Sex, age, race and intervention type in clinical studies of HIV cure: a systematic review

Johnston RE¹, Heitzeg MM².

¹ amfAR, The Foundation for AIDS Research, New York, NY, USA

² University of Michigan, Dept. Psychiatry

Running title: Demographics & interventions in HIV cure [studies]

Corresponding author:

Rowena Johnston, Ph.D.
amfAR, The Foundation for AIDS Research
120 Wall St, 13th floor
New York, NY 10005
Rowena.johnston@amfar.org
Abstract

*Introduction and objectives:* This systematic review was undertaken to determine the extent to which adult subjects representing sex (female), race (non-White) and age (>50) categories are included in clinical studies of HIV curative interventions and thus, by extension, the potential for data to be analyzed that may shed light on the influence of such demographic variables on safety and/or efficacy.

*Methods:* English-language publications retrieved from PubMed and from references of retrieved papers describing clinical studies of curative interventions were read and demographic, recruitment year and intervention-type details were noted. Variables of interest included participation by sex, age and race; changes in participation rates by recruitment year; differences in participation by intervention type.

*Results:* Of 151 publications, 23% reported full demographic data of study enrollees, and only 6% reported conducting efficacy analyses by demographic variables. Included studies recruited participants from 1991-2011. No study conducted safety analyses by demographic variables. The representation of women, older people and non-Whites did not reflect national or international burdens of HIV infection. Participation of demographic subgroups differed by intervention type and study location. Rates of participation of demographic groups of interest did not vary with time. Limited data suggest efficacy, particularly of early therapy initiation followed by treatment interruption, may vary by demographic variables, in this case sex.

*Conclusion:* More data are needed to determine associations between demographic characteristics and safety/efficacy of curative interventions. Studies should be powered to conduct such analyses and cure-relevant measures should be standardized.
Introduction

Efforts to “cure” HIV - defined as either achieving a state in which the virus is controlled in the absence of medication (sometimes called “functional cure”) or the disabling or removal of all replication-competent virus (sometimes called “sterilizing cure”) - have been active since the 1990s. However, optimism for the potential to cure HIV has increased since the report of the apparent cure of an HIV patient in Berlin after transplantation of stem cells homozygous for the CCR5 delta-32 mutation.

The limited number of cure cases to date makes it impossible to know which intervention(s) will prove most effective at curing HIV, or whether effectiveness will be equivalent across different groups of patients. Little is known about the potential for demographic factors to affect the safety or efficacy of curative interventions. Sex, age or race may influence the establishment or persistence of reservoirs of virus that are not targeted by ART or the immune system, or efforts to disrupt them, via differences and fluctuations in hormones, X versus Y chromosome gene expression, HLA type, response to drugs (e.g. PK or PD), route of acquisition of infection, duration of infection before treatment initiation, immune activation/inflammation, and/or the presence of opportunistic and/or co-infections and morbidities, as outlined below.

Sex differences potentially related to cure

Clinical observations and studies have indicated that female sex is associated with lower viral load, higher CD4 count, more frequent ART side effects and discontinuation, faster progression to AIDS, higher D-dimer levels, greater likelihood of elite control (Steven Deeks, personal communication), more pronounced immune and vaccine responses and higher immune activation. In vitro data,
where ART exposure can be more easily controlled, also suggest possible effects of sex on measures of interest in curing HIV, including infection of and replication in cells, transcription and HIV promoter activity\textsuperscript{26, 28-35}. Biologically-based sex differences, especially in immunity, may also arise from differential gene expression associated with the X chromosome\textsuperscript{19, 36-39} and sex differences may also be associated with access to or outcomes of bone marrow or stem cell transplantation (HCT)\textsuperscript{40-45}.

Evidence suggests that a smaller persistent viral reservoir may be associated with longer time to rebound after therapy interruption (e.g. refs.\textsuperscript{1, 46, 47}). Factors determining size of the reservoir may include duration of infection before ART initiation\textsuperscript{48} or CD4 nadir\textsuperscript{49}. In addition to biological differences between males and females, socio-behavioral factors (such as access to care or level and duration of viral suppression) may also affect these variables and thus the safety or efficacy of curative interventions. A 2009 publication reported that in Europe women were more likely to receive ART than men\textsuperscript{50}. A recent WHO/UNAIDS update extended this finding, by reporting that globally, women are more likely than men to be accessing ART\textsuperscript{51}, likely due largely to increased ART uptake in antenatal care clinics. Recent data issued by the US national government on the HIV continuum of care confirm that while women have slightly higher rates than men of new HIV diagnoses and linkage to care, women and men are equivalent in terms of viral suppression\textsuperscript{52}.

\textit{Age differences potentially related to cure}

Clinical observations indicate that older age is associated with shorter time to death, smaller CD4 increase on ART, lower viral load but faster viral load increase off ART, altered PK and increased immune
activation\textsuperscript{53-61}. Older patients may also have decreased ability to mount immune responses against the virus\textsuperscript{62-67}.

Move up paragraph here Age differences are strongly associated with access to or outcomes of HCT\textsuperscript{42, 68-72}.

Socio-behavioral factors may also be important for age. Older adults in North America are likely to present for care at lower CD4 counts than younger adults\textsuperscript{73}, and the same is true in Europe\textsuperscript{74, 75}.

According to US continuum of care data, linkage to care is lowest in young adults, but rates of viral suppression in those aged 65 and older compare unfavorably to rates in those aged 45-64\textsuperscript{52}.

\begin{flushright}
\textbf{Race differences potentially related to cure}
\end{flushright}

In the context of HIV, race has been associated with differences in plasma viral load\textsuperscript{76, 77}, CD4 count\textsuperscript{6, 76}, rate of CD4 decline\textsuperscript{78}, D-dimer levels\textsuperscript{20}, HLA-associated vaccine response or viral escape\textsuperscript{79, 80}, disease progression and death\textsuperscript{81}, immune and vaccine responses\textsuperscript{82, 83}, and increased systemic immune activation\textsuperscript{84-86}. Move up paragraph here Associations between race and the function of enzymes involved in the metabolism of drugs, including ART, have also been described\textsuperscript{87-91}. Although HLA-B\textsuperscript{*}5701 may be a better predictor of abacavir sensitivity than race\textsuperscript{92}, a study in the UK demonstrated a substantially lower prevalence of the B\textsuperscript{*}5701 HLA type in Africans compared to Whites\textsuperscript{93}. Race differences in pharmacodynamics have also been noted for responses to therapeutic interferon-alpha in HIV/hepatitis C co-infected subjects\textsuperscript{94}. Results from the VaxGen trial indicated that Blacks and Asians generated higher levels of antibodies against HIV than other participants\textsuperscript{95}.
As with sex, race differences may also be associated with access to or outcomes of bone marrow or HCT. African Americans are less likely than Caucasians to undergo HCT,\textsuperscript{40} In a retrospective analysis of more than three thousand patients who received HCT between 1992-2000, Black was the only race associated with significantly higher mortality and grant-versus-host disease (GVHD) than Whites.\textsuperscript{96}

Socio-behavioral differences between races have also been observed. In this context, differences between races considered “majority” – either numerically or, more importantly, according to socioeconomic advantage – versus “minority/other” are of particular interest. In Europe, those not of European origin were more likely to present for care later.\textsuperscript{73, 74} At all points along the HIV continuum of care in the US - particularly in new diagnoses, engagement in care and suppression of viral load - Blacks are doing more poorly than other races.\textsuperscript{52}

**Demographic predictors in hepatitis C cure**

The only viral disease for which a medical “cure” (more commonly referred to as sustained virological response, SVR) is regularly attained is hepatitis C (HCV). Some evidence indicates that demographic characteristics might be associated with variable rates of SVR in HCV. A recent report described treatment outcomes in older (60 years or above) versus younger patients treated with peginterferon alpha-2a and ribavirin. Patients in the older group had to stop treatment more often due to virological failure and adverse events, and were significantly less likely to reach SVR, despite higher compliance and less loss-to-follow up.\textsuperscript{97} Malnick et al. (2014) reviewed several clinical trials of various HCV treatments and found that those aged 65 or older were often excluded from trials, but where data were available, older patients discontinued treatment at a higher rate and/or achieved SVR at a lower rate than younger study participants.\textsuperscript{98} Narciso-Schiavon et al. (2010) reported that although women had more adverse events than men, SVR rates were similar between the sexes.\textsuperscript{99} In those receiving liver transplants and experiencing HCV recurrence, SVR rates in genotype 1 patients were lower in women than men.\textsuperscript{100} More
recently, the use of sofosbuvir has improved SVR rates. In the NEUTRINO and FISSION studies, age (above or below 50) predicted SVR rates with different treatment combinations, and while male sex and black race appeared to be associated with lower SVR, these differences were not significant.\textsuperscript{101}

**Study objectives**

The National Institutes of Health (NIH), the world’s largest single funder of clinical research, has recognized the importance of including representatives of diverse demographics in clinical studies and requires investigators to include, and examine differential effects by sex and diverse racial and ethnic groups in studies conducted with human subjects or on material of human origin such as tissues or specimens.\textsuperscript{102} More recently, they have issued a policy requiring researchers to balance female and male animals and cells in pre-clinical research studies.\textsuperscript{103}

Although sex, age and race are associated with measures thought to be important in curing HIV, via biological and/or socio-behavioral effects, it is currently unknown whether these demographic factors are associated with the safety or efficacy of any curative intervention. This systematic review was undertaken to determine the extent to which adult subjects representing different sex, race and age categories are included in clinical studies of curative interventions and thus, by extension, the potential for data to be analyzed that may shed light on the influence of such demographic factors on safety and/or efficacy.

**Materials and Methods**

**Search methodology**

Clinical intervention studies for which adult males and females were eligible, published through end 2013, whose ultimate goal was to inform efforts to cure HIV either by enhancing control of HIV in the
absence of ART (i.e. “functional cure” or “remission”) and/or perturbing the persistent viral reservoir by disabling, reducing or removing replication-competent virus (i.e. “sterilizing cure” or “eradication”) were included.

PubMed was used to search for relevant English-language publications, and the references of retrieved papers were also scanned. Publications were included if they involved at least one post-hoc analysis of a measure pertinent to curing HIV (e.g. reservoir size) even if other publications referring to the same study stated that the original intent of the study was not cure-related. In order to maximize the number of studies, all clinical studies, not just those that were randomized and/or blinded and/or controlled, were included. Case studies of single patients were not included, nor were conference abstracts. Similarly, explicitly pediatric studies were excluded, but the inclusion criteria of some included studies included participants ages 13 years or older. One publication per study was included in the analysis. When more than one publication referred to a single study, the publication judged to have the most complete demographic data was used.

Variables

Various demographic and other data were collected from each of the analyzed publications. The total number of participants included or enrolled into the study was noted, without regard to whether participants received any or all of the intended interventions or whether they were included in analyses, as the main measure of interest for this systematic review was to assess the access or willingness to participate, and/or the willingness/ability of researchers to recruit, participants who are female, older or non-White. Where data were available, the number of women (including male-to-female transgender), “older” (defined as 50 years or over) and non-white participants included were noted. Reported analyses of safety or efficacy differences in the three demographic variables of interest were also noted.
The year in which the study started recruiting was judged to be a more accurate indication of the timing of the study than the year of publication because of potential variations in time to publication. Where data were not available in the publication, data on recruitment period were gathered on the basis of NCT numbers (or other publications describing the same study), when available. The country(ies) in which the studies took place were also noted.

**Cure interventions**

Cure interventions were classified according to seven categories: ATI, cell therapy, early treatment, immune modulation, intensification, reactivation, and therapeutic vaccine. ATI included studies in which ART was interrupted as an intervention (i.e. to enhance the likelihood of viral control in the absence of ART by e.g. boosting immune responses), or when it was used as a measure of the success of the intervention. Treatment interruption studies in which the stated goal was to save cost or minimize toxicities etc. were not included. Cell therapy included studies of gene therapy and/or cell transplantation. Early treatment involved the initiation of ART during a time period classified by study authors as early in infection. Immune modulation included interventions thought to enhance the function of the immune system independently of classic vaccine responses, including agents that affect the number or function of CD4, CD8 or other immune cells. Intensification was defined as the addition of ART agents beyond standard medical practice. Reactivation studies employed agents such as histone deacetylase inhibitors (and other drug classes) aimed at increasing transcription from proviral DNA. Therapeutic vaccine studies employed vaccines aimed at stimulating the host immune system to generate more effective cellular and/or humoral control of existing infection. Single studies that included more than one category of intervention were assigned to each of the categories described in the study.

**Statistical analyses**
Statistical analyses were performed using SPSS (Version 19.0, Armonk, NY: IBM Corp). Trends across time and differences across interventions were investigated using a general linear model (GLM). Linear regression was used to determine trends across time, with recruitment year (1991-2008) coded as a continuous variable, beginning at zero (year 1991). Linear models were fit to enable detection of increasing or decreasing trends and quadratic models (which included the linear term) were fit to detect curvilinear trends. To determine differences across interventions, one-way analyses of variance (ANOVA) were used with intervention type as the independent variable. The remaining analyses were conducted using non-parametric tests. To determine differences in sex and non-White participation in interventions, one-sample binomial z-tests were conducted using the proportion of total subjects per intervention as the expected value and proportion of women or non-Whites as the observed value. Differences across countries were investigated with Pearson Chi-square tests. For comparisons across the seven interventions, type I error was controlled using a Bonferroni-corrected threshold to establish significance (.05/7=.007). For all other comparisons, significance was established at α<.05.

Results

General

151 publications were retrieved describing 159 studies that were judged non-duplicative. (Six publications described 2 studies, and 1 publication described 3 studies.) Studies were published between 1995-2013. Recruitment periods were available for 76 publications, with start dates ranging 1991-2011. Time from start of recruitment to publication ranged 0-13 years (mean 5.7, median 5). The 159 studies included a total of 14,345 participants. The number of participants per study ranged from 2-4111 (mean 94, median 25). Each study included 1-3 (mean 1.3, median 1) interventions (e.g. patients were treated early in infection, then administered a therapeutic vaccine, followed by a treatment interruption).
The 137 studies conducted within single countries took place in: Argentina (1); Australia (4); Belgium (2); Brazil (2); Canada (4); Denmark (1); France (16); Germany (6); Hungary (1); Italy (5); Japan (1); Netherlands (4); Norway (1); Spain (11); Sweden (4); Switzerland (1); Thailand (2); UK (4); USA (67).

Table 1 provides an overview of publications and studies by intervention, demographics, and number of study countries.

**Interventions**

Of the 14,345 participants in clinical studies, fewer than 1% joined reactivation studies, whereas over 50% were included in immune modification studies. More data are shown in Table 2. While the average number of participants in each type of intervention varied widely (fig 1), a one-way ANOVA found that differences across all seven intervention types were not significant (\(F_{6,198} = 1.0, p = 0.403\)). This is likely due to broad intra-intervention variability.

In terms of the number of participants, Intensification (\(R^2 = 0.277, df = 1, 16, p = 0.025\)) and Tx vaccine (\(R^2 = 0.237, df = 1, 16, p = 0.041\)) showed linear increases from 1991 to 2008, whereas ATI (\(R^2 = 0.552, df = 2, 15, p = 0.002\)) and Immune Mod (\(R^2 = 0.403, df = 2, 15, p = 0.021\)) showed quadratic trajectories. However, only the trajectory for ATI passed threshold for multiple comparisons (\(\alpha = 0.007\)). Studies of Early tx, Cell tx and Reactivation interventions did not change over time (\(p's > 0.21\)). To visualize these effects, years were divided into 3 epochs: pre-cART (up to and including 1995), triple cART (1996-2002) and multi-cART (2003-2008). A fourth epoch, post-Berlin patient (2009-present) was not included in analyses as there were too fewer studies retrieved (consistent with the observation that the mean time from recruitment to publishing was 5.7 years). As Figure 2 illustrates, recruitment into ATI and Immune Mod intervention studies peaked during the triple-cART epoch, whereas recruitment into Intensification and Tx vaccine studies continued to increase over epochs into the multi-cART epoch.

**Demographics – general**
23% of publications provided some detail of the demographics of participants on all three measures – sex, inclusion of older participants, and race. 42% of publications provided details on two demographic variables, 18% provided details on one demographic variable, and the remaining 17% of publications provided no demographic details. See Table 1. No study reported conducting analyses of safety/adverse event outcomes by any of the demographic variables investigated in this systematic review. There were no significant differences in the number of demographic variables reported across intervention types (F<sub>6,190</sub>=0.8, p=0.609), nor were there linear (R<sup>2</sup>=0.158, df=1.16, p=0.102) or quadratic (R<sup>2</sup>=0.163, df=2.15, p=0.264) changes across the timeframe of 1991 to 2008.

**Sex**

Data on the sex of participants was available in 119 (79%) of 151 publications, 125 of 159 (79%) of studies. Of the 12,946 participants in those studies, 2,323 (17.9%) were women. Percent women in studies ranged 0-89% (mean 14%, median 11%). Of the 125 studies where sex data were available, 32 (26%) reported no women in the study. 140 publications described studies conducted entirely within (n=65) or entirely outside (n=75) the USA. There was a trend for publications reporting zero women, or where the number of women was not reported, to be conducted in the United States (χ<sup>2</sup>=3.2, p=0.075; USA 34/65, outside 28/75; fig 3).

Of the 2,323 women included in studies, fewer than 1% participated in cell therapy or reactivation studies. More than 50% of all female participants took part in studies of immune modulation. Compared with all participants, women were less likely to take part in Early tx (z=4.85, p<0.001), Cell tx (z=3.76, p<0.001), Intensification (z=4.48, p<0.001) and Tx vaccine (z=5.23, p<0.001) interventions, and more likely to be in Immune mod (z= 3.53, p<0.001) and ATI (z=5.82, p<0.001) interventions. All survive correction for multiple comparisons (α=0.007). There was no difference for Reactivation (z=1.04, p=0.298) intervention studies (fig 4).
As a median percentage of all participants in each intervention type, studies of immune modulation and reactivation recruited the greatest number of women, while cell therapy studies recruited the fewest. There were no significant differences in percent of woman participants across intervention type ($F_{6,158}=1.1, p=0.356$, fig 5a). However, exploratory independent samples t-tests show differences between intervention types (fig 5b). More data are shown in Table 2.

There were no significant linear ($R^2=0.050$, df=1,16, $p=.372$) or quadratic ($R^2=0.171$, df=2,15, $p=0.245$) changes in percent of women participants from 1991 to 2008.

87 (57.6%) publications reported including at least one woman. Of these, 7 (8%) reported conducting sex difference analyses, of which two reported that there were significant sex differences in efficacy outcomes.

**Age**

Data on the inclusion of participants aged 50 years or older was available in 95 (63%) of 151 publications, 98 (62%) of 159 of studies. The number of older participants was available for only 45 publications and studies. Of these, 21 publications and studies reported no older participants. Percent older participants ranged from 0-60 (mean 13.4, median 8.3). 49 publications (52 studies) reported that participants aged 50 or more were included, but not how many. 6 publications (7 studies) reported conducting age difference analyses, and no differences were found.

Because only 24 publications and studies reported the number of older participants (n=72), no further analyses were conducted on this variable.

**Race**

Data on the number of non-White participants included in studies were available in 46 (30%) of 151 publications, 47 of 159 (30%) of studies. Of the 8,074 participants in those studies, 1,904 (23.6%) were
non-White. Percent non-White in studies ranged 0-60% (mean 25.5%, median 23%). Of the 47 studies where non-White data were available, 4 (8.5%) reported no non-Whites in the study. 140 publications described studies conducted entirely within (n=65) or entirely outside (n=75) the USA. Publications reporting zero non-Whites, or where the number of non-Whites was not reported, were less likely to be conducted in the United States than outside the USA ($\chi^2=12.7$, $p<0.001$; USA 38/65, outside 64/75; fig 3).

Of the 1,904 non-White participants included in studies, none were included in reactivation studies, while more than 70% took part in studies of immune modulation. Compared with all participants, non-Whites were less likely to take part in Early tx ($z=9.32$, $p<0.001$), ATI ($z=9.98$, $p<0.001$) and Tx vaccine ($z=7.27$, $p<0.001$) and more likely to be in Cell tx ($z=2.85$, $p=.004$) and Immune mod ($z=20.45$, $p<0.001$) intervention studies (fig 6). All survive correction for multiple comparisons ($\alpha=0.007$). There was no difference for Intensification ($z=0.44$, $p=0.660$).

As a median percentage of all participants in each intervention type, studies of treatment intensification and treatment interruption recruited the greatest number of non-White participants, while reactivation studies recruited none. Among the six intervention types that did recruit non-White participants, there were no significant differences in percent of non-White participants across intervention type ($F_{5,57}=0.2$, $p=0.943$; fig 7). Furthermore, exploratory independent samples t-tests did not reveal differences between intervention types ($p’s>0.21$). More data are shown in Table 2. Included in these data are two studies (one Early tx and one Tx Vaccine) that had 100% non-White: one was conducted in Thailand and the other in Japan (in which all subjects were either Thai or Japanese, respectively). Because the aim of this review was to identify the rate at which subjects considered “minority” are recruited into studies, we conducted the analyses without these studies and results did not change (ANOVA: $F_{5,55}=0.9$, $p=0.497$; t-tests: $p’s>0.24$).
There were no significant linear ($R^2=0.051$, df=1,14, p=0.403) or quadratic ($R^2=0.180$, df=2,13, p=0.276) changes in percent of non-White participants from 1991 to 2008.

2 (1.3%) publications/3 (1.9%) studies reported conducting race difference analyses, and no differences were found.

Although it would have been interesting to conduct an interaction analysis to test whether combinations of demographic characteristics were associated with any of the variables of interest, almost no publications provided demographic data in sufficient detail to allow such an analysis.

Discussion

Recruitment into clinical studies

Of adults living with HIV in the United States in 2010, 33% were female, 67% were non-White and 19% were older than 55. Worldwide, women account for 50% of all those living with HIV older adults constitute 10% and 85% are living in regions of the world where White is not the predominant race.

Despite the populations in which HIV infection is prevalent, and funder policies requiring representative participation in clinical studies (e.g. ref102), this systematic review suggests that participation in cure studies does not accurately reflect national or international burdens of infection in women, older adults, or non-Whites.

If the safety and efficacy profiles of curative interventions were known to be equivalent across different demographic groups, it may not be as important that studies are conducted largely in young white males. However, this systematic review reveals that the current data are insufficient to draw this
conclusion. Moved from paragraph below: It is currently believed that an HIV cure may initially be more feasible in those whose virus is well-suppressed. In the United States, this tends to be found more often in Whites, equally in males and females, and more often in those aged 55-64. The GRACE (gender, race and clinical experience) trial demonstrated that it is possible to recruit women and racial minorities into HIV clinical research. At this relatively preliminary time in the development of curative strategies, it will be crucial to understand if differences associated with demographics exist.

Are there demographic differences?

Although many data points were missing in the current review, making broader generalizations difficult, some interesting observations emerged. Two publications noted sex differences in intervention efficacy, and both were studies of early treatment followed by treatment interruption. Desquilbet et al. (2004) described 58 patients, including 13 women, in the French PRIMO cohort who started therapy during acute (none or one Western blot band) or early (two or more Western blot bands) infection. There was a median 1.5 years of viral suppression before patients elected to stop therapy for various reasons. At 12 months post-interruption, women had significantly lower viral loads (by a log) than men. As previously described in other studies, women also had lower baseline viral loads, but even when adjusted for baseline viral load, women still had lower off-therapy viral loads. Hoen et al. (2005) conducted the PRIMSTOP study, also in France, in 26 patients (5 women) who initiated ART (plus hydroxyurea) for 34 weeks during primary infection (3 or fewer Western blot bands), followed by alternating periods of treatment interruption and ART. In a multivariate analysis, female sex was the only variable that
predicted virological success (defined as VL < 1000) 6 months after ART discontinuation. Virological success was also associated with greater gains in CD4 T cells. As in other studies, these authors also noted that women also started with lower baseline viral loads\textsuperscript{111}.

Replicating these studies of early treatment in women in other settings may prove challenging. It is commonly believed that women are rarely diagnosed and treated during early infection, although national data confirming this appear to be lacking. However, data from cohorts and studies indicate that such women are in fact available. Data from the Duke-UNC Acute HIV Infection Research Consortium indicate that 12% of the cohort is female\textsuperscript{112}, somewhat lower than the 22% of HIV cases in females in North Carolina in 2012\textsuperscript{113}. Similarly, the primary infection Options cohort in San Francisco consists of about 4% women\textsuperscript{114}, reflecting the low contribution (about 8%\textsuperscript{115}) of women to that city’s epidemic. However, a 2006 publication\textsuperscript{116} reported that 33% of patients who enrolled in the Acute Infection and Early Disease Research Program (AIEDRP) at the site in Baltimore, Maryland were female, close to the 35% of all HIV cases in Baltimore in 2014\textsuperscript{117}. In France, in the PRIMO cohort consisting of 1267 patients enrolled during primary HIV infection, 16% are women\textsuperscript{118}, lower than the national 29% of all HIV cases\textsuperscript{119}. Although not necessarily reflective of national averages, the SPARTAC trial conducted across 8 countries (United Kingdom, South Africa, Brazil, Uganda, Spain, Australia, Italy, Ireland) consisted of a cohort of 371 subjects identified during primary infection, of whom 40% were women\textsuperscript{120}.

It is possible that some combination of demographic variables may also have been associated with participation or efficacy. For example, it may be possible that a certain intervention is more efficacious in Black females, or that older males are more or less likely to participate in studies than other subjects. Study subjects can be described by, and embody the characteristics of, multiple demographic and other
descriptors. Data available in the publications described in this review did not often enough provide the level of detail that would be required to conduct such analyses.

**Factors associated with participation of under-represented groups**

Moved from recruitment section above: Historically, women, older people and non-whites have been under-represented in HIV clinical studies and many reasons for this have been invoked, ranging from medical mistrust to poverty to childcare and employment challenges. In addition, care provider attitudes towards and perceptions of female, older and non-White patients also tend to be more negative, which may negatively influence the likelihood of solicitation into clinical research. Taking these factors together, it is perhaps not surprising that women, older people and non-Whites are under-represented in HIV cure-related clinical studies. Despite NIH guidelines, there was a tendency for publications reporting zero female participants, or where the number of women was not reported, to be conducted more often inside than outside the US. However, studies including non-Whites were significantly more likely to take place inside the US than outside. The participation of these two groups in clinical studies of cure interventions has not changed in the period 1991-2008. It should be noted that it is conceivable that some studies may be relevant only to circumscribed populations. In an attempt to correct for this bias, this study undertook to review only those studies for which both men and women were eligible. However, there may be limitations, such as CCR5 genotype, that might guide the exclusion of some types of patients, especially non-Whites, in whom the delta32 allele would not be expected to be present.

It is interesting to note that women and non-Whites are more and/or less likely than participants overall to take part in certain types of cure studies. Women and non-Whites were both more likely to take part in immune modulation studies when compared to all participants. It is difficult to know whether such
studies are more appealing to these participants or if so, why. From the opposite perspective, immune modulation and reactivation studies had a greater percent women included than other types of studies, while studies of treatment intensification and treatment interruption recruited more non-Whites than did other types of studies. Again, it is difficult to understand the differences in the motivations or behaviors of the researchers recruiting participants to these studies. One might speculate that physicians or health care providers serving as a link between these populations and different types of cure studies may make assumptions about health literacy or self-efficacy associated with sex or race and that these assumptions may either guide the ways in which they present studies to potential participants (leading to differences in the appeal of various kinds of studies), or the extent/range of efforts they employ to recruit such participants (leading to differences in percent participation). For reasons that would be more difficult to understand, there may also be systematic differences between study types in methods or locations of advertising and recruitment efforts (e.g. inner city clinics versus college campuses). There was no apparent relationship between the geographic location of particular study types, either between or within countries, and the likelihood of inclusion of women or non-Whites.

**Potential to generate data on demographics**

Despite the general paucity of analyses by demographics in this dataset, it may nonetheless be possible – with access to the raw data - to conduct post-hoc analyses (or meta-analyses), even if only descriptive (especially in those studies with insufficient participant numbers to conduct formal analyses). Many of the studies in this dataset anecdotally described unusual safety or efficacy outcomes, but rarely provided demographic data of those participants. It is remarkable that no studies conducted analyses by demographics on safety outcomes.
Analyses of biomedical outcomes according to demographic variables can be challenging to interpret. Sex, although arguably the most clear-cut of the three demographic variables investigated here because of its categorical nature, can be complicated by age (e.g. pre- versus post-menopausal) and/or use of hormonal contraceptives. For this systematic review, 50 years of age was chosen as a cut-off for “older” participants, to mirror many epidemiological reports. However, age is a continuous variable with no major biological differences expected in relevant measures in individuals close to the cut-off on each side. Age may also have very different associations with outcomes depending on whether the participant acquired HIV at a younger age and has aged with the virus, or whether the virus was acquired in older age. Race is potentially the most complicated of these variables. For these analyses, White vs. non-White was chosen as the comparator, but there is little a priori reason not to choose some other category demarcation, for example, Black vs non-Black. In fact, more genetic variation among Blacks has been reported than among all other races\textsuperscript{131}, so the latter analysis may have more biological relevance. In the context of biomedical HIV research, HLA may be better related to safety and/or efficacy than race, especially for interventions that require an immune response associated with particular HLA types. However, reporting the race of participants would still provide a useful indicator of the access to or willingness to participate in cure studies. In order to conduct such analyses, participants representing each of these demographics will need to be recruited to studies in adequate numbers. Investigators interested in probing associations between demographic variables and safety or efficacy will face many challenges.

Limitations
Although steps were taken to identify duplicates, we cannot rule out the possibility that some publications described studies that were at least partially overlapping with others, and as such some participant groups may be over-represented. There may also be other publications describing studies that fit the inclusion criteria that we did not identify, possibly resulting in under-representation of other participant groups. In addition, it is possible that in cases where single studies were described in multiple publications, the publications not included in this systematic review may have included demographics analyses that were not identified, although efforts were made to minimize this. Finally, identifying clinical studies that fit the stated cure definition was in many cases a judgment call, and others may have classified the included and excluded publications differently.

One of the greatest limitations of this systematic review was the lack of data concerning some key variables. Data on the number of women, older people and non-Whites were missing for 21%, 84% and 70% of publications respectively, limiting our ability to analyze the associations between these variables and participation in studies of different interventions over time. In some studies where these demographics were reported, no participants in those categories were included, further limiting the analysis. In addition, the year in which recruitment started was available for only 50% of publications. Despite these limitations, significant differences were still found on a number of measures. It is not possible to know whether the results reported in this systematic review would change if these data were available.

**Recommendations**

Several recommendations emerge from this study: 1) study authors should report at least basic (sex, age, race) demographic data of those who enroll into studies and those whose data are analyzed; 2) outcomes – both safety and efficacy – should be analyzed by demographic variables and results of such
analyses should be reported; 3) to the extent possible, studies should be powered to explore demographic differences; 4) when studies are not sufficiently powered to allow such analyses, meta-analyses among similar interventions may provide important information; 5) to facilitate such meta-analyses, standardized endpoint measures should be taken, including for example changes in viral load, CD4 count, and proviral DNA, at a minimum; and 6) animal studies may provide preliminary indications of potential demographic differences, at least concerning sex and age.

Acknowledgments

The authors would like to thank Shirin Heidari, Samuela Rossi, Mirjam Cuomo, Matt Price and Ioannis Hodges-Mameletzis for thoughtful discussion on participation of women in HIV research. The authors also thank Jeffrey Laurence and Greg Millett for feedback on an earlier version of this manuscript.

The authors declare that they have no conflicts of interest to report.

Reprint requests:

Rowena Johnston

amfAR, The Foundation for AIDS Research

120 Wall St, 13th floor

New York, NY 10005

USA
REFERENCES


7. Delmas MC, Jadand C, De Vincenzi I, et al. Gender difference in CD4+ cell counts persist after HI


36. Pinheiro I, Dejager L, Libert C. X-chromosome-located microRNAs in immunity: might they explain male/female differences? The X chromosome-genomic context may affect X-located


73. Althoff KN, Gebo KA, Gange SJ, et al. CD4 count at presentation for HIV care in the United States and Canada: are those over 50 years more likely to have a delayed presentation? *AIDS Res Ther.* 2010;7:45.


114. Hecht FM. Personal Communication. UCSF School of Medicine.


FIGURE LEGENDS
Figure 1. Mean number of participants in each type of intervention. Error bars indicate ±1 standard error.
Figure 2. Number of studies of each intervention type by epoch. The recruitment year variable was divided into 4 epochs: pre-cART (up to and including 1995); triple cART (1996-2002); multi-cART (2003-2008); post-Berlin patient (2009 to present). The most recent epoch was not included in this analysis as there were too few studies included in this review. *p<.05, **p<.007.
Figure 3. Participation of women and non-Whites in studies conducted entirely within or outside the United States. + p<0.08, **p<0.001.
Figure 4. Sex differences in participation by intervention. One-sample binomial z-tests were conducted using the proportion of total subjects per intervention as the expected value (black bars) and proportion of women as the observed value (gray bars). Proportions are represented as percentages. **p<0.007.
Figure 5. A. Mean percent of women recruited into each type of intervention. Error bars indicate ±1 standard error. Lines indicate significant (at p<.05) differences between interventions from pairwise t-tests. B. P-values of pairwise t-tests showing differences between interventions in the mean percent of women participants. *p<.05, **p<.007.
Figure 6. Race differences in participation by intervention. One-sample binomial z-tests were conducted using the proportion of total subjects per intervention as the expected value (black bars) and proportion of non-Whites as the observed value (gray bars); proportions are represented as percentages.

**p<0.007.
Figure 7. Mean percent of non-Whites recruited into each type of intervention. Error bars indicate ±1 standard error. No non-Whites were reported as participants in Reactivation interventions. Pairwise t-tests revealed no significant differences between interventions.
Table 1. Overview of publications and studies by demographics, intervention type and study countries.

<table>
<thead>
<tr>
<th>Intervention</th>
<th># publications (of 151)</th>
<th>% pubs</th>
<th># studies (of 159)</th>
<th>% studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>early treatment</td>
<td>34</td>
<td>22.5</td>
<td>37</td>
<td>23.3</td>
</tr>
<tr>
<td>cell therapy</td>
<td>17</td>
<td>11.3</td>
<td>17</td>
<td>10.7</td>
</tr>
<tr>
<td>immune modification</td>
<td>38</td>
<td>25.2</td>
<td>42</td>
<td>26.4</td>
</tr>
<tr>
<td>treatment intensification</td>
<td>18</td>
<td>11.9</td>
<td>19</td>
<td>11.9</td>
</tr>
<tr>
<td>reactivation</td>
<td>6</td>
<td>4.0</td>
<td>6</td>
<td>3.8</td>
</tr>
<tr>
<td>treatment interruption&lt;sup&gt;c&lt;/sup&gt;</td>
<td>51</td>
<td>33.8</td>
<td>54</td>
<td>34.0</td>
</tr>
<tr>
<td>therapeutic vaccine</td>
<td>40</td>
<td>26.5</td>
<td>41</td>
<td>25.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demographics</th>
<th># publications (of 151)</th>
<th>% pubs</th>
<th># studies (of 159)</th>
<th>% studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>women</td>
<td>32</td>
<td>21.2</td>
<td>34</td>
<td>21.4</td>
</tr>
<tr>
<td>none</td>
<td>32</td>
<td>21.2</td>
<td>32</td>
<td>20.1</td>
</tr>
<tr>
<td>at least 1 participant</td>
<td>87</td>
<td>57.6</td>
<td>93</td>
<td>58.5</td>
</tr>
<tr>
<td>older&lt;sup&gt;d&lt;/sup&gt;</td>
<td>56</td>
<td>37.1</td>
<td>61</td>
<td>38.4</td>
</tr>
<tr>
<td>insufficient data</td>
<td>56</td>
<td>37.1</td>
<td>61</td>
<td>38.4</td>
</tr>
<tr>
<td>none</td>
<td>56</td>
<td>37.1</td>
<td>61</td>
<td>38.4</td>
</tr>
<tr>
<td>at least 1 participant</td>
<td>87</td>
<td>57.6</td>
<td>93</td>
<td>58.5</td>
</tr>
<tr>
<td>non-White</td>
<td>105</td>
<td>69.5</td>
<td>112</td>
<td>70.4</td>
</tr>
<tr>
<td>insufficient data</td>
<td>105</td>
<td>69.5</td>
<td>112</td>
<td>70.4</td>
</tr>
<tr>
<td>none</td>
<td>4</td>
<td>2.6</td>
<td>4</td>
<td>2.5</td>
</tr>
<tr>
<td>at least 1 participant</td>
<td>42</td>
<td>27.8</td>
<td>43</td>
<td>27.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># demographic variables reported&lt;sup&gt;e&lt;/sup&gt;</th>
<th># publications (of 151)</th>
<th>% pubs</th>
<th># studies (of 159)</th>
<th>% studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>16.6</td>
<td>27</td>
<td>17.0</td>
</tr>
<tr>
<td>1</td>
<td>27</td>
<td>17.9</td>
<td>29</td>
<td>18.2</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
<td>42.4</td>
<td>68</td>
<td>42.8</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>23.2</td>
<td>35</td>
<td>22.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study country</th>
<th># publications (of 151)</th>
<th>% pubs</th>
<th># studies (of 159)</th>
<th>% studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>single</td>
<td>133</td>
<td>88.1</td>
<td>137</td>
<td>86.2</td>
</tr>
<tr>
<td>multiple</td>
<td>14</td>
<td>9.3</td>
<td>16</td>
<td>10.1</td>
</tr>
<tr>
<td>no data</td>
<td>4</td>
<td>2.6</td>
<td>6</td>
<td>3.8</td>
</tr>
</tbody>
</table>

<sup>a</sup> number of studies is greater than number of publications because 6 publications reported on 2 studies and 1 publication reported on 3 studies

<sup>b</sup> interventions do not add up to total numbers of publications and studies because many publications and studies included more than one type of study

<sup>c</sup> included studies where treatment interruption was an intervention or was used to evaluate success of an intervention

<sup>d</sup> defined as 50+ years of age

<sup>e</sup> among: number women; whether included participants 50+ yrs old; number of non-White race
Table 2. Participants per intervention type.

<table>
<thead>
<tr>
<th>Intervention type</th>
<th># participants</th>
<th>Median # participants</th>
<th>Median # women (# studies missing data)</th>
<th>Median # non-White (# studies missing data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>early treatment</td>
<td>3285</td>
<td>34</td>
<td>5(5)</td>
<td>8 (23)</td>
</tr>
<tr>
<td>cell therapy</td>
<td>245</td>
<td>8</td>
<td>0 (5)</td>
<td>4 (9)</td>
</tr>
<tr>
<td>immune modulation</td>
<td>7516</td>
<td>30</td>
<td>5 (10)</td>
<td>7 (30)</td>
</tr>
<tr>
<td>treatment intensification</td>
<td>394</td>
<td>13</td>
<td>1 (2)</td>
<td>13 (14)</td>
</tr>
<tr>
<td>reactivation</td>
<td>113</td>
<td>11</td>
<td>3 (1)</td>
<td>n/a (6)</td>
</tr>
<tr>
<td>treatment interruption</td>
<td>4566</td>
<td>40</td>
<td>7 (9)</td>
<td>14 (36)</td>
</tr>
<tr>
<td>therapeutic vaccine</td>
<td>2381</td>
<td>36</td>
<td>3 (8)</td>
<td>5 (24)</td>
</tr>
</tbody>
</table>