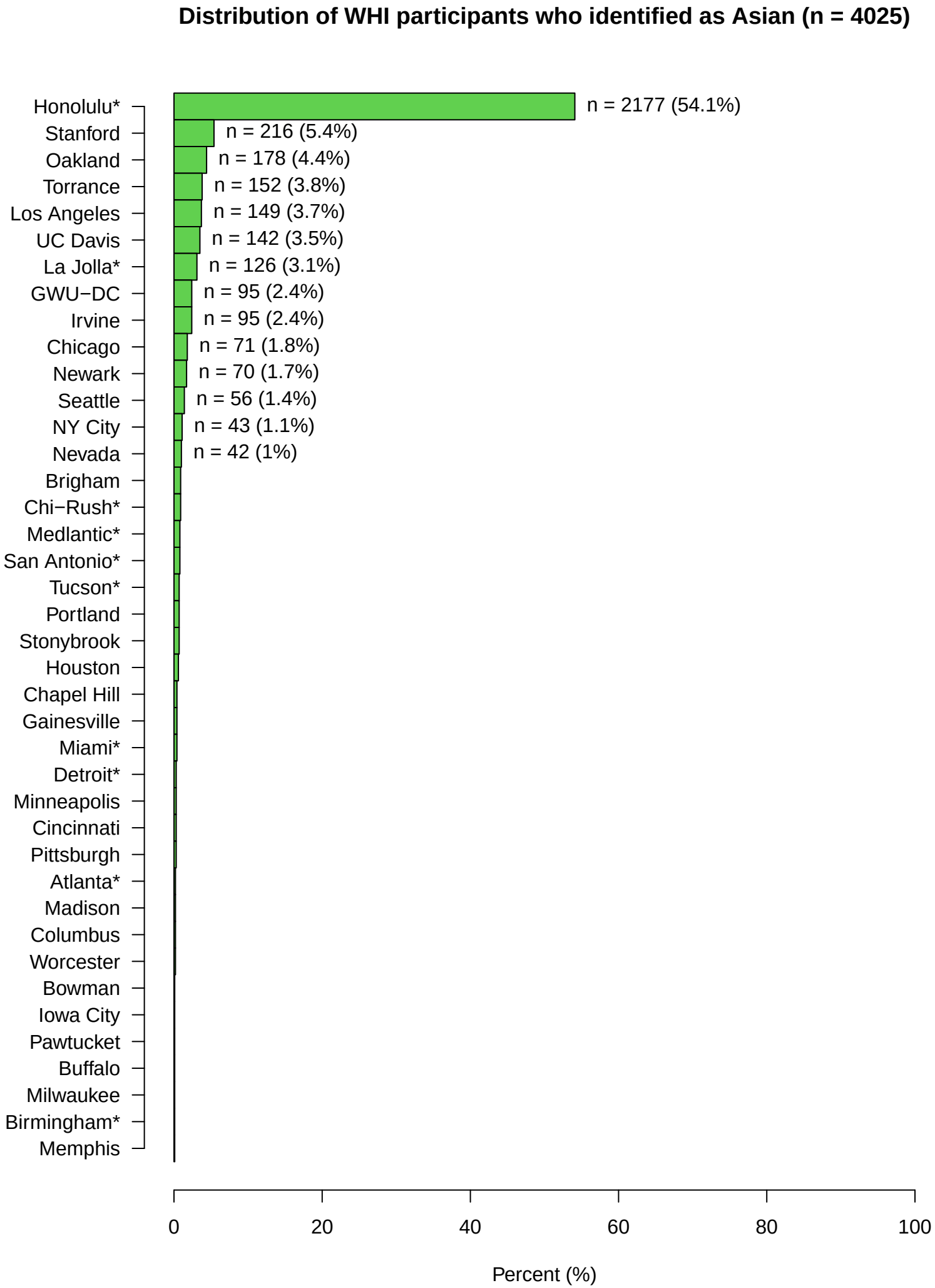
Appendix C3. Distribution of WHI participants who identified as a single race by racial groups (N= 156,657).

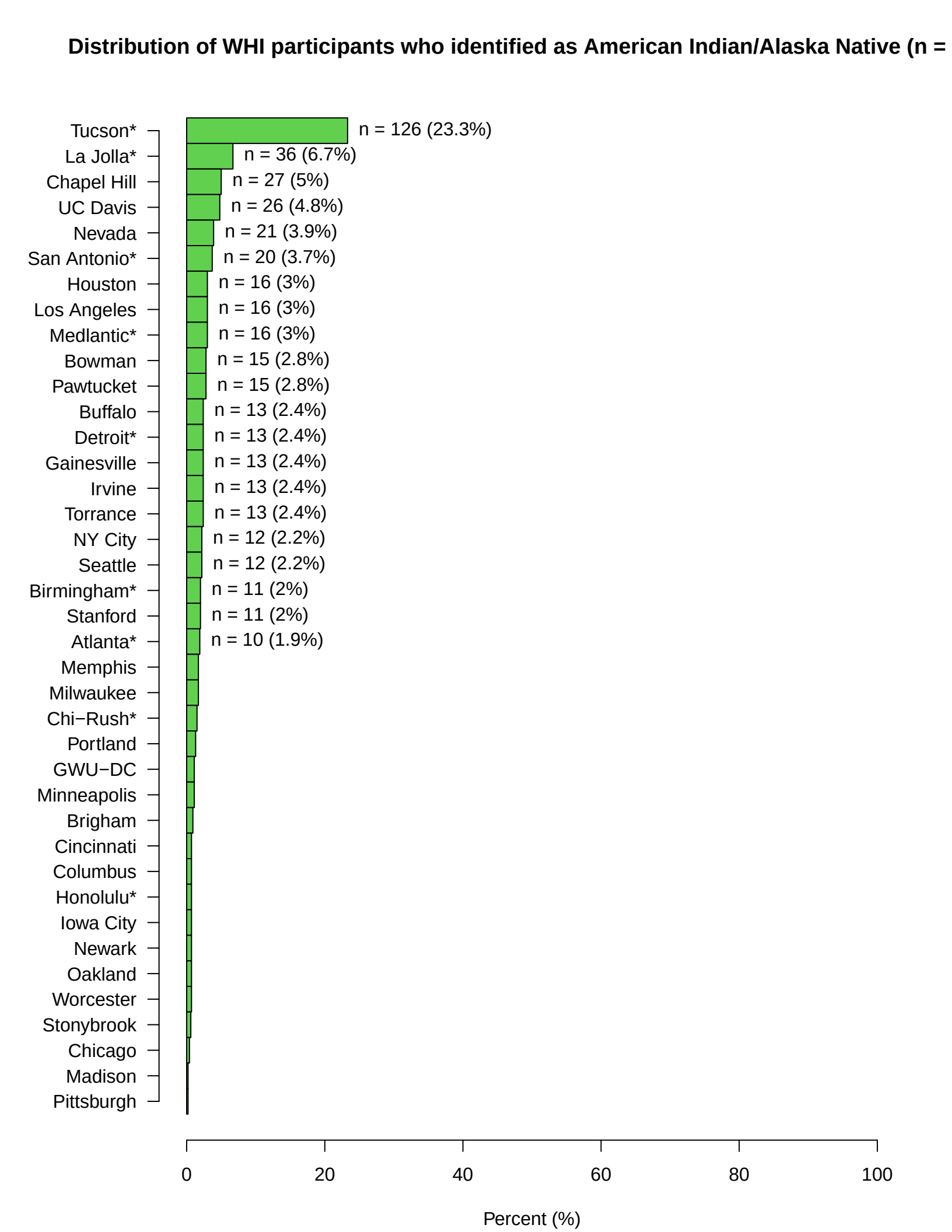


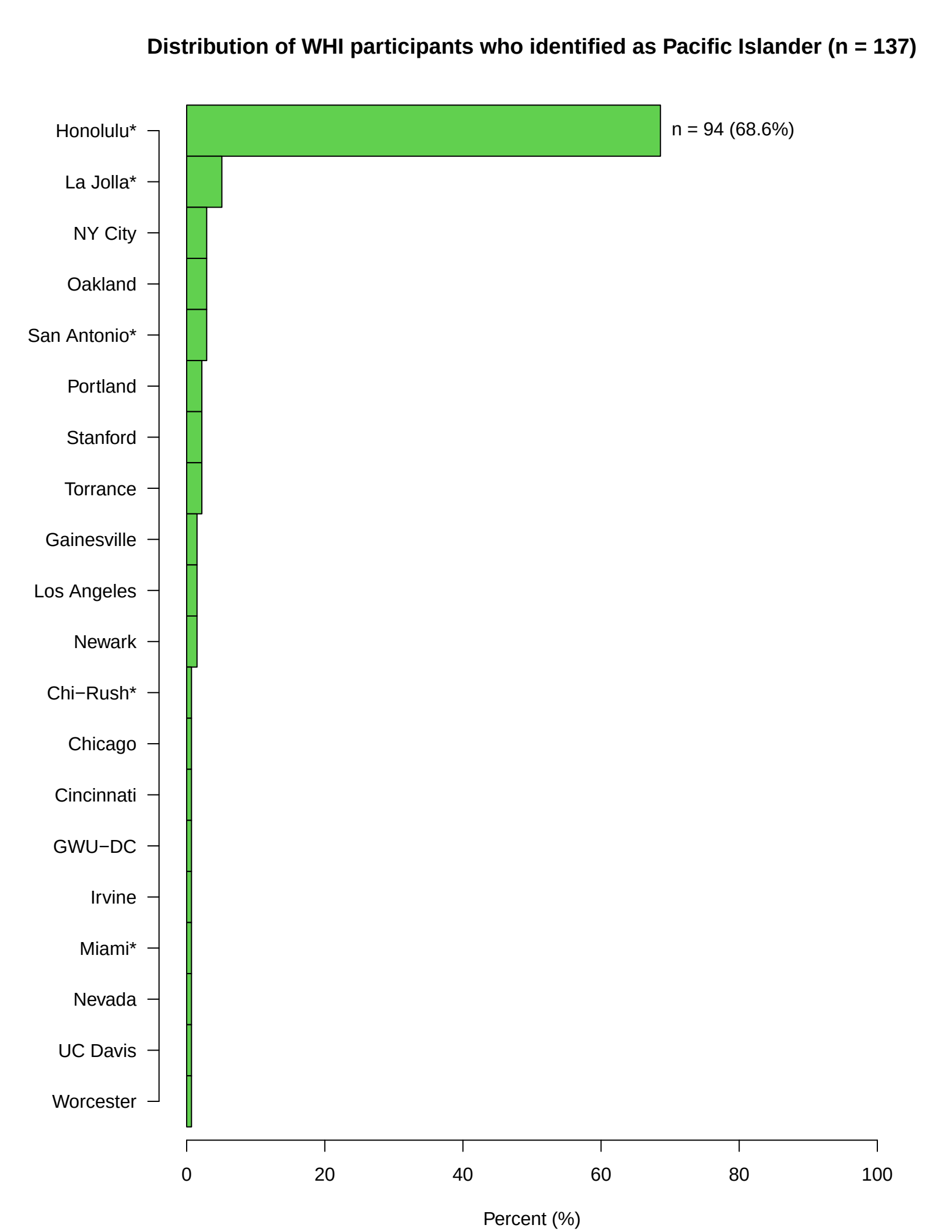
Clinical centers were ordered by total percent of enrollment for targeted racial groups, beginning with 10 'minority recruitment centers', followed by the 30 remaining clinical centers.
More than one race reported by n = 1880 participants; race unknown for n = 3271

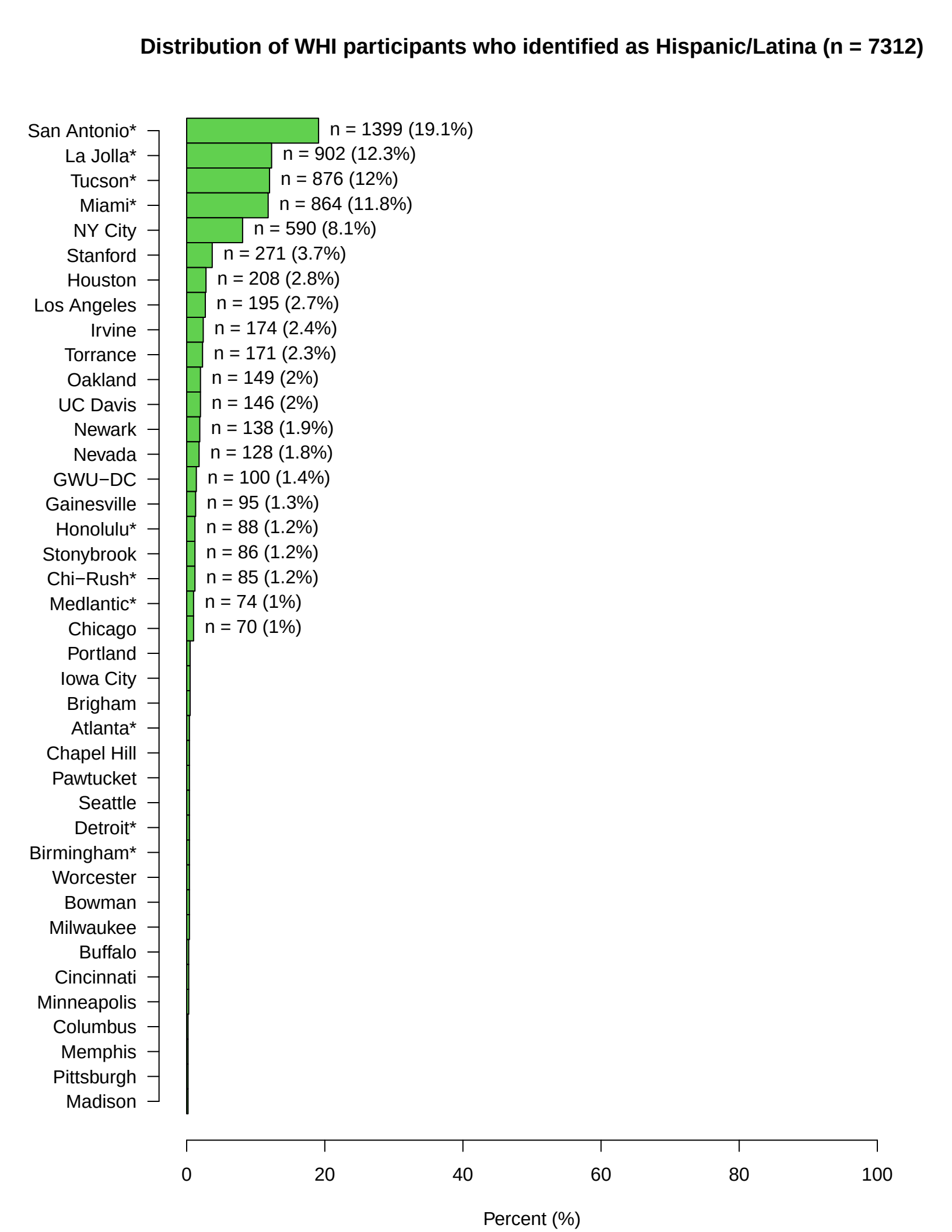

Distribution of WHI participants who identified as white (n = 137,628)
Pawtucket
Iowa City
Newark
Brigham
La Jolla*
Minneapolis
Gainesville
Stanford
Tucson*
Milwaukee
Worcester
Buffalo
UC Davis
Columbus
Portland
Irvine
Madison
Nevada
Memphis
Pittsburgh
Bowman
Los Angeles
GWU−DC
NY City
Cincinnati
Chapel Hill
Seattle
Stonybrook
Atlanta*
Chicago
Oakland
Houston
Birmingham*
San Antonio*
Detroit*
Miami*
Torrance
Medlantic*
Chi−Rush*
Honolulu*
n = 6062 (4.4%)
n = 5495 (4%)
n = 5006 (3.6%)
n = 4989 (3.6%)
n = 4712 (3.4%)
n = 4618 (3.4%)
n = 4392 (3.2%)
n = 4130 (3%)
n = 3938 (2.9%)
n = 3792 (2.8%)
n = 3791 (2.8%)
n = 3721 (2.7%)
n = 3647 (2.6%)
n = 3600 (2.6%)
n = 3597 (2.6%)
n = 3554 (2.6%)
n = 3505 (2.5%)
n = 3485 (2.5%)
n = 3426 (2.5%)
n = 3371 (2.4%)
n = 3345 (2.4%)
n = 3311 (2.4%)
n = 3294 (2.4%)
n = 3277 (2.4%)
n = 3266 (2.4%)
n = 3253 (2.4%)
n = 3248 (2.4%)
n = 3238 (2.4%)
n = 3181 (2.3%)
n = 3043 (2.2%)
n = 2957 (2.1%)
n = 2934 (2.1%)
n = 2907 (2.1%)
n = 2862 (2.1%)
n = 2451 (1.8%)
n = 2169 (1.6%)
n = 2040 (1.5%)
n = 1586 (1.2%)
n = 1584 (1.2%)
0
20
40
60
80
100
Percent (%)


Distribution of WHI participants who identified as Black/African American (n = 14,32
Medlantic*
Chi−Rush*
Birmingham*
NY City
Detroit*
Atlanta*
Memphis
Newark
Oakland
Chicago
Bowman
Cincinnati
GWU−DC
Los Angeles
Chapel Hill
Miami*
Houston
Gainesville
Torrance
La Jolla*
Pittsburgh
Brigham
Columbus
UC Davis
Portland
Buffalo
San Antonio*
Tucson*
Stonybrook
Stanford
Milwaukee
Pawtucket
Seattle
Iowa City
Worcester
Irvine
Minneapolis
Nevada
Madison
Honolulu*
n = 1966 (13.7%)
n = 1641 (11.5%)
n = 1395 (9.7%)
n = 1140 (8%)
n = 929 (6.5%)
n = 927 (6.5%)
n = 772 (5.4%)
n = 707 (4.9%)
n = 361 (2.5%)
n = 341 (2.4%)
n = 330 (2.3%)
n = 290 (2%)
n = 284 (2%)
n = 280 (2%)
n = 276 (1.9%)
n = 270 (1.9%)
n = 263 (1.8%)
n = 260 (1.8%)
n = 215 (1.5%)
n = 205 (1.4%)
n = 174 (1.2%)
n = 142 (1%)
n = 140 (1%)
0
20
40
60
80
100
Percent (%)


Distribution of WHI participants who identified as Asian (n = 4025)
Honolulu*
Stanford
Oakland
Torrance
Los Angeles
UC Davis
La Jolla*
GWU−DC
Irvine
Chicago
Newark
Seattle
NY City
Nevada
Brigham
Chi−Rush*
Medlantic*
San Antonio*
Tucson*
Portland
Stonybrook
Houston
Chapel Hill
Gainesville
Miami*
Detroit*
Minneapolis
Cincinnati
Pittsburgh
Atlanta*
Madison
Columbus
Worcester
Bowman
Iowa City
Pawtucket
Buffalo
Milwaukee
Birmingham*
Memphis
n = 2177 (54.1%)
n = 216 (5.4%)
n = 178 (4.4%)
n = 152 (3.8%)
n = 149 (3.7%)
n = 142 (3.5%)
n = 126 (3.1%)
n = 95 (2.4%)
n = 95 (2.4%)
n = 71 (1.8%)
n = 70 (1.7%)
n = 56 (1.4%)
n = 43 (1.1%)
n = 42 (1%)
0
20
40
60
80
100
Percent (%)


Distribution of WHI participants who identified as American Indian/Alaska Native (n =
Tucson*
La Jolla*
Chapel Hill
UC Davis
Nevada
San Antonio*
Houston
Los Angeles
Medlantic*
Bowman
Pawtucket
Buffalo
Detroit*
Gainesville
Irvine
Torrance
NY City
Seattle
Birmingham*
Stanford
Atlanta*
Memphis
Milwaukee
Chi−Rush*
Portland
GWU−DC
Minneapolis
Brigham
Cincinnati
Columbus
Honolulu*
Iowa City
Newark
Oakland
Worcester
Stonybrook
Chicago
Madison
Pittsburgh
n = 126 (23.3%)
n = 36 (6.7%)
n = 27 (5%)
n = 26 (4.8%)
n = 21 (3.9%)
n = 20 (3.7%)
n = 16 (3%)
n = 16 (3%)
n = 16 (3%)
n = 15 (2.8%)
n = 15 (2.8%)
n = 13 (2.4%)
n = 13 (2.4%)
n = 13 (2.4%)
n = 13 (2.4%)
n = 13 (2.4%)
n = 12 (2.2%)
n = 12 (2.2%)
n = 11 (2%)
n = 11 (2%)
n = 10 (1.9%)
0
20
40
60
80
100
Percent (%)


Distribution of WHI participants who identified as Pacific Islander (n = 137)
Honolulu*
La Jolla*
NY City
Oakland
San Antonio*
Portland
Stanford
Torrance
Gainesville
Los Angeles
Newark
Chi−Rush*
Chicago
Cincinnati
GWU−DC
Irvine
Miami*
Nevada
UC Davis
Worcester
n = 94 (68.6%)
0
20
40
60
80
100
Percent (%)


Distribution of WHI participants who identified as Hispanic/Latina (n = 7312)
San Antonio*
La Jolla*
Tucson*
Miami*
NY City
Stanford
Houston
Los Angeles
Irvine
Torrance
Oakland
UC Davis
Newark
Nevada
GWU−DC
Gainesville
Honolulu*
Stonybrook
Chi−Rush*
Medlantic*
Chicago
Portland
Iowa City
Brigham
Atlanta*
Chapel Hill
Pawtucket
Seattle
Detroit*
Birmingham*
Worcester
Bowman
Milwaukee
Buffalo
Cincinnati
Minneapolis
Columbus
Memphis
Pittsburgh
Madison
n = 1399 (19.1%)
n = 902 (12.3%)
n = 876 (12%)
n = 864 (11.8%)
n = 590 (8.1%)
n = 271 (3.7%)
n = 208 (2.8%)
n = 195 (2.7%)
n = 174 (2.4%)
n = 171 (2.3%)
n = 149 (2%)
n = 146 (2%)
n = 138 (1.9%)
n = 128 (1.8%)
n = 100 (1.4%)
n = 95 (1.3%)
n = 88 (1.2%)
n = 86 (1.2%)
n = 85 (1.2%)
n = 74 (1%)
n = 70 (1%)
0
20
40
60
80
100
Percent (%)