2022

Representativeness of Data from an Online Sample of Individuals with Severe Alcohol Use Disorder

Michelle Eglovitch

Representativeness of Data from an Online Sample of Individuals with Severe Alcohol Use Disorder

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

by
Michelle Eglovitch, MPH
Boston University School of Public Health

Director: Dace S. Svikis, Ph.D.
Professor, Department of Psychology

Virginia Commonwealth University
Richmond, Virginia
December, 2022

Acknowledgements

I would like to thank my advisor and committee chair, Dr. Dace Svikis, for her guidance, mentorship, and cheerleading throughout this process. I would also like to thank my other committee members, Drs. Jared Keeley and Alexis Edwards, for their advice throughout the completion of this thesis. I would also like to acknowledge all of my friends and family who supported me every step of the way. I would like to thank my sweet cats, Olive and Popcorn, for demanding attention and treats at the exact moment I needed a break from writing. Finally, to my husband and best friend, Bobby, for the constant support and love.

Table of Contents

Abstract

Representativeness of Data from an Online Sample of Individuals with Severe Alcohol Use Disorder

By Michelle Eglovitch, MPH

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2022

Director: Dace S. Svikis, Ph.D., Professor, Department of Psychology

Understanding the characteristics of research participants who provide DNA is crucial to ensuring sample representativeness and generalizability of findings of genetics studies of substance use disorders. Using anonymous survey data, the present study had a unique opportunity to compare participants who provided DNA (n=2,414) with those that did not provide DNA (n=1,515). The present study found numerous differences between groups across demographic, substance use, and personality domains. Using multivariate regression, the most parsimonious model found that being male, a non-smoker, and endorsing perseverance was associated with providing DNA. The present study provides benchmark data on sample representativeness in a genome wide association study.

Representativeness of Data from an Online Sample of Individuals with Severe Alcohol Use Disorder

Introduction

While sample representativeness has received attention in clinical trials, it has been less of a focus in genetics studies of psychiatric disorders, including substance use disorders (SUDs). There is a need for representativeness in research, as the conclusions drawn from research are applicable in clinical settings only to the extent that the participants resemble the patient population of interest (Humphreys et al., 2013). Without adequate representation across demographic characteristics and other important factors, the results derived from such studies are only applicable to the populations sampled. Research demonstrates that several demographic factors are associated with lack of participation in clinical trials, within categories such as race, age, gender, physical and mental health status, and socioeconomic status (SES). However, factors beyond demographics, such as personality, are not well described in the literature as they relate to participation in research. Furthermore, it is unclear if these issues of representation extend beyond clinical trials to other types of research, such as survey and genetics studies.

Representativeness becomes even more complex with the advent of biospecimens, such as saliva. Biospecimens are often used to help corroborate self-reported health behaviors, assess severity of illness, and provide other metrics of health (Sakshaug et al., 2014). Initially, biospecimens were collected in a clinical setting as part of research study procedures. This ensured participation as well as safety, but such procedures were limiting in the ability to reach populations broadly, as the provision of in-person biospecimens requires time and resources on the behalf of the participant. Thus, there began a broader shift to collecting biospecimens, particularly ones that provide DNA, using mail-in procedures (Sakshaug et al., 2014).

However, the evidence on who provides mail-in biospecimens as part of genetics study procedures is limited, and the data that is available is mixed: some studies have found significant differences by factors such as race, age, and gender, where others have not found any significant group differences (see Table 1). In addition, there is a paucity of data on this topic in individuals with alcohol and other drug use disorders. Research on the factors associated with participation and nonparticipation in the provision of DNA related in studies of such disorders is scarce, and much of the literature on responsiveness to requests of DNA samples have been focused predominantly on smoking and cancer. Representativeness becomes an even more salient issue when considering evolving technology and the mapping of the human genome. Providing DNA now raises concerns over issues of privacy and autonomy in providing such samples, and concerns of an ability to control access and use over such data (Clayton et al., 2019).

The purpose of the proposed study was to identify demographic, psychosocial and personality variables associated with provision of saliva DNA specimens as part of a genome wide association study (GWAS) of severe alcohol use disorder (AUD). The sample consists of N=3,927 individuals recruited online through Facebook. All participants met DSM-5 criteria (lifetime) for severe AUD and completed an online survey about their alcohol and other drug use and related variables. Upon finishing the survey, individuals were mailed a kit for collection of a saliva DNA sample and were asked to return the sample in a pre-addressed, postage-paid envelope with a $10 electronic gift card (e-gift card) as compensation for study participation. The proposed research is important, as only 61.4% of survey completers returned their saliva samples and there is a paucity of data on the representativeness of GWAS participants; particularly for socially stigmatizing conditions such as AUD. To our knowledge, this is one of the first studies investigating predictors of biospecimen provision within this specific population.

The specific aims of the proposed research were as follows:

Specific Aim 1: Describe the demographic, psychosocial and personality characteristics of the individuals with severe AUD who completed the on-line survey and agreed to provide a saliva DNA sample.

Specific Aim 2: Compare GWAS survey completers who did (n = 2,412) and did not (n = 1,515) subsequently provide a DNA saliva sample on a variety of demographic, psychosocial and personality variables using univariate analyses. Based on the literature, the following hypotheses will be tested:

H1: Individuals self-identifying as non-white, racial minorities will be less likely to return DNA samples than individuals self-identifying as white.

H2: Men will be less likely to return DNA samples than women.

H3: Individuals with a lower SES will be less likely to return DNA samples than individuals of higher SES.

Specific Aim 3: Taking the demographic, psychosocial and personality variables for which univariate analyses detected group differences (or trends at the $p < .05$ level), use a multivariate logistic regression to identify the most parsimonious model.

Review of the literature

Research recruitment and enrollment

History

Historically, the typical human research participant was a white male. Female subjects were generally excluded from biomedical research and members of minoritized groups were often underrepresented in research as well (Bennett, 1993). This did not change until 1993, when the National Institutes of Health, through the NIH Revitalization Act, mandated that

women and minoritized groups be included in all government-funded clinical research and randomized control trials (RCTs) (Freedman et al., 1995). While change has been slow, progress has been made.

In addition to actively recruiting women and minority group members, the NIH Revitalization Act and other policies focused attention on sample bias, and the extent to which RCT participants adequately represented the broader target populations from which they were ascertained. Interest was kindled in research comparing individuals who enrolled in clinical trials to those who chose not to participate and strategies for recruiting representative samples.

Variations in sample characteristics by recruitment method have been reported in studies ranging from cigarette smoking (Harris et al., 2003) to Alzheimer's disease (Andersen et al., 2010). Harris and colleagues (2003), for example, found that subjects recruited using reactive methods (e.g., flyers encouraging potential participants to call a study line) had higher levels of education and income, and reported being in better health than subjects recruited using proactive methods (e.g., in-person recruitment by research staff and health care providers). Similarly, in a study of Alzheimer's disease, patients recruited with reactive methods (e.g., by mail) were more likely to be younger and male, with greater self-reliance and higher mini mental state exam (MMSE) sum scores than patients recruited using proactive methods (Andersen et al., 2010).

Factors associated with participation and non-participation in survey research

While the issue of sample representativeness has received considerable attention in clinical research, the focus has not been limited to RCTs and it is important to look at this issue across a broader array of study designs. Survey research, for example, has long been a convenient, inexpensive, and reliable way to gather cross-sectional as well as longitudinal data. Thus, despite such advances in recruitment strategies and other efforts to diversify the

representativeness of samples, it is unknown if such efforts to expand reach translates to the representativeness of who replies to surveys and why. Survey non-response behavior is complex and poorly understood (Smith, 2008). Further, survey non-response can be influenced by a variety of variables, including method of recruitment, survey length and format, and use of follow-up reminders (Sheehan, 2001).

Taking such methodological diversity into account, little is known about who elects to participate in a study and who does not. The representativeness of survey research participants has public health and clinical implications, as little is known about the characteristics of those who consent to research participation, and how accurately they reflect the overall population of interest. Most studies do not report the characteristics of participants who were approached about such studies but elected not to participate. While limited, the following section summarizes what is currently known about demographic variables and their association with who does and does not consent to survey research participation.

Socio-demographics

Race: In survey research, non-white racial minority populations may participate in research at lower rates when compared to white populations (Voigt et al., 2002; Triplett et al., 1996). This trend is reflected in other domains of research, such as clinical trials (Anwuri et al., 2013; George et al., 2014). These racial differences in the willingness to participate is due to many reasons, amongst them being systemic issues around medical mistrust. Such mistrust is notably rooted in historical events, a prominent example being the US Public Health Services (USPHS) Syphilis Study at Tuskegee (Tuskegee Study), in which treatment was withheld for Black patients during the study, although penicillin was largely available. Medical mistrust may also be rooted in the perception that such research will primarily benefit white populations, a fear

of purposeful mistreatment, and the perception that by signing the informed consent they are relinquishing their rights and providing the researcher with legal protection against any harm that may be inflicted on the participants (George et al., 2014).

SES: Individuals of lower SES are less likely to consent to research participation than higher SES individuals. Low SES, measured by various metrics such as educational attainment, income status, employment status, insurance coverage, housing situation, and others, have consistently been demonstrated to have a negative correlation with research participation (Ford et al., 2008; Galea & Tracey, 2007). This trend is reflected in national household surveys such as the Wisconsin Longitudinal Study (WLS), where level of education was positively correlated with rate of survey participation (Dykema et al., 2017).

Gender: Gender is another sociodemographic factor associated with consenting to research participation. Assuming a binary of men vs. women as identified on surveys, many studies have demonstrated that women are more likely to participate in scientific studies than men (see Galea & Tracy, 2007 for a list of such studies). Several studies done with survey research have found similar trends, with women participating at higher levels than men (Hauser, 2005; Smith, 2008).

Several theoretical frameworks have been proposed for why this gender disparity exists. England (1989) posits that males are more likely to possess or place a high value on separative characteristics than females, while females, on the other hand, are more likely to value characteristics more consistent with connective selves, such as empathy or emotional closeness. This would indicate that women are more likely to participate in research if it is seen of value to a connective, empathic self (Smith, 2008). In addition, pre-existing social influences of gender may impact decision making processes for women differently compared to men, which in turn

may influence their willingness to participate in research as compared to men (Lobato et al., 2014). Thus, women may be more likely to participate in research for a variety of reasons linked to social constructs such as societal expectation, or latent social constructs such as connectiveness that are more prevalent in women.

Personality characteristics

While other factors have been described as potential contributors to survey participation, little systematic research is available to support such assertions. For example, personality traits may affect study participants' decisions to engage in research, yet they are rarely measured. One study has analyzed personality traits from the Five Factor Model (FFM) and how they related to missing data from a large RCT (Jerant et al., 2009). These five factors include Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The researchers found that missing data was significantly less likely to occur among participants displaying higher levels of Openness, Conscientiousness, and Agreeableness. This indicates that those displaying greater levels of Openness (a trait associated with being intellectually curious), Conscientiousness (being planning and achievement-driven), and Agreeableness (cooperativeness and compassion) were more likely than those with lower scores in these three domains to follow up in the study and had fewer missing data associated with their participation (Jerant et al., 2009). Although this is seemingly one of the only studies to explicitly look at personality factors as they relate to participation in research, these findings align with what is known about such personality traits. Higher levels of Conscientiousness are associated with dutifulness and treatment adherence broadly (see Jerant et al. for a list of such studies). Such factors are likely to reduce RCT attrition (Jerant et al., 2009).

Beyond these studies, there is little research regarding the effects of personality traits on participation in survey research. However, personality, amongst other factors, remains one of many potential and unexplored contributors that could affect representativeness of survey research samples. This information would provide insight into potential differences between individuals who decide to partake in research and those who do not, and without this knowledge, survey research runs the risk of making broad conclusions about populations based on data from only a portion of the target population.

Adherence to research protocol

In longitudinal or RCT research, once people give consent and enroll in the research, there may be additional factors that affect continued participation over multiple time points, which in turn can limit sample representativeness. Although researchers have been encouraged to report specific study retention strategies and identify factors that influence study retention, little information exists about associations of demographic, clinical and psychosocial factors on study attrition (Bambs et al., 2013; Robinson et al., 2007). Amongst the limited body of literature, the sociodemographic factors that affect consent to participate in research may parallel those that predict protocol adherence and longitudinal survey research.

In addition, certain personality traits may influence adherence to research protocol procedures. For example, impulsivity is an important risk factor for the development of certain disorders such as substance use disorders (Moeller & Dougherty, 2002). Specifically, within the impulsivity domain is lack of perseverance, which refers to the inability to maintain focus on tasks. In longitudinal research which requires following through with multiple activities, one could posit that lack of perseverance would impact research participation, although at present there is insufficient research to support this. If impulsivity scores are associated with higher rates

of study dropouts, this could impact the remaining sample and reduce representativeness and generalizability of study findings.

Genetics research

Factors associated with participation in genetics research

Genetics research refers to the field of study of the role of genes in traits such as the development of disease. Genetics research is an increasingly important field, as precision medicine and the tailoring a person's health treatment utilizing their own genetic data, is becoming more prominent. Genetics research can encompass a variety of domains, ranging from genome wide association studies (GWAS) where sets of DNA are mapped to see if any particular genetic variant is associated with a trait of interest, to genetic testing in specific populations for particular biomarkers of a given disease. Biospecimens, such as saliva, are often collected as part of procedures to obtain genetic information from study participants via DNA.

Genetics research may have its own trends that influence study participation, and it is less clear if associations found other research domains also hold true for genetics research (Cuccaro et al., 2020). These trends may be complicated by the participant receiving results of genetic tests as part of research procedures. Such information benefits the participant directly, and thus may contribute to being more likely to participate in such research. Indeed, Sanderson and colleagues (2013) found that participants were more likely to report that they would participate in genomics research if personal results were offered than if they would not receive such data. However, research does demonstrate that even without this tangible incentive of receiving results, participants may be motivated by the altruistic value of research (Carrera et al., 2018). Specifically, Dang and colleagues (2014) found that participants were willing to participate in genetic biobanking and having their biospecimen used to benefit others.

Because genetics studies often require bio-specimen sampling as part of study enrollment procedures, little is known about refusals beyond demographics such as age, gender, and race. Thus, it is difficult to gain a complete understanding of how these two groups— ones who provide bio-specimen samples for genetics research specifically and ones who do not— may differ, and if these factors are different than those related to survey research participation. However, research does demonstrate certain consistent patterns in sociodemographic characteristics across participation in genetics and other research domains.

Race: Reflecting trends in other domains of research, individuals from minority/nonwhite groups are less likely to participate in genetics studies as well, specifically genome wide association studies (GWAS) (Haga et al., 2010). In addition, Ford and colleagues (2006) found that patients from minoritized groups were significantly less likely to enroll in a study of cancer genetics than white patients. Interestingly, Mezuk and colleagues (2008) found in their analyses from the Baltimore Epidemiologic Catchment Area Follow-Up study that although there was no association between race and the consent to donate a specimen, Black participants were less likely to consent to DNA storage for future research as compared with members of other racial groups (Mezuk et al., 2008).

Less is known about who participates in studies that offer personalized genetic DNA results, and it may vary by the disease being studied. For example, Satia and colleagues (2006) found that interest levels are similar between African American and white populations when the opportunity to receive personalized genomics information on colon cancer was available. In the same research in a qualitative component of the same study, however, African American respondents expressed concerns about potential abuses of genetic information (Satia et al., 2006).

These findings illustrate the fraught relationship between genetics research and racial minority populations in the U.S. Given the history of medical mistrust and eugenic sentiment among researchers in the U.S., developments in medical technology may intersect with existing racial relations. An example of this intersection are the privacy and consent issues prevalent in Henrietta Lacks' story, in which Lacks, a Black woman, unwittingly was the source of one of the most important cell lines used in medical research, yet she did not provide informed consent and was not compensated for their extraction or use. Stories such as this may have bolstered feelings of mistrust in relation to genetics research, in addition to fostering other important complex barriers to minority participation in genetics research (Jones et al., 2017). Similar to the impact of the Tuskegee Syphilis Experiment, stories' such as Lacks provide examples of abuses of power by primarily white researchers, and the negative relationship with research entities that has resulted.

Gender: Research on gender as it relates to genetics study participation is mixed. A study of participants of the Mayo Clinic Biobank found that identifying as female was positively associated with biobank participation compared to non-participation (Ridgeway et al., 2013). However, in larger household surveys, such as the National Health and Nutrition Examination Survey (NHANES), women were less likely to consent during a time period when they were told the specimen would be used for later genetic research (McQuillan et al., 2006). Similarly, a population-based study in Japan found that being female was a factor for nonparticipation in a population-based cohort study involving genetic research (Matsui et al., 2005).

The findings are also mixed for research that provides personalized genetics results to study participants. A U.K. population-based survey examining ascertaining interest in genetic testing for susceptibility to heart disease and cancer found that men were more likely to express

interest than women (Sanderson et al., 2004). In contrast, Alford and colleagues (2011) found no difference by gender in interest and participation in a genetic test. Such limited and mixed evidence across multiple domains may indicate that the type of genetics study, study methods, and other factors could underly the relationship between gender and genetics research participation.

Family history: Family history of health conditions may be a motivator to participate in genetics research. For instance, Ford and colleagues (2006) examined participation rates in cancer genetics research and found that family history of cancer was a predictor of study enrollment (Ford, 2006). Qualitative research has affirmed that participants in genetics research may be motivated by a family history of a condition, with beliefs that such research participation may improve healthcare for the specific population and may benefit current or future generations of the participant's family (Hallowell et al., 2010). This pattern has been reported across a range of studies and does not seem to be impacted by whether or not a participant receives individualized results from the genetic testing as part of study participation. However, much of this research about familial motivations to participate in genetics research has been done in the context of cancer genetics, which may be different than the motivations to participate from other conditions such as substance use disorders.

Participation in substance use genetics research

Psychiatric diseases carry unique factors that may further affect perceptions of genetics research relating to such conditions. However, there are a few factors that may be related to participation in psychiatric genetics research: one study found that intentions to test for psychiatric disease as part of psychiatric genetics research was positively associated with having

children, trusting the investigators, and optimism that awareness of their genetic status will better prepare them to address conditions (Laegsgaard et al., 2009).

Although there is a limited body of literature about representativeness within psychiatric genetics research, even less is known about what factors may be associated with participation in substance use genetics research specifically. This is particularly important to explicate as genetic factors have been implicated in SUD etiology (Nestler, 2001). Thus, the dearth of research on representativeness in substance use and genetics research has large implications, particularly as there is a genetic propensity for substance use disorders such as AUD, and evidence demonstrates that minority adults are disproportionately affected by negative alcohol-related consequences (Chartier et al., 2014; Zemore et al., 2018).

In addition, stigma may play a particularly important role in the study of participation in substance use genetics research. Scott and colleagues (2020) conducted a study that surveyed individuals of African descent about their willingness to participate in genetic research on various disorders. They found that people were less willing to participate in a genetic research study of alcohol use disorder (73%) as compared to other conditions such as cancer (87%), diabetes (89%), and Alzheimer's disease (88%) (Scott et al., 2020). The researchers posited that the lower percentage of willingness to participate in research related to alcohol use disorders could relate to stigma, as well as a known family history of alcohol use disorder, or complex memories and associations with alcohol (Scott et al., 2020).

Chartier and colleagues (2021) looked specifically at perceptions of genetics research as it related to alcohol use. All who gave consent reported one or more reasons for agreeing to participate. Primary reasons included thinking that genetic research is important (22.8%); participating in research can help their community (15.9%); alcohol has affected them or

someone close to them (15.2%); the topic is interesting (13.8%); or they feel an obligation to do surveys (11.7%). About one-third of the respondents who declined to participate gave reasons such as not wanting to give a DNA sample (39.5%); not being interested or not wanting to participate in a survey (27.9%); or lacking time (13.9%).

Biological measures in genetics research

DNA for genetics research has historically been conducted in person in a medical setting. However, in more recent decades, prominent US data collection efforts have adopted biotechnology, opting to collect DNA by mail by having respondents provide a saliva sample through the mail (Gatney, Couper, & Axinn, 2013). While less expensive, this approach provides another opportunity for study attrition and raises concern about representativeness. To date, little research attention has focused on the possible nonresponse bias resulting from differential compliance with biospecimen requests in large-scale surveys.

Provision of biospecimens for research protocols

Genetics studies that involve a survey and then require follow through to provide a biospecimen for DNA by mail create a longitudinal component and multiple layers of complexity. First, issues that may affect attrition in a longitudinal research protocol, such as psychiatric illness and personality traits may not only be at play, but the factors that may affect willingness to provide a sample for DNA research may have impact as well. Although participants may agree to provide a biospecimen, once they receive the activity that requires them to put genetic material in a tube and send such personal information through the mail, the decision may shift. Thus, although a person may initially consent to provision of survey responses and a sample for DNA at the outset, multiple factors may impact their follow-through in providing the biospecimen.

Dykema and colleagues (2017), through their work with the WLS, identified factors that may affect the provision of biospecimens; these dimensions include respondents' understanding of data collection procedures, methods for recruitment, informed consent processes, the type of biospecimen sought and invasiveness of such measurement, and other survey-based characteristics (e.g., use of incentives).

Task-based characteristics are particularly important to understanding participation. These characteristics span domains such as respondents' understanding of data collection procedures, methods for recruitment, informed consent language and processes, the type of biospecimen sought, the invasiveness of the method of biospecimen measurement, the setting for measurement (e.g., respondent's home, clinic), who is responsible for taking the measurement (e.g., respondents themselves, respondents with assistance, another person), and other survey-based characteristics (e.g., use of incentives) (Dykema et al., 2017).

Thus, these factors weave into a complex relationship that affects the representativeness of those who contribute biological samples. Context is important when considering the factors that may contribute to the provision of biospecimens, but the contextual factors become particularly imperative when considering biospecimen provision as part of population-based surveys. Biospecimens collected as part of such surveys provide a unique and unprecedented insight into population-level health and disease, and the implications are far-reaching.

Rates and characteristics of return of biospecimens

A select number of studies have looked explicitly at the rates of return of samples. Finding rates of return is challenging as oftentimes these rates are noted only in the study methods of published research. Table 1 summarizes 18 studies that have assessed mail-in requests of oral samples for DNA. These studies included various forms of oral samples, such as

saliva (which was obtained using saliva itself or a mouthwash technique) and buccal (which was obtained using cheek swabs). Most of the studies identified were part of ongoing research registries or longitudinal study samples. In many of these cases, participants completed an initial survey, either in-person or via phone, as part of parent cohort or registry procedures. Randomly selected cohort members were then invited to participate in a sub-study involving DNA and were sent a collection kit. They were asked to send a saliva sample back through the mail to the researchers using a pre-paid mail-in system. Researchers then looked at rates of biospecimen return, and some of these studies looked at factors related to rates of return.

Overall, rates of response widely varied from 40-80% across those studies identified in the literature. This variability was likely due to a variety of factors including study methods, populations sampled, and the timing of return of samples, among others. For example, some of the studies employed reminder mailings or phone calls to non-responders, which may have an impact on rates of return, whereas other studies did not. For example, Margolis and colleagues (2011) sent a study participation package to randomly selected members of a registry. The package included a saliva collection kit, a small incentive, IRB-approved consent forms, and an FAQ but did not mention reminders in their study methods. This study had a rate of response of 22.4% (Margolis et al., 2011). Alternatively, Hansen and colleagues (2007) employed a similar methodology, in that they randomly selected members of an ongoing cohort and sent a saliva collection kit, IRB-approved consent forms, and an FAQ document to identified members. However, this study sent two reminders to non-responders, and achieved a response rate of 72.0% (Hansen et al., 2007).

Yet other studies provided monetary incentives to those that returned biospecimens, whereas other studies did not utilize incentives, and some studies did not report on incentive

usage at all. Incentives may not only affect rates of return, but the characteristics of those who do versus do not return, as monetary compensation may be a stronger motivator for individuals of a lower SES. However, this may be countered by generally lower rates of participation amongst those with a lower SES. Finally, some studies employed different protocols throughout the data collection process regarding reminders and incentives, thus further contributing to variation not only between studies, but within studies, as far as study methodologies.

Of the seventeen studies, 5 studies did not assess differences between responders and non-responders (Etter et al., 2005; Freeman et al., 1997; Hansen et al., 2007; Margolis et al. 2011; Rylander Rudqvist et al, 2006). Most of these studies focused primarily on the quality of the biospecimens themselves and feasibility of collecting samples, rather than characteristics of responders vs. non-responders. Amongst the remaining twelve studies that did compare those returning and not returning their DNA, some found significant differences by demographic factors such as age, minority status, level of education, level of interaction with health care systems, and a prior history of health conditions (Boyle et al., 2010; Crider et al., 2006; Dykema et al., 2017; Fix et al. 2010; Kang et al., 2011; Kozlowski et al., 2002; Ness et al., 2010; Waespe et al., 2021). Others reported no statistically significant differences in follow through between those providing and not providing biospecimens (Chartier et al., 2021; Cozier et al., 2003; Fix et al., 2010; Gatney, Couple, & Axinn, 2013; Le Marchand et al., 2001).

Even within studies that found statistically significant differences, findings were mixed in terms of factors that related to provision vs. non-provision of specimens. For example, Dykema and colleagues (2017) found that of a sample of participants of the WLS, odds of provision of salivary DNA were lower among women. In contrast, Ness and colleagues (2010) found that

among invitees of the Childhood Cancer Survivor Study (CCSS) Cohort, being female was associated with the provision of a sample.

Given the lack of assessment in some studies and the contradicting results of others, the literature provides no clear conclusions on sociodemographic differences between respondents who do and do not comply with requests for self-collected biospecimens, and little elucidation on non-demographic variables associated with providing a biological specimen.

Statement of the Problem and Hypotheses

While sample representativeness has received attention in clinical trials and survey research, it has been less of a focus in genetics studies, particularly of mental disorders including SUDs. Such groups are often stigmatized, and it is important to obtain representative samples of these populations. Without adequate representation across demographic characteristics and other important factors, the results derived from such studies may only apply to the populations sampled. However, there is a paucity of research on who contributes biospecimen samples amongst individuals with substance use disorders such as AUD. Research on the factors associated with participation and nonparticipation in survey research related to AUD is scarce and it is often difficult to even determine actual rates of participation because published results often do not report such data. The implications for this dearth of research are large, particularly as the implications for AUD as it affects populations in the context of race and other sociodemographic factors is substantial.

The purpose of the proposed study was to identify demographic, psychosocial and personality variables associated with provision of saliva DNA specimens as part of a genome wide association study (GWAS) of severe AUD. The sample consisted of 3,927 individuals recruited online through Facebook. All participants met DSM-5 criteria (lifetime) for severe

AUD and completed an online survey about their alcohol and other drug use and related variables. Upon finishing the survey, individuals were mailed a kit for collection of a saliva DNA sample and were asked to return the sample in a pre-addressed, postage-paid envelope and $10 e-gift card compensation. The proposed research is important, as only 61.4% of survey completers returned their saliva samples and there is a paucity of data on the representativeness of GWAS participants, particularly for socially stigmatizing conditions such as AUD. To our knowledge, this is one of the first studies investigating predictors of biospecimen provision within this specific population.

The specific aims of the proposed research were as follows:

Specific Aim 1: Describe the demographic, psychosocial and personality characteristics of the individuals with severe AUD who completed the on-line survey and agreed to provide a saliva DNA sample.

Specific Aim 2: Compare GWAS survey completers who did (n = 2,412) and did not (n = 1,515) subsequently provide a DNA saliva sample on a variety of demographic, psychosocial and personality variables using univariate analyses. Based on the literature, the following hypotheses will be tested:

H1: Individuals self-identifying as non-white, racial minorities will be less likely to return DNA samples than individuals self-identifying as white.

H2: Men will be less likely to return DNA samples than women.

H3: Individuals with a lower SES will be less likely to return DNA samples than individuals of higher SES.

Specific Aim 3: Taking the demographic, psychosocial and personality variables for which univariate analyses detected group differences (or trends at the $p < .05$ level), a multivariate

logistic regression was used to identify the most parsimonious model of variables that predict participant follow-through with the DNA component of the study.

Methods

Parent study

This is a secondary analysis of data from an ongoing Genome Wide Association Study (GWAS) of severe AUD (R01 AA026750). The parent study is funded by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) (co-principal investigators: Drs. Kenneth Kendler and Dace Svikis). The purpose of the primary study is to identify biomarkers of risk for severe AUD through phenotypic survey data and DNA saliva samples; trace such biomarkers to their relevant risk genes; and use such findings to provide insight into biological etiologic pathways to AUD risk. Data collection began in 2018 and is ongoing. The VCU institutional review board (IRB) approved all study procedures.

Recruitment

Data collection for the parent study takes place both in-person through a network of participating AUD/SUD treatment programs and on-line through a variety of web-based groups and social media. Participants are recruited online through advertisements on such sites as Faces and Voices of Recovery (FVR), an online recovery community; Facebook, a social media platform; and Research Match, a national recruitment registry. Brief information blurbs and a link to the online survey are placed on the websites themselves. While the parent study also recruits individuals in-person from participating AUD treatment programs, the proposed study only utilized data from participants recruited online through Facebook. Given that participants from in-person treatment facilities and ResearchMatch may represent populations incentivized to participate in research, Facebook may encapsulate a broader and more generalized population.

Screening

For online recruitment, when potential participants click on the link to access the study, they are directed to a welcome screen in REDcap, a data collection software (Harris et al., 2019). They first receive an introduction to the study and are then asked to complete an online screening survey. Participants who give informed consent to participate in screening are asked eligibility questions. Eligibility criteria include: 1) being at least 18 years of age; 2) meeting the criteria for either current or lifetime severe AUD, and 3) being able to understand English. To qualify for a diagnosis of severe AUD, the participant must meet six of eleven DSM-5 diagnostic criteria during the screening process (Diagnostic and Statistical Manual of Mental Disorders, 5th edition, American Psychiatric Association, 2013). Those who do not meet criteria reach a landing page that thanks them for their participation in screening and instructs them to close out of the site.

Consent

Individuals meeting criteria proceed to an information screen that asks whether they would complete an anonymous survey and provide a salivary DNA sample. They are offered a link to a downloadable document with Frequently Asked Questions about DNA in research. Those who agree to the survey and saliva sample are informed that they are eligible to participate in the study. To continue, they must read and agree to the survey consent form, which describes the activities conducted as part of the study, and notes that they may withdraw consent at any time. Individuals who are not willing to provide a saliva (DNA) sample see a screen informing them that they are not eligible to participate.

Survey

After meeting eligibility criteria and consenting to participate in the study, participants then complete a 15-20-minute survey in REDcap, which includes questions on demographics,

family history of alcohol problems, other substance use, personality, depressive symptoms, impulsivity, and antisocial behavior. Participants can skip questions, excluding those items that factor into a decision tree. After completing the survey, participants are asked to provide contact information for mailing DNA saliva kits and, if samples are returned, provision of $10 e-gift cards for participation. In addition, respondents providing DNA informed consent are given an opportunity to consent to future contact for future genetic research and if agreeable, to supply contact information. This information is stored separately from their survey and DNA data. They are told that if contacted in the future, they can elect not to participate in any follow-up studies.

DNA Samples

After the initial survey, participants are sent a kit for collecting saliva using a test-tube and requested to provide and send back the sample through the mail. Each mailer includes a DNA kit containing a tube labeled with a barcode ID, a postage paid cardboard return envelope with a mailing address label and business reply label. They also receive a packet of forms including a DNA FAQ document with study contact information, a DNA kit instruction sheet, and the DNA study consent form. Research staff routinely retrieve returned samples from the VCU mail system. These returned samples are logged into a database and sorted by adequate volume sample, low volume sample, or no saliva. Tubes without saliva samples are discarded; the remainder of the samples are transferred to the laboratory for analysis. When the saliva samples are returned and logged in by the research team, participants receive a $10 e-gift card, or they can request that one be mailed to them through the U.S. postal service.

Participants who did not return the kits are sent reminders using a text message or e-mail procedure. Based on participant preferences, those who do not return kits receive a text message or email approximately 3 weeks from the date the kit was sent to thank them for participating in

the study and remind them to return the kit to receive compensation. If their kit has not been returned and they have not responded by 4 weeks after their kit was mailed, the study team sends a second reminder via text message or email. If 5 weeks have passed with no sample received, participants are sent a reminder e-mail. If their kit has not been returned and they have not responded by 6 weeks after their kit was mailed, they are sent a reminder letter, which is the final attempt to reach a participant. Additional kits are sent to participants upon request.

Proposed study

The proposed study analyzed data from participants recruited online through Facebook from March 2019 to December 2021. Specifically, the sample included those who provided informed consent to both the survey and saliva sample, completed the initial study survey, and were mailed the saliva kit. This recruitment window was chosen to provide sufficient time and reminders for the return of the saliva DNA kits. Participants who returned the saliva kit (n = 2,412) were compared to participants who did not return the saliva kit (n = 1,515) on a variety of demographic and psychosocial variables using chi-square and *t*-tests. Those significant at the 0.05 level were included in a multivariate logistic regression model to identify the most parsimonious model of predictors of DNA sample return.

Measures

Demographic Variables. The survey included the following demographic variables: race, ethnicity, age, gender, education, marital status, employment status, and living situation.

***Current Alcohol Use Behavior*.** Quantity and frequency of alcohol use was assessed. For frequency, participants reported how often they drank. For quantity, a drink was defined as one twelve-ounce beer, one five-ounce glass of wine, or one shot of liquor. Participants also reported

how often they consumed 6 or more drinks on one occasion. In addition, participants reported on the age at which they first began to experience alcohol problems.

Lifetime Smoking Behavior. Lifetime cigarette smoking endorsement was categorized into one of three groups: 1 = 100 or more cigarettes, 2 = Less than 100 cigarettes, and 3 = I have never smoked cigarettes. The Fagerstom Test for Nicotine Dependence (FTND), a 6-item standardized measure of nicotine dependence, was administered (Heatherton et al., 1991). The FTND has demonstrated high reliability and validity (Pomerleau et al., 1994). Items include the number of cigarettes smoked per day, time after waking in the morning to first cigarette, which cigarette would be the most difficult to give up, whether it is difficult to refrain from smoking, urge to smoke while sick, and times of smoking throughout the day. Sum totals using the Fagerstrom were examined.

Alcohol Use Disorder (AUD) Criteria. Participants were assessed for 11 DSM-5 criteria for AUD using questions converted from the actual criteria, and endorsement of at least 6 criteria was required for study enrollment. Participants were instructed to think about the 12-month period when their drinking was at its worst. Specific symptoms reported and number of criteria endorsed were analyzed.

Family History of Alcohol Problems. Problem alcohol use was assessed for first-degree relatives (parents, siblings, children) and second-degree relatives (grandparents).

Lifetime Substance Use. Lifetime use of marijuana, cocaine, stimulants, sedatives, and hallucinogens was categorized into one of three groups: never, 1-5 times, and 6 or more times (Spitzer et al., 1992).

Personality (Conscientiousness, Extraversion, Agreeableness, and Neuroticism). Personality was assessed using the 15-item Big Five Inventory Short Form (BFI-2-XS) (John &

Srivastava, 1999). The BFI-2-XS allows for the brief assessment of each of the Big Five domains. At the domain level, the BFI-2-XS retains much of the full measure's validity (Soto & John, 2017). Individual items from four of the five domains within the BFI-2-XS were examined. Openness scale items were not included in the survey for the primary study due to its poor predictive validity for substance use disorders (Kotov et al., 2010).

Impulsivity. This study utilized the 15 item Urgency, Premeditation (lack of), Perseverance (lack of), Sensation seeking, Positive urgency (UPPS-P) impulsive behavior scale (Cyders et al., 2014). The UPPS-P includes three items for each of the 5 dimensions of impulsivity. Individual items from the UPPS-P were examined.

Data Analysis Plan

Statistical analyses were performed with SPSS version 28 (SPSS Inc., Chicago, IL, USA). Descriptive statistics were run for demographic data including gender, age, race, and ethnicity. Hypotheses were tested using chi-square tests and *t*-tests for race and SES, with employment and education level used as proxies for SES. Providers and non-providers of biospecimens were compared on these variables to determine if trends in research participation hold true in a population with severe AUD. For race, it was hypothesized that non-white racial minorities would be less likely to provide DNA compared to white individuals. For gender, it was hypothesized that men would be less likely to provide DNA than women. Finally, for SES, it was hypothesized that individuals of lower SES would be less likely to participate compared to those of a higher SES.

Univariate analyses were used to identify other correlates of DNA provision, including substance use, family history, and personality. The other correlates of DNA return were identified with chi square tests for categorical variables and *t*-tests for continuous variables from

domains of interest including drug and alcohol use, family history of substance use, and personality. Sum scores were calculated for each of the following domains: conscientiousness (Big 5), lack of premeditation (UPPS-SF), lack of perseverance (UPPS-SF), and negative urgency (UPPS-SF).

Variables that reached a significance level of 0.05 in the univariate analysis were included in a final multivariate logistic regression model. If multiple items from the same domain reached significance, only one predictor was included in the multivariate analysis to avoid issues of multicollinearity. Correlations were run prior to entering variables into the regression to identify variables to represent the overall category. Only personality domains that were significant across each of the variables were included in the multivariate analysis for consistency. Significance was set at 0.05 for the multivariate analysis. The outcome of interest was the dichotomous variable of whether participants provided or did not provide biospecimens. The data was treated as is, with missing values excluded from the analyses.

Results

Demographics

Demographic characteristics of DNA providers and non-providers are summarized in Table 2, with significant variables shown in bold. The first hypothesis that non-white racial minorities would be less likely to provide DNA compared to white individuals was not supported. The two groups did not significantly differ in race; approximately 90% of both DNA providers and DNA non-providers were white (87.9% and 88.9%, respectively), $\chi^2(6) = 5.248$, $p = 0.512$. While the two groups did differ on gender, the difference was not in the hypothesized direction. Instead, men were more likely to provide DNA, as 37.7% of the providers were male, compared to 32.9% of the non-providers, $\chi^2(1) = 10.536$, $p = 0.005$.

The third hypothesis, that individuals of lower SES would be less likely to participate compared to those of higher SES, was supported, with DNA providers twice as likely to report a graduate level education (31.6%), compared to 15.9% of DNA non-providers, $\chi^2(8) = 35.940$, $p < .001$. In addition, DNA providers and non-providers differed significantly in employment status, with unemployment reported by 13.1% of DNA non-providers, as compared to 10.4% for DNA providers, $\chi^2(7) = 15.604$, $p = 0.029$.

DNA providers and non-providers differed on their living situations, with DNA providers being more likely to live with their spouse only, and DNA non-providers being more likely to live with children only, $\chi^2(7) = 17.334$, $p = 0.015$. For other demographic measures, there were no differences between DNA providers and non-providers. Both groups were of a similar age, $\chi^2(5)= 10.156$, $p = 0.180$, and most DNA providers (35.8%) and non-providers (31.8%) were married, $\chi^2(5)= 6.652$, $p = 0.248$.

Table 2: Demographic Characteristics of DNA Providers and Non-Providers

Variable	DNA Providers (N=2,412)	DNA Non-Providers (N=1,515)	P-value
Age			.180
20-29	332 (13.8%)	204 (13.5%)	
30-39	560 (23.2%)	337 (22.2%)	
40-49	513 (21.3%)	355 (23.4%)	
50-59	504 (20.9%)	350 (23.1%)	
60-69	339 (14.1%)	202 (13.3%)	
70-79	82 (3.4%)	33 (2.2%)	
Gender (% male)	910 (37.7%)	499 (32.9%)	**.005**
Race			.512
White	2,121 (87.9%)	1,347 (88.9%)	
Black	123 (5.1%)	65 (4.3%)	
Asian	52 (2.2%)	28 (1.8%)	
Marital Status			.248
Single	563 (23.3%)	375 (24.8%)	
In a relationship	486 (20.1%)	330 (21.8%)	
Married	863 (35.8%)	482 (31.8%)	
Divorced	391 (16.2%)	256 (16.9%)	
Widowed	85 (3.5%)	57 (3.8%)	

Living Situation			**.015**
Children/spouse	572 (23.7%)	365 (24.1%)	
Spouse only	689 (28.6%)	395 (26.1%)	
Children only	172 (7.1%)	145 (9.6%)	
Other family	223 (9.2%)	138 (9.1%)	
Friends	145 (6.0%)	104 (6.9%)	
Alone	547 (22.7%)	325 (21.5%)	
Homeless	17 (.7%)	20 (1.3%)	
Employment			**.029**
Full time	967 (40.1%)	604 (39.9%)	
Part time	328 (13.6%)	180 (11.9%)	
On disability	309 (12.8%)	213 (14.1%)	
Unemployed	252 (10.4%)	198 (13.1%)	
Retired	281 (11.7%)	146 (9.6%)	
Student	108 (4.5%)	55 (3.6%)	
Education			**<.001**
Grade 12/GED	299 (12.4%)	256 (16.9%)	
Some college	561 (23.3%)	391 (25.8%)	
Bachelors	670 (27.8%)	389 (25.7%)	
Masters/PhD	763 (31.6%)	241 (15.9%)	

Recent (Past Year) Alcohol Use Behavior

Measures of quantity and frequency of past year alcohol use is summarized in Table 3. Providers and non-providers differed significantly on all three variables. DNA providers endorsed higher frequency of recent use than non-providers, with 31.0% of providers reporting drinking 4+ times and 20.7% drinking 2-3 times per week, compared to 20.9% of non-providers drinking 4+ times and 17.8% drinking 2-3 times a week, $\chi^2(4) = 11.847$, $p = 0.019$.

For typical drinks per occasion, non-providers endorsed higher quantities consumed than providers, with 10.3% of the DNA non-providers reported drinking 10+ drinks on drinking days over the past year, compared to 7.2% of the DNA providers. In addition, 35.3% of DNA providers reported drinking 1-2 drinks per drinking day, compared with 30.3% of the DNA non-providers, $\chi^2(4) = 16.059$, $p = 0.003$. Finally, more DNA non-providers reported daily drinking events of 6+ drinks (12.1%) compared to DNA providers (9.3%), $\chi^2(4) = 13.074$, $p = 0.011$.

Table 3: Recent (Past Year) Alcohol Use of DNA Providers and Non-Providers

Variable	DNA Providers	DNA Non-Providers	P-value
Frequency of alcohol consumption over the past year			**.019**
Never	296/2,154 (13.7%)	157/1,347 (11.7%)	
Monthly or less	353/2,154 (16.4%)	219/1,347 (16.3%)	
2-4 times/month	392/2,154 (18.2%)	250/1,347 (18.6%)	
2-3 times/week	445/2,154 (20.7%)	240/1,347 (17.8%)	
4+ times/week	668/2,154 (31.0%)	281/1,347 (20.9%)	
Number of drinks on typical day of drinking over the past year			**.003**
1-2 drinks	678/1,920 (35.3%)	376/1,239 (30.3%)	
3-4 drinks	581/1,920 (30.3%)	366/1,239 (29.5%)	
5-6 drinks	349/1,920 (18.2%)	243/1,239 (19.6%)	
7-9 drinks	174/1,920 (9.1%)	127/1,239 (10.3%)	
10+ drinks	138/1,920 (7.2%)	127/1,239 (10.3%)	
Frequency of drinking events of 6+ drinks over the past year			**.011**
Never	525/1,960 (26.8%)	298/1,246 (23.9%)	
Less than monthly	566/1,960 (28.9%)	323/1,246 (25.9%)	
Monthly	324/1,960 (16.5%)	214/1,246 (17.2%)	
Weekly	363/1,960 (18.5%)	260/1,246 (20.9%)	
Daily/almost daily	182/1,960 (9.3%)	151/1,246 (12.1%)	

Heaviest Alcohol Use (Lifetime)

Rates of endorsement for individual criteria that contribute to DSM-5 diagnosis of lifetime AUD in DNA providers and non-providers are summarized in Table 4. Overall, rates of item endorsement ranged from 60.8% (withdrawal) to 97.7% (drank more than planned). When criteria were summed, DNA non-providers endorsed more DSM-5 criteria for severe AUD (M = 8.99, SD = 1.75) than DNA providers (M=8.86 (1.77), ($t(3,927) = 2.208$, $p = .027$). DNA providers and non-providers differed on only one criterion, craving, with DNA non-providers being more likely to report craving a drink or finding they could not think of anything but a drink (74.8%) than DNA providers (70.6%), $\chi^2(1) = 8.104$, $p = .004$. There were no significant differences between both groups in lifetime alcohol use behavior for the remaining 10 of 11 DSM domains.

Table 4: DSM-5 Criteria for AUD (lifetime)

Variable	DNA Providers (N=2,412)	DNA Non-Providers (N=1,515)	P-value
Item 1: Drank more than planned (% yes)	2,356 (97.7%)	1,480 (97.7%)	.981
Item 2: Unsuccessful efforts to reduce/stop (% yes)	1,822 (75.5%)	1,157 (76.4%)	.953
Item 3: Spent substantial time obtaining or recovering from drinking (% yes)	2,039 (84.5%)	1,283 (84.7%)	.899
Item 4: Craving alcohol (% yes)	1,703 (70.6%)	1,133 (74.8%)	**.004**
Item 5: Interfered with responsibilities (% yes)	1,882 (78.0%)	1,214 (80.1%)	.131
Item 6: Problems with friends/family (% yes)	1,918 (79.5%)	1,236 (81.6%)	.113
Item 7: Less time in valued activities (% yes)	1,876 (77.8%)	1,195 (78.9%)	.416
Item 8: Risky behavior while intoxicated (% yes)	1,818 (75.4%)	1,126 (74.3%)	.460
Item 9: Drank despite health problems (% yes)	2,292 (95.0%)	1,444 (95.3%)	.682
Item 10: Developed tolerance (% yes)	2,173 (90.1%)	1,380 (91.1%)	.300
Item 11: Withdrawal symptoms (% yes)	1,467 (60.8%)	954 (63.0%)	.054

Cigarette Smoking and Tobacco Use

Prevalence of lifetime and current cigarette smoking are summarized in Table 5. DNA providers and non-providers did not differ in lifetime smoking, with approximately two thirds of both groups reporting smoking 100+ cigarettes. Within this sub-group, DNA non-providers were more likely to report being current smokers at the time of the survey compared to providers (31.2% and 23.7%, respectively), χ^2 (1) = 19.727, p < .001. In addition, a calculated t-test in total Fagerstrom scores showed no difference in total score, (*t*(2,434) = -.759, p = .448).

Table 5: Tobacco Use of DNA Providers and Non-Providers

Variable	DNA Providers (N=1,708)	DNA Non-Providers (N=1,146)	P-value
Lifetime smoking behavior			.073
100+ cigarettes	1,018 (69.6%)	725 (63.3%)	
<100 cigarettes	246 (14.4%)	164 (14.2%)	
Never smoker	444 (26.0%)	256 (22.3%)	
Current smoker (% yes)	404 (23.7%)	357 (31.2%)	**<.001**

Other Substance Use

Other lifetime substance use is summarized in Table 6. The two groups did not significantly differ across all six substances. Cannabis was the most prevalently used substance (approximately 80%), followed by cocaine, stimulants, and sedatives, which approximately 30% of both groups had used, and hallucinogens were the least prevalently used substance, with 19% of both groups endorsing lifetime use of hallucinogens.

Table 6: Lifetime Substance Use of Providers and Non-Providers

Variable	DNA Providers (N=2,412)	DNA Non-Providers (N=1,515)	P-value
Lifetime use: Marijuana			.082
Never	302 (12.5%)	159 (10.5%)	
1-5 times	496 (20.6%)	296 (19.5%)	
6 or more times	1,614 (66.9%)	1,060 (70.0%)	
Lifetime use: Cocaine			
Never	1,195 (49.5%)	718 (47.4%)	.422
1-5 times	490 (20.3%)	322 (21.3%)	
6 or more times	727 (30.1%)	475 (31.4%)	
Lifetime use: Stimulants			.496
Never	1,306 (54.1%)	795 (52.5%)	
1-5 times	444 (18.4%)	279 (18.4%)	
6 or more times	662 (27.4%)	441 (29.1%)	
Lifetime use: Sedatives			
Never	1,275 (52.9%)	744 (49.1%)	.072
1-5 times	465 (19.3%)	314 (20.7%)	
6 or more times	672 (27.9%)	457 (30.2%)	
Lifetime use: Hallucinogens			.795
Never	1,335 (55.3%)	825 (54.5%)	
1-5 times	614 (25.5%)	400 (26.4%)	
6 or more times	463 (19.2%)	290 (19.1%)	

Family History of Alcohol Problems

Rates of parental and grandparental alcohol problems in DNA providers and non-providers are summarized in Table 7. The two groups did not differ on any category or relative, with over half of participants reporting alcohol problems in biological father (53.7%-54.8%) and biological grandfather (51.1%-52.1%).

Table 7: Family History of Alcohol Use

Variable (Alcohol Problems In):	DNA Providers	DNA Non-Providers	P-value
Biological father (% yes)	1,284/2,389 (53.7%)	822/1,499 (54.8%)	.570
Biological mother (% yes)	686/2,388 (28.7%)	469/1,502 (31.2%)	.083
Any grandfather (% yes)	1,244/2,390 (52.1%)	767/1,500 (51.1%)	.505
Any grandmother (% yes)	541/2,391 (22.6%)	368/1,497 (24.6%)	.239

Personality and Impulsivity

Four of the Big 5 domains of personality (Conscientiousness, Extraversion, Agreeableness, and Neuroticism) are summarized in Table 8. DNA providers were more likely to endorse conscientiousness, with more DNA providers reporting being capable of doing a thorough job, being a reliable worker, and being efficient compared to DNA non-providers. However, DNA providers and non-providers were similar across all variables of extraversion, endorsing similar levels of talkativeness, quietness, and sociability.

Regarding agreeableness, DNA providers were more likely to disagree with being rude to others compared to DNA non-providers (45.7% and 42.8%), $\chi^2(4) = 14.862$, $p = .005$, yet DNA non-providers were more likely to endorse being considerate and kind compared to DNA providers (82.6% and 85.4%), $\chi^2(4) = 11.191$, $p = .024$. In addition, there were no differences between levels of rating of being helpful and unselfish, with approximately 82% of DNA providers and DNA non-providers endorsing similar levels.

Finally, DNA non-providers endorsed higher levels of neuroticism on 5 of the 6 variables. A higher proportion of DNA non-providers were more likely to endorse worrying a lot, getting nervous easily, and being moody compared to DNA providers. DNA non-providers were also more likely to disagree with the statement of handling stress well.

Table 8: Personality (Big Five) Measures of DNA Providers and Non-Providers

Variable	DNA Providers	DNA Non-Providers	P-value

Conscientiousness			
Does a thorough job (% agree/strongly agree)	2,056/2,371 (86.7%)	1,236/1,492 (82.8%)	**.012**
Reliable worker (% agree/strongly agree)	2,046/2,378 (86.0%)	1,227/1,496 (82.0%)	**.016**
Efficient (% agree/strongly agree)	1,861/2,371 (78.5%)	1,104/1,490 (74.1%)	**<.001**
Extraversion			
Quiet (% disagree/strongly disagree)	1,092/2,371 (46.1%)	659/1,488 (44.3%)	.512
Talkative (% agree/strongly agree)	1,405/2,369 (59.3%)	934/1,493 (62.6%)	.236
Outgoing and sociable (% agree/strongly agree)	1,219/2,372 (51.4%)	752/1,492 (50.4%)	.059
Agreeableness			
Rude to others (% disagree/strongly disagree)	1,086/2,376 (45.7%)	638/1,490 (42.8%)	**.005**
Helpful and unselfish (% agree/strongly agree)	1,951/2,373(82.2%)	1,229/1,488 (82.6%)	.588
Considerate and kind (% agree/strongly agree)	2,028/2,376 (85.4%)	1,289/1,487 (86.7%)	**.024**
Neuroticism			
Handles stress well (% disagree/strongly disagree)	996/2,388 (41.7%)	675/1,495 (45.2%)	**.033**
Worries a lot (% agree/strongly agree)	1,685/2,376 (70.9%)	1,149/1,492 (77.0%)	**<.001**
Gets nervous easily (% agree/strongly agree)	1,408/2,381 (59.1%)	964/1,486 (64.9%)	**.001**
Can be tense (% agree/strongly agree)	1,650/2,371 (69.6%)	1,046/1,491 (70.2%)	**.011**
Emotionally stable (% disagree/strongly disagree)	901/2,382 (37.8%)	607/1,494 (40.6%)	.060
Moody (% agree/strongly agree)	1,704/2,636 (64.6%)	1,130/1,485 (76.1%)	**.002**

Impulsivity from the UPPS-SF is summarized in Table 9. Overall, the two groups differed for 11 of the 15 total variables. This included all variables in the lack of perseverance, negative urgency, and lack of premeditation domains. However, both groups endorsed similar levels of positive urgency, except for DNA non-providers reporting higher levels of others worrying about the things they do. In addition, both groups endorsed similar levels of sensation seeking, with the exception of DNA providers endorsing welcoming exciting experiences.

Table 9: UPPS-SF (Impulsivity) Measures for Providers and Non-Providers

Variable	DNA Providers	DNA Non-Providers	P-value
Lack of perseverance			
See things through (% disagree)	301/2,376 (12.7%)	233/1,496 (15.6%)	**.004**
Bothered by unfinished tasks (% disagree)	564/2,366 (23.8%)	424/1,492 (28.3%)	**.011**
Finish what is started (% disagree)	564/2,380 (23.7%)	448/1,487 (30.1%)	**<.001**
Negative urgency			
Do things that lead to regret (% agree)	1,540/2,378 (65.8%)	1,038/1,488 (69.8%)	**<.001**
Acts without thinking when upset (% agree)	1,400/2,377 (58.9%)	945/1,492 (63.3%)	**.003**
Say things that lead to regret (% agree)	1,318/2,380 (55.4%)	880/1,492 (59.0%)	**.004**

Positive urgency			
Lose control when in good mood (% agree)	744/2,377 (31.3%)	533/1,495 (35.7%)	.097
Others worry about the things I do (% agree)	675/2,372 (28.5%)	468/1,486 (31.5%)	**.035**
Acts without thinking when excited (% agree)	1,041/2,371 (43.9%)	692/1,488 (46.5%)	.359
Lack of premeditation			
Careful and purposeful thinking (% disagree)	529/2,378 (22.2%)	394/1,491 (26.4%)	**<.001**
Thinks carefully before acting (% disagree)	691/2,370 (29.2%)	498/1,484 (33.6%)	**.030**
Stops and thinks before doing things (% disagree)	609/2,368 (25.7%)	443/1,490 (29.7%)	**.035**
Sensation seeking			
Enjoys taking risks (% agree)	1,139/2,378 (47.9%)	748/1,496 (50.0%)	.283
Welcome exciting experiences (% agree)	1,618/2,372 (68.2%)	995/1,491 (66.7%)	**.015**
Would enjoy skiing fast (% agree)	985/2,376 (41.5%)	611/1,488 (41.1%)	.409

Multivariate analysis

Table 10 includes all of the significant variables, with the variables entered into the multiple regression bolded. Chosen variables included: 2 demographic variables, 3 substance use variables, 1 personality variable, and 3 impulsivity variables. For demographics, only the variables relevant to the first hypothesis were entered into the regression. For recent alcohol use, only one variable was selected for inclusion in the regression analysis to represent the overall category. For the personality and impulsivity variables, sum scores were calculated for domains in which each variable was significant as to only include the most salient domains.

Findings from the multivariate regression are shown in Table 11. Three variables were significant. Significant variables included: being male and not being a current smoker, which were associated with higher odds of providing DNA. In addition, higher levels of lack of perseverance were associated with lower odds of providing a sample. The model explained between 2.6% (Cox & Snell R Square) and 3.5% (Nagelkerke R Square) of the variance of subjects, and correctly classified 58.9% of cases.

Table 10: Variables included in the multivariate logistic regression

Demographics	**Substance use**	**Personality**	**Impulsivity**
Gender	Recent alcohol use (frequency of alcohol consumption)	**Conscientiousness**	**Lack of perseverance**
Education	Recent alcohol use (number of drinks on typical drinking day)	Extraversion	**Negative urgency**
Race	**Recent alcohol use (frequency of binge drinking)**	Agreeableness	Positive urgency
Living situation	**DSM-5 AUD symptom (craving)**	Neuroticism	**Lack of premeditation**
Employment	**Current smoker**		
Education			

*Bold indicates variable that was entered into the regression

Table 11: Multivariate logistic regression

Variable	Reference Category	Estimate	Standard Error	P-value	Odds Ratio (95% CI)
Gender					
Female	*Male*	.290	.100	**.013**	1.297 (1.055-1.594)
Education	*High*			.284	
College	*school*	-.244	.181	.177	.783 (.550-1.116)
Graduate		-.054	.148	.715	.948 (.710-1.265)
Binge drinking	*Never*			.669	
< monthly		.147	.173	.396	1.158 (.825 -1.626)
Monthly		.080	.168	.636	1.083 (.779 -1.507)
Weekly		.099	.188	.598	1.104 (.764-1.597)
Daily		-.065	.173	.706	.937 (.667-1.315)
Smoker					
Yes	*No*	.272	.103	**.008**	1.312 (1.073-1.605)
Lifetime craving					
Yes	*No*	.107	.112	.337	1.113 (.895-1.385)
Personality and Impulsivity					
Conscientiousness		-.011	.026	.674	.989 (.940-1.040)
Lack of perseverance		-.098	.032	**.002**	.907 (.851-.965)
Lack of premeditation		.019	.029	.551	1.019 (.963-1.079)
Negative urgency		-.043	.023	**.**058	.958 (.916-1.001)

Discussion

The present study was a secondary analysis of an ongoing GWAS study for severe AUD. It compared DNA providers and non-providers on a variety of demographic, psychosocial and

personality variables. All participants (N=3,927) were recruited via Facebook. Of this sample, 2,412 (61.4%) provided salivary DNA, with the remainder (n=1,515) not providing a sample. Three hypotheses about demographic variables associated with research participation were tested. In addition, univariate analyses were used to identify additional psychosocial and personality correlates of provision of DNA. Finally, variables significant in univariate analyses at the $p < .05$ level were examined, using multivariate logistic regression to determine the most parsimonious model to predict provision of DNA.

Summary of Findings

Groups compared on demographic, psychosocial, and personality variables differed on multiple domains. There were three specific hypotheses tested within demographics, and only one was supported. Other differences were found within the substance use and personality domains, including smoking status, lifetime alcohol use, conscientiousness, and impulsivity.

Discussion of Findings

Hypothesis 1: The first hypothesis, that non-white minorities would be less likely to return DNA samples, was not supported. Racial minorities were not less likely to provide DNA for this research study compared to white individuals. However, racial minorities did appear to consent to survey completion at a lower rate given that the overall sample of this study was around 90% white and 5% Black. This may not reflect the true prevalence of AUD by race, as data from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) indicates that among the first wave of data collection, 34.0% of the sample with lifetime AUD were white, and 20.6% of the sample with lifetime AUD were black (Hasin & Grant, 2015). This pattern demonstrated in the GAP sample is consistent with past research demonstrating that minority populations are less likely to participate in research than white populations (George et

al., 2014). A review by Wendler et al. (2005) found that low rates of minority participation in research were less a function of willingness to do so, and more a result of structural barriers to accessing and participating in research opportunities.

However, among those consenting to the research, there was an equal likelihood between white and nonwhite racial minorities of providing DNA. Thus, present study findings provide evidence that among those consenting, there was equal willingness to follow through and provide saliva samples. Further research is needed to better understand factors influencing study participation with continued attention on best methods for recruitment of minoritized populations.

Hypothesis Two: The second hypothesis, which posited that men would be less likely to return DNA samples than women, was also not supported in the analyses. In fact, results went in the opposite direction with men more likely to provide salivary DNA than women. The hypothesis was formulated because women are generally more likely than men to participate in research across a broad range of domains, and it has been suggested that societal expectations of gender may play a role in these rates (Lobato et al., 2014). However, there is mixed evidence in the literature regarding participation in genetics research by men and women. For example, men are more likely to participate specifically in cancer research than women. Ford and colleagues (2006) have posited that this may be due in part to men being at greater risk for developing cancer, thus making it more salient. Similarly, men are more likely to have alcohol problems than women (Centers for Disease Control and Prevention, 2022). Perhaps the relevancy of the study condition to men contributes to present study fundings. Future research should continue to explore whether such factors increase motivation to provide salivary DNA.

Hypothesis Three: The present study supported the third hypothesis, which was that individuals with a lower SES would be less likely to return DNA samples than individuals of a higher SES as measured by employment status and education level. In the present study, DNA providers were more likely to be employed and twice as likely to have a graduate degree than non-providers. This finding aligns with previous research, as education may affect health literacy and a broader understanding of clinical research. This in turn may affect what an individual participant understands about the purpose of the research and study procedures (Asare et al., 2017; Scanlon et al., 2021). This is amplified in genetics studies, where individuals with lower levels of education are less likely to participate and have less favorable attitudes toward genetic research than those with higher education (Scherr et al, 2019). For the present study, greater understanding of the research and health literacy may contribute to more familiarity with salivary DNA specimen collection and long and short-term implications of study participation.

Substance use: Recent (past year) alcohol use quantity and frequency measures yielded interesting patterns, with DNA providers being more likely to report higher frequency of use but lower quantities consumed per occasion, and lower numbers of binge drinking events. It is possible that drinking events and after-effects could influence memory or impede completion of tasks (e.g., follow through with DNA sampling), in the same way such events can contribute to missing events or neglecting social responsibilities. Further research should explore this relationship between heavy and binge drinking and the provision of DNA.

DNA non-providers were also more likely to be current smokers, which is consistent with findings by Kozlowski and colleagues (2002) who reported that smokers were less likely to return DNA samples through the mail after agreeing to participate in a larger study than non-smokers. Long-term smoking status does cause saliva production to decrease and increases the

chances of oral disorders relating to dry mouth (Rad et al., 2010). Thus, difficulty of production of saliva may have impacted willingness to to collect, pack, and send saliva through the mail.

In contrast, there were no differences in rates of problem drinking by family members for DNA providers and non-providers. At first this might seem inconsistent with broader literature indicating family history may be a motivator to participate in genetics studies, perhaps due to the personal nature of such conditions (Ford et al., 2006; Hallowell et al., 2010). The present study, however, only recruited individuals with severe AUD and overall rates of alcohol problems in family members were higher than those in the general population. Further, the present study did not focus on initial interest and engagement in research; rather it looked at post-consent follow through with provision of salivary DNA.

Personality factors: The present study found that DNA providers rated higher on conscientiousness and lower on neuroticism than DNA non-providers. These findings are consistent with previous research describing associations between conscientiousness and participation in research as well as follow-through with research protocols (Jerant et al., 2009; O'Cleirigh et al. 2007). Present study findings warrant further attention, as conscientiousness and neuroticism are associated with health and disease processes, including longevity, morbidity-related risk factors, and other health related behaviors (Bogg & Roberts, 2013; Lahey, 2009). Findings should be interpreted with caution, however, as the magnitude of score differences between DNA provider and non-provider groups, while statistically significant, were modest. Future research is warranted, with a focus on personality measures and participation in other types of public health focused genetic research.

Multivariate analyses: While univariate analyses found statistically significant differences for DNA providers and non-providers across a variety of domains, the final

multivariate model accounted for only 2% of the variance in DNA provision and correctly classified 58.9% of cases. The most parsimonious model predicting provision of DNA included nine variables, but only three were significant: being a male, non-smoker, and perseverance.

The present study is among the first to examine characteristics associated with DNA provision in a sample recruited through Facebook and focused solely on individuals with severe AUD. The overall rate of return of saliva DNA through the mail was lower than that found with other online recruitment sources in this study (e.g., ResearchMatch). Strategies to increase this rate are needed and would contribute to sample representativeness and generalizability. At present rates of DNA sample returns, current study findings are important as they shed light on sample representativeness in a GWAS study of severe AUD. While statistically significant differences were found for the two groups, the magnitude of DNA provider and non-provider differences were in many cases quite modest and should be interpreted with caution.

Study Strengths and Limitations

The present study had several strengths. First, the large sample size was sufficient to detect group differences. Second, recruitment through Facebook provided a diverse pool of participants, one not often accessed in genetic studies of AUD. More often participants are recruited through SUD treatment programs, where participants present with greater severity and increased rates of co-morbid mental health problems. Finally, data for the primary study were collected using standardized measures with demonstrated reliability and validity.

Despite these strengths, there are limitations. First, the survey data relied on self-reported information, which may be influenced by self-report and social desirability bias and underreporting of socially stigmatized behaviors (Kelpin et al., 2019; Svikis & Reid-Quinones, 2003). Second, the univariate analyses required many comparisons across groups, which

increased the probability of a committing a Type I error. However, the exploratory nature of the study necessitated many comparisons to fully explore what domains may be of interest. Third, although recruiting via Facebook allowed a wider and more diverse geographical spread, such advertisements are targeted to people who are active on social media which may affect generalizability of findings to populations that are not as active.

Study Implications and Applications

The present study provides benchmark data on representativeness in a GWAS study of severe AUD. The secondary analysis of the primary dataset offered a unique opportunity to compare providers and non-providers of DNA across a wider arrange of variables than has been reported. Overall, there were differences between those who provided and did not provide DNA. Study findings are important, as they inform the primary study about sample representativeness and generalizability.

While racial minorities may have enrolled in the study at lower rates, it was nonetheless reassuring to find no differences in rates of follow through with DNA sample provision. However, the lack of differences across race is consistent with the finding of Chartier and colleagues (2019), who found no differences in rates of consenting to provide salivary DNA across a range of racial and ethnic groups. Regardless, lack of ethnic diversity in genetic research across a wide range of conditions, including AUD, remains an area of concern and has been described as a barrier to advancement in treatment and etiology (Popejoy and Fullerton 2016).

The finding that those with graduate degrees were more likely to provide salivary DNA than those without graduate degrees is important, and it is critical that we address the educational level gap. The complexity of a GWAS makes it challenging to explain genomic science in terms that the public may understand. Dykema et al. (2017) recommend studies develop materials that

respond to a range of education levels, including providing study information written at a level appropriate for individuals with low education and low health literacy. This should be considered for future studies that include a salivary DNA component.

In the review of the literature for the present study, significant differences in research procedures across a range of studies was clearly observed. Rates of DNA sample return reflect the myriad of factors that may influence an individual's decision whether to provide a specimen. Dykema (2017) offers a conceptual model that includes understanding of data collection procedures, survey-based characteristics, informed consent process, and other factors that may inform someone's choice to provide a biospecimen. Future studies may want to consider such factors in development of study procedures, including targeted follow-up strategies and for under-represented groups. For instance, DNA non-providers in this specific study were more likely to be unemployed, suggesting that while research must not be coercive, financial incentives may be more effective with this population. A better understanding of such factors could help tailor recruitment efforts and improve the representativeness of the sample.

Conclusion

In summary, the present study offered a unique opportunity to compare DNA providers and non-providers in a GWAS study of severe AUD across a large array of psychosocial and demographic variables. Overall, the two groups differed on many variables across the domains surveyed. The magnitude and significance of these differences warrant further study. With that caveat, study findings should be considered when representativeness and generalizability of primary GWAS findings are described. The present study serves as a preliminary analysis of factors involved with the provision of DNA and highlights the need to further explore representativeness of participation in future genetics research.

Appendix: Table 1: Rates of response for mail-in requests of oral samples for DNA

Auth./Yr.	Spec.	Study methodology	Rate of response	Sample characteristics	Differences found in DNA providers vs. DNA non-providers
Bauer et al. 2004	Saliva	1) Sent an invitation letter. 2) Mailed pre-incentive (2/3 received incentive), collection kit, and informed consent form. 3) Used follow-up calls.	N=300 Rate: 37.3% (N=112)	Randomly selected smokers from the Community Intervention Trial for Smoking Cessation	No significant differences were found in demographic variables (age, gender, minority status, education, income level).
Boyle et al. 2010	Saliva	1) Participants asked if they would provide a specimen after telephone interview and offered an incentive. 2) Mailed a collection kit.	N=1540 Rate: 41.3% (N=637)	Individuals from Florida counties directly affected by 2004 hurricanes	No significant differences found by gender. Those aged 60 and older were more likely provide DNA than those aged 40-59/18-39; White participants were more likely to provide DNA than non-White participants.
Chartier et al. 2021	Saliva	1) Participants consented to survey and biospecimens. 2) Participants were mailed collection kit, and pre-incentive. 3) Up to two reminders sent. 4) Post-incentive sent.	N=128 Rate: 60.2% (N=77)	Participants of the 2015 US National Alcohol Survey and US National Alcohol's Harm to Others Survey	No significant differences were found in demographic variables (age, education, minority status).
Crider et al. 2006	Buccal	1) Sent an invitation letter with pre-incentive. 2) Followed up with phone call and consented to survey and biospecimen. 3) Received incentive for completing survey. 3) Mailed collection kit. 4) Post-incentive sent.	N=1606 Rate: 47.6% (N=764)	Randomly selected participants of the National Birth Defects Prevention Study	Minority groups were less likely to provide DNA. Among White mothers, higher education, pregnancy intention, and having a child with a birth defect were associated with increased participation. Among Hispanic mothers, an English-language interview, higher education, and receipt of the redesigned packet and incentive were associated with increased provision.
Cozier et al. 2003	Buccal	1) Mailed an introductory letter, swab collection kit, and consent.	N=644 Rate: 36.8% (N=237)	Randomly selected participants of the Black Women's Health Study	No significant differences found in demographic variables (age and education).

Dykema et al. 2017	Saliva	1) Mailed an introductory letter, collection form, kit, and instructions. 2) Reminders made via call or postcard.	N=8,081 Rate:54.0% (N=4,365)	Participants of the Wisconsin Longitudinal Study	Odds of participation were lower among females, less education, those more socially isolated, those in poorer health and with less contact with the healthcare system, and those who are more religious.
Etter et al. 2005	Saliva	1) E-mail invitation sent. 2) Participants who consented and completed survey were mailed a collection kit.	N=392 Rate: 80.4% (N=315)	Visitors of a smoking cessation site who agreed to be contacted by e-mail.	Not assessed.
Freeman et al. 1997	Buccal	1) Mailed a collection kit and explanatory material.	N=126 Rate: 92% (N=116)	Randomly selected Participants of the Twins Early Development Study (UK)	Not assessed.
Fix et al. 2010	Saliva	1) Surveyed and asked to provide DNA. 2) Those who consented were mailed a collection kit, consent, and pre-incentive. 3) Used reminder phone calls and an additional letter.	N=400 Rate: 51.8% (N=207)	Randomly selected smokers from the International Tobacco Control (ITC) Study cohort	No significant differences in age, gender, and nicotine dependence status. Significant differences found for age and country of residence.
Gatney, Couper, & Axinn 2013	Saliva	1) Mailed a collection kit, explanatory letter, pre-incentive, instruction sheet, and consent forms.	N=150 Rate: 65% (N=97)	First cohort respondents to end a romantic relationship in the Relationship Dynamics and Social Life Study	No significant differences found in certain demographics (minority status, education level, marital status). Those who provided a biospecimen were more likely to be receiving public assistance.
Hansen et al. 2007	Saliva	1) Mailed a fact sheet, consent, questionnaire, and a collection kit. 2) Two reminders sent.	N=100 Rate: 72% (N=72)	Randomly selected participants of the Danish Nurse Cohort	Not assessed.
Kang et al. 2011	Buccal	1) Mailed letters of invitation, a questionnaire, a consent, and collection kits.	N=44,773 Rate: 29.2% (N=13,084)	Randomly selected Participants in the Korea Medical Insurance	In men, participation was positively associated with old age, a family history of disease, poor health status, and regular exercise, and negatively associated with

		2) Surveys were re-sent to non-responders after 3 months.		Corporation (KMIC) (SK)	smoking status. In women, old age was associated with participation.
Kozlowski et al. 2002	Buccal	1) Administered a survey and asked if they agreed to receive an information kit. 2) Those who agreed were mailed consent, fact sheet, collection kit, and pre-incentive. 3) One reminder postcard sent.	N=3,383 Rate: 25.7% (N=870)	Telephone respondents of the Smoker and Non-smoker Study	Older, better educated, and White participants were more likely to return DNA samples, but current smokers were less likely to do so.
Le Marchand et al. 2001	Buccal	1) Mailed a letter of invitation, fact sheet, consent, and a collection kit. 2) Up to four reminders (two mailings, two phone calls) sent.	N=355 Rate: 67.3% (N=239)	Randomly selected participants of the Multi-Ethnic Cohort in Hawaii	No significant differences found in demographic variables (age, education, and smoking status).
Margolis et al. 2011	Buccal	1) Mailed a collection kit, pre-incentive, and consent forms.	N=3,264 Rate: 22.4% (N=732)	All children from the Pediatric Eczema Elective Registry Program	Not assessed.
Ness et al. 2010	Buccal	1) Mailed a collection kit, fact sheet and instructions, and consent forms.	N=10,356 Rate: 53.5% (N=5,537) (survivors); 39.0% (N=1,230) (siblings)	All survivor and sibling members of the Childhood Cancer Survivor Study (CCSS) Cohort	Female sex, white race/ethnicity, college graduation, never-smoking status, recent contact with a healthcare provider, and having a confirmed second malignant neoplasm were associated with returning a sample.
Rylander-Rudqvist et al. 2006	Saliva	1) Received preliminary notification. 2) Mailed collection kit, invitation letter, instructions, and consent form 2) Up to two reminders sent.	N=611 Rate: 80.2% (N=490)	Geographic sample of cohort of Swedish men aged 53-87	Not assessed.
Waespe et al. 2021	Saliva	1) Mailed collection kit and consent. 2) Up to two reminders sent by mail.	N=928 Rate: 29.9% (N=463)	Randomly selected survivors from the Swiss Childhood Cancer Registry	Foreign nationality, survivors aged 30-39 years at study versus other age groups, and survivors with cancer predisposition were less likely to provide DNA samples.

References

Alford, S. H., McBride, C. M., Reid, R. J., Larson, E. B., Baxevanis, A. D., & Brody, L. C. (2011). Participation in genetic testing research varies by social group. *Public health genomics*, *14*(2), 85-93.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA.

Andersen, F., Engstad, T. A., Straume, B., Viitanen, M., Halvorsen, D. S., Hykkerud, S., & Sjøbrend, K. (2010). Recruitment methods in Alzheimer's disease research: general practice versus population-based screening by mail. *BMC medical research methodology*, *10*(1), 35.

Asare, M., Flannery, M., & Kamen, C. (2017, January). Social determinants of health: a framework for studying cancer health disparities and minority participation in research. In *Oncology nursing forum* (Vol. 44, No. 1, p. 20). NIH Public Access.

Bambs, C. E., Kip, K. E., Mulukutla, S. R., Aiyer, A. N., Johnson, C., McDowell, L. A., ... & Reis, S. E. (2013). Sociodemographic, clinical, and psychological factors associated with attrition in a prospective study of cardiovascular prevention: the Heart Strategies Concentrating on Risk Evaluation study. *Annals of epidemiology*, *23*(6), 328-333.

Beebe, T. J. (2007, March). Challenges of collecting survey-based biomarker and genetic data. In *Proceedings of the 9th Conference on Health Survey Research Methods*.

Bennett, J. C. (1993). Inclusion of women in clinical trials--policies for population subgroups. *New England Journal of Medicine*, *329*(4), 288-292.

Bogg, T., & Roberts, B. W. (2013). The case for conscientiousness: Evidence and implications for a personality trait marker of health and longevity. *Annals of Behavioral Medicine*, *45*(3), 278-288.

Boyle, J.M., Kilpatrick, D., Acinerno, R., Ruggiero, K., Resnick, H., Galea, S., & Gelernter, J. (2010) Biological specimen collection in an RDD telephone survey: 2004 Florida hurricanes gene and environment study. In: Aday LA, Cynamon M (eds) Ninth Conference on Health Survey Research Methods. National Center for Health Statistics, Hyattsville, pp 176–184.

Burlew, K., Larios, S., Suarez-Morales, L., Holmes, B., Venner, K., & Chavez, R. (2011). Increasing racial minority participation in substance abuse clinical trials: lessons learned in the National Institute on Drug Abuse's Clinical Trials Network. *Cultural Diversity and Racial minority Psychology*, *17*(4), 345.

Carrera, J. S., Brown, P., Brody, J. G., & Morello-Frosch, R. (2018). Research altruism as motivation for participation in community-centered environmental health research. *Social science & medicine*, *196*, 175-181.

Centers for Disease Control and Prevention. (2022). *Excessive alcohol use is a risk to men's health.* Fact sheet, online website.

Chang, M. C., Glei, D. A., Goldman, N., & Weinstein, M. (2007). Biosocial surveys: Committee on advances in collecting and utilizing biological indicators and genetic information in social science surveys.

Chartier, K., & Caetano, R. (2010). Ethnicity and health disparities in alcohol research. *Alcohol Research & Health*, *33*(1-2), 152.

Chartier, K. G., Scott, D. M., Wall, T. L., Covault, J., Karriker-Jaffe, K. J., Mills, B. A., ... & Arroyo, J. A. (2014). Framing ethnic variations in alcohol outcomes from biological pathways to neighborhood context. *Alcoholism: clinical and experimental research*, *38*(3), 611-618.

Chartier, K. G. (2019). The paucity of genetic studies of alcohol phenotypes in diverse populations: what are the consequences and opportunities? *Addiction*, *114*(7), 1147-1149.

Chartier, K. G., Martinez, P., Cummings, C., Riley, B. P., & Karriker-Jaffe, K. J. (2021). Recruiting for diversity: a pilot test of recruitment strategies for a national alcohol survey with mail-in genetic data collection. *Journal of Community Genetics*, 1-10.

Chen, D. T., Rosenstein, D. L., Muthappan, P., Hilsenbeck, S. G., Miller, F. G., Emanuel, E. J., & Wendler, D. (2005). Research with stored biological samples: what do research participants want? *Archives of Internal Medicine*, *165*(6), 652-655.

Clayton, E. W., Evans, B. J., Hazel, J. W., & Rothstein, M. A. (2019). The law of genetic privacy: applications, implications, and limitations. *Journal of Law and the Biosciences*, *6*(1), 1-36.

Couper, M. P. (2008). *Designing effective Web surveys*. Cambridge University Press.

Cuccaro, M. L., Manrique, C. P., Quintero, M. A., Martinez, R., & McCauley, J. L. (2020). Understanding Participation in Genetic Research Among Patients With Multiple Sclerosis: The Influences of Ethnicity, Gender, Education, and Age. *Frontiers in genetics*, *11*, 120.

Cyders, M. A., Littlefield, A. K., Coffey, S., and Karyadi, K. A. (2014). Examination of a short English version of the UPPS-P Impulsive Behavior Scale. *Addictive Behaviors*, *39*(9), 1372–1376.

Dang, J. H., Rodriguez, E. M., Luque, J. S., Erwin, D. O., Meade, C. D., & Chen, M. S. (2014). Engaging diverse populations about biospecimen donation for cancer research. *Journal of community genetics*, *5*(4), 313-327.

Dunn, E. C., Brown, R. C., Dai, Y., Rosand, J., Nugent, N. R., Amstadter, A. B., & Smoller, J. W. (2015). Genetic determinants of depression: recent findings and future directions. *Harvard review of psychiatry*, *23*(1), 1.

Dykema, J., DiLoreto, K., Croes, K. D., Garbarski, D., & Beach, J. (2017). Factors associated with participation in the collection of saliva samples by mail in a survey of older adults. *Public Opinion Quarterly*, *81*(1), 57-85.

Fischer, E. H., Dornelas, E. A., & Goethe, J. W. (2001). Characteristics of people lost to attrition in psychiatric follow-up studies. *The Journal of nervous and mental disease*, *189*(1), 49-55.

Fix, B. V., O'Connor, R., Hammond, D., King, B., McNeill, A., Thrasher, J., ... & Hyland, A. (2010). ITC "spit and butts" pilot study: The feasibility of collecting saliva and cigarette butt samples from smokers to evaluate policy. *Nicotine & tobacco research*, *12*(3), 185-190.

Ford, B. M., Evans, J. S., Stoffel, E. M., Balmaña, J., Regan, M. M., & Syngal, S. (2006). Factors associated with enrollment in cancer genetics research. *Cancer Epidemiology and Prevention Biomarkers*, *15*(7), 1355-1359.

Ford, J. G., Howerton, M. W., Lai, G. Y., Gary, T. L., Bolen, S., Gibbons, M. C., ... & Powe, N. R. (2008). Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. *Cancer*, *112*(2), 228-242.

Freedman, L. S., Simon, R., Foulkes, M. A., Friedman, L., Geller, N. L., Gordon, D. J., & Mowery, R. (1995). Inclusion of women and minorities in clinical trials and the NIH Revitalization Act of 1993—the perspective of NIH clinical trialists. *Controlled clinical trials*, *16*(5), 277-285.

Galea, S., & Tracy, M. (2007). Participation rates in epidemiologic studies. *Annals of epidemiology*, *17*(9), 643-653.

Gatney, H. H., Couper, M. P., & Axinn, W. G. (2013). New strategies for biosample collection in population-based social research. *Social science research*, *42*(5), 1402-1409.

George, S., Duran, N., & Norris, K. (2014). A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. *American journal of public health*, *104*(2), e16-e31.

Haga, S. B. (2010). Impact of limited population diversity of genome-wide association studies. *Genetics in Medicine*, *12*(2), 81-84.

Hallowell, N., Cooke, S., Crawford, G., Lucassen, A., Parker, M., & Snowdon, C. (2010). An investigation of patients' motivations for their participation in genetics-related research. *Journal of Medical Ethics*, *36*(1), 37-45.

Halbert, C. H., McDonald, J., Vadaparampil, S., Rice, L., & Jefferson, M. (2016). Conducting precision medicine research with African Americans. *PLoS One*, *11*(7), e0154850.

Hansen, T. V. O., Simonsen, M. K., Nielsen, F. C., & Hundrup, Y. A. (2007). Collection of blood, saliva, and buccal cell samples in a pilot study on the Danish nurse cohort: comparison of the response rate and quality of genomic DNA. *Cancer Epidemiology and Prevention Biomarkers*, *16*(10), 2072-2076.

Harris, K. J., Ahluwalia, J. S., Catley, D., Okuyemi, K. S., Mayo, M. S., & Resnicow, K. (2003). Successful recruitment of minorities into clinical trials: The Kick It at Swope project. *Nicotine & Tobacco Research, 5*(4), 575-584.

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., et al. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics, 95*, 103208.

Hasin, D. S., & Grant, B. F. (2015). The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) Waves 1 and 2: review and summary of findings. *Social psychiatry and psychiatric epidemiology*, *50*(11), 1609-1640.)

Hauser, R. M. (2005). Survey response in the long run: The Wisconsin Longitudinal Study. *Field Methods*, *17*(1), 3-29.

Henderson, S., Stacey, C. L., & Dohan, D. (2008). Social stigma and the dilemmas of providing care to substance users in a safety-net emergency department. *Journal of health care for the poor and underserved*, *19*(4), 1336-1349.

Humphreys, K., Maisel, N. C., Blodgett, J. C., & Finney, J. W. (2013). Representativeness of patients enrolled in influential clinical trials: a comparison of substance dependence with other medical disorders. *Journal of studies on alcohol and drugs*, *74*(6), 889-893.

Jerant, A., Chapman, B. P., Duberstein, P., & Franks, P. (2009). Is personality a key predictor of missing study data? An analysis from a randomized controlled trial. *The Annals of Family Medicine*, *7*(2), 148-156.

John, O. P., and Srivastava, S. (1999). The big-five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* (2nd ed.). New York, NY: Guilford Press.

Johnson, T. P. (Ed.). (2014). *Handbook of health survey methods* (Vol. 565). John Wiley & Sons.

Jones, B. L., Vyhlidal, C. A., Bradley-Ewing, A., Sherman, A., & Goggin, K. (2017). If we would only ask: how Henrietta Lacks continues to teach us about perceptions of research and genetic research among African Americans today. *Journal of racial and ethnic health disparities*, *4*(4), 735-745.

Kelpin, S. S., Ondersma, S. J., Weaver, M., & Svikis, D. S. (2018). Representativeness of patients enrolled in a primary care clinical trial for heavy/problem substance use. *Substance abuse*, *39*(4), 469-475.

Kelpin, S.S., Rusteikas, S.G., Karjane, N.W. and Svikis, D.S. Screening for at-risk alcohol and drug use in the perinatal period: How do young women compare to older adult women? Journal of Pediatric and Adolescent Gynecology. 2019 Jun;32(3): 325-329. PMID 30584914.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychological bulletin*, *136*(5), 768.

Laegsgaard, M. M., Kristensen, A. S., & Mors, O. (2009). Potential consumers' attitudes toward psychiatric genetic research and testing and factors influencing their intentions to test. *Genetic testing and molecular biomarkers*, *13*(1), 57-65.

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, *64*(4), 241.

Le Marchand, L., Lum-Jones, A., Saltzman, B., Visaya, V., Nomura, A. M., & Kolonel, L. N. (2001). Feasibility of collecting buccal cell DNA by mail in a cohort study. *Cancer Epidemiology and Prevention Biomarkers*, *10*(6), 701-703.

Lobato, L., Bethony, J. M., Pereira, F. B., Grahek, S. L., Diemert, D., & Gazzinelli, M. F. (2014). Impact of gender on the decision to participate in a clinical trial: a cross-sectional study. *BMC Public Health*, *14*(1), 1-9.

Malone, T., Catalano, P. J., O'Dwyer, P. J., & Giantonio, B. (2002). High rate of consent to bank biologic samples for future research: the Eastern Cooperative Oncology Group experience. *Journal of the National Cancer Institute*, *94*(10), 769-771.

Matsui, K., Kita, Y., & Ueshima, H. (2005). Informed consent, participation in, and withdrawal from a population based cohort study involving genetic analysis. *Journal of Medical Ethics*, *31*(7), 385-392.

McDonald, J. A., Weathers, B., Barg, F. K., Troxel, A. B., Shea, J. A., Bowen, D., ... & Halbert, C. H. (2012). Donation intentions for cancer genetics research among African Americans. *Genetic testing and molecular biomarkers*, *16*(4), 252-258.

McQuillan, G. M., Pan, Q., & Porter, K. S. (2006). Consent for genetic research in a general population: an update on the National Health and Nutrition Examination Survey experience. *Genetics in Medicine*, *8*(6), 354-360.

Menke, A. (2018). Precision pharmacotherapy: psychiatry's future direction in preventing, diagnosing, and treating mental disorders. *Pharmacogenomics and personalized medicine*, *11*, 211.

Meshberg-Cohen, S., Presseau, C., Thacker, L.R., Hefner, K. and Svikis, D.S. Posttraumatic stress disorder, health problems, and depression among AfricanAmerican women in residential substance use treatment. Journal of Women's Health. 2016 July:25(7):729-737. PMID:27387193

Mezuk, B., Eaton, W. W., & Zandi, P. (2008). Participant characteristics that influence consent for genetic research in a population-based survey: the Baltimore epidemiologic catchment area follow-up. *Public Health Genomics*, *11*(3), 171-178.

Moeller, F. G., & Dougherty, D. M. (2002). Impulsivity and substance abuse: what is the connection? *Addictive Disorders & Their Treatment*, *1*(1), 3-10.

National Institute on Alcohol Abuse and Alcoholism. Rethinking Drinking. Washington, DC: US Department of Health and Human Services; 2009. Report no. 09-3770.

Ness, K. K., Li, C., Mitby, P. A., Radloff, G. A., Mertens, A. C., Davies, S. M., ... & Robison, L. L. (2010). Characteristics of responders to a request for a buccal cell specimen among survivors of childhood cancer and their siblings. *Pediatric blood & cancer*, *55*(1), 165-170.

Nestler, E. J. (2001). Molecular basis of long-term plasticity underlying addiction. *Nature reviews neuroscience*, *2*(2), 119-128.

O'Cleirigh, C., Ironson, G., Weiss, A., & Costa Jr, P. T. (2007). Conscientiousness predicts disease progression (CD4 number and viral load) in people living with HIV. *Health Psychology*, *26*(4), 473.

Pomerleau, C. S., Carton, S. M., Lutzke, M. L., Flessland, K. A., & Pomerleau, O. F. (1994). Reliability of the Fagerstrom tolerance questionnaire and the Fagerstrom test for nicotine dependence. *Addictive behaviors*, *19*(1), 33-39.

Rad, M., Kakoie, S., Brojeni, F. N., & Pourdamghan, N. (2010). Effect of long-term smoking on whole-mouth salivary flow rate and oral health. *Journal of dental research, dental clinics, dental prospects*, *4*(4), 110.

Radler, B. T., & Ryff, C. D. (2010). Who participates? Accounting for longitudinal retention in the MIDUS national study of health and well-being. *Journal of aging and health*, *22*(3), 307-331.

Ridgeway, J. L., Han, L. C., Olson, J. E., Lackore, K. A., Koenig, B. A., Beebe, T. J., & Ziegenfuss, J. Y. (2013). Potential bias in the bank: what distinguishes refusers, nonresponders and participants in a clinic-based biobank?. *Public Health Genomics*, *16*(3), 118-126.

Robinson, K. A., Dennison, C. R., Wayman, D. M., Pronovost, P. J., & Needham, D. M. (2007). Systematic review identifies number of strategies important for retaining study participants. *Journal of clinical epidemiology*, *60*(8), 757-e1.

Sakshaug, J. W., Ofstedal, M. B., Guyer, H., & Beebe, T. J. (2014). The collection of biospecimens in health surveys. *Health Survey Methods*, 383-419.

Sanderson, S. C., Wardle, J., Jarvis, M. J., & Humphries, S. E. (2004). Public interest in genetic testing for susceptibility to heart disease and cancer: a population-based survey in the UK. *Preventive medicine*, *39*(3), 458-464.

Sanderson, S. C., Diefenbach, M. A., Zinberg, R., Horowitz, C. R., Smirnoff, M., Zweig, M., ... & Richardson, L. D. (2013). Willingness to participate in genomics research and desire for personal results among underrepresented minority patients: a structured interview study. *Journal of community genetics*, *4*(4), 469-482.

Satia, J. A., McRitchie, S., Kupper, L. L., & Halbert, C. H. (2006). Genetic testing for colon cancer among African Americans in North Carolina. *Preventive Medicine*, *42*(1), 51-59.

Scanlon, J. K., Wofford, L., Fair, A., & Philippi, D. (2021). Predictors of Participation in Clinical Research. Nursing research, 70(4), 289.

Scherr, C. L., Ramesh, S., Marshall-Fricker, C., & Perera, M. A. (2019). A review of African Americans' beliefs and attitudes about genomic studies: opportunities for message design. *Frontiers in genetics*, *10*, 548.

Schmidt, L. A., Ye, Y., Greenfield, T. K., & Bond, J. (2007). Ethnic disparities in clinical severity and services for alcohol problems: results from the National Alcohol Survey. *Alcoholism: Clinical and Experimental Research*, *31*(1), 48-56.

Scott, D. M., Thomas, V. G., Otado, J., Rockcliffe, F., Olopoenia, O., Johnson, D., & Callier, S. (2020). Attitudes and experiences regarding genetic research among persons of African descent. *Journal of community genetics*, *11*(1), 65-72.

Sheehan, K. B. (2001). E-mail survey response rates: A review. *Journal of computer-mediated communication*, *6*(2), JCMC621.

Smith, G. (2008). Does gender influence online survey participation?: A record-linkage analysis of university faculty online survey response behavior. *ERIC document reproduction service no. ED 501717*.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69-81.

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1992). The structured clinical interview for DSM-III-R (SCID): I: history, rationale, and description. *Archives of general psychiatry*, *49*(8), 624-629.

Stewart, C., & Pepper, M. S. (2017). Cystic fibrosis in the African diaspora. *Annals of the American Thoracic Society*, *14*(1), 1-7.

Svikis, D.S. and ReidQuinones, K. Screening and prevention of alcohol and drug use disorders in women. Obstet Gynecol Clin North America, 2003, 30(3), 447-68.

Triplett T, Blair J, Hamilton T, et al. Initial cooperators vs. converted refusers: are there response behavior differences? In: 1996 Proceedings of the Section on Survey Research Methods, American Statistical Association. Alexandria, VA: American Statistical Association, 1996:1038–41.

Voigt, L. F., Koepsell, T. D., & Daling, J. R. (2003). Characteristics of telephone survey respondents according to willingness to participate. *American journal of epidemiology*, *157*(1), 66-73.

Waespe, N., Strebel, S., Marino, D., Mattiello, V., Muet, F., Nava, T., ... & Ansari, M. (2021). Predictors for participation in DNA self-sampling of childhood cancer survivors in Switzerland. *medRxiv*.

Wendler, D., Kington, R., Madans, J., Wye, G. V., Christ-Schmidt, H., Pratt, L. A., ... & Emanuel, E. (2006). Are racial and racial minorities less willing to participate in health research? *PLoS medicine*, *3*(2), e19.

Zemore, S. E., Karriker-Jaffe, K. J., Mulia, N., Kerr, W. C., Ehlers, C. L., Cook, W. K., ... & Greenfield, T. K. (2018). The future of research on alcohol-related disparities across US racial/ethnic groups: A plan of attack. *Journal of studies on alcohol and drugs*, *79*(1), 7-21.