DNA Collection in a Social Science Study: A Pilot Study of Peer Impacts on Attitudes and Drinking Behavior

Jonathan K. Daw[1], Hedwig Lee[1], Amy Lucas[1], Craig Owen[1], and Emily McKendry-Smith[1]

[1] *Department of Sociology, University of North Carolina at Chapel Hill, NC 27599*

- Key Words: DNA collection, saliva, genetic studies, social survey
- Direct correspondence to: Guang Guo, CB#3210, University of North Carolina, Chapel Hill, NC 27599-3210. Email: guang_guo@unc.edu.

ABSTRACT.

We report the experiences and participation rates of a pilot study (involving 200 individuals) of college drinking behavior incorporating both web survey and DNA collection components. Of the eligible sample, 78% completed the web survey, and 83.2% of those submitted DNA samples, for an overall participation rate of 64.9%. No evidence of lower participation rates along gender or racial/ethnic lines was observed, although there is some suggestive evidence of different levels of participant comfort with the process by race/ethnicity. Our experiences indicate that genetic data can be successfully collected for social science studies. We provide details of experiences in the data collection process and offer recommendations to inform future endeavors in DNA collection.

Social scientists today increasingly acknowledge and investigate the potential role of biological factors in relevant social processes (Committee on Population 2001; HERNANDEZ *et al.* 2006). Studies incorporating genetic variation into the study design – whether implicitly through twin studies, or explicitly through the use of molecular genetic information – hold the potential to separate the roles of social and biological factors in human behaviors of interest to sociologists. In particular the latter design potentially allows researchers to investigate gene-environment interactions -- how the influence of environmental factors depends on one's genetic makeup or how the influence of the human genome depends on one's social environment.

In this article, we describe the design and DNA collection processes of a pilot study of social and genetic effects on health behaviors in a college campus setting. We expect this overview will prove useful for researchers seeking to implement such a design, use such data collected by others, or evaluate other's research using this type of data. The specific procedures and experiences of DNA collection in social science studies have rarely been described in detail. Although the pioneering DNA collection incorporated by the National Longitudinal Study of Adolescent Health (Harris et al. 2003) was designed by academic researchers it was implemented by professional data collection organizations. Accordingly the substantive focus of Harris et al. (2003) is to provide services to Add Health users; many details of the DNA collection efforts are not reported.

We begin in Section 1 by situating DNA collection within the historical context of previous genetic research. In Section 2 we discuss the steps leading up to DNA collection in the pilot study and in Section 3, we briefly describe the web study carried out with the DNA collection. The process of DNA collection itself is described in Section 4, followed by DNA

processing and genotyping in Section 5. Finally, in Sections 6 and 7, we discuss the experiences of participants in the DNA collection and draw some preliminary conclusions.

1. Background

1.1. Recent Calls for Integrating Social, Behavioral, and Genetic Research Efforts

As DNA collection and analysis methods have increased in quality and affordability, a number of national committees composed of experts in the biological and social sciences have called for further integration of social, behavioral, and genetic research efforts. Most notably, the National Institutes of Health (NIH)'s Office of Behavioral and Social Science Research (OBSSR) asked that the Institute of Medicine (IOM) in the National Academies of Sciences undertake a study to examine the state of the science on gene-environment interactions that affect human health, focusing on the social environment. The subsequent committee report (HERNANDEZ et al. 2006) identified approaches to doing so, recommending that all three disciplines consider the others' work in their own research, develop more rigorous interaction models, and develop new datasets with the information needed to pursue these aims.

1.2. Concerns over Collecting Genetic Data

The contemporary efforts in social science research examining biological influences must be viewed in light of the historical legacy of the eugenics movement in the United States, Nazi Germany's racial 'hygiene' campaign, and more recently, ethically questionable medical research in the U.S. Although the term 'eugenics' is primarily used pejoratively today, it was an influential scientific movement until about 40 years ago. Eugenic research motivated state laws forbidding marriage and compelling the sterilization of the mentally ill, and this research also rationalized laws prohibiting interracial marriage (Kevles 1985; Halbert *et al.* 2006). The U.S. Supreme Court upheld these laws in 1927, not overturning this decision until 1967. The Nazis

similarly cited eugenic perspectives to justify their brutal efforts to 'purify' the German race by eliminating Jews, Gypsies, homosexuals, and the mentally handicapped (Proctor 1988).

Some recent policy efforts are also questionable, such as the sickle cell screening programs established in the United States by the National Sickle Cell Anemia Control Act of 1972 . Although members of all ethnic groups may carry the genetic sickle cell trait and present the disease, these screening programs nearly exclusively targeted African Americans. Furthermore, these screening programs failed to distinguish between the sickle cell genetic trait and the disease itself, and legislation erroneously referred to the disease as "communicable," suggesting that sickle cell disease is infectious, rather than inheritable. These screening programs overestimated the prevalence of the disease and ultimately led to the stigmatization of trait carriers in the health insurance and job markets (Tapper 1999). The historical case of sickle cell screening indicates the importance of social and political factors in any genetic research, in particular the differential impact such research might exert on members of different racial, ethnic, and gender groups.

1.3. Differential Participation in DNA Collection in Medical Studies

The bulk of genetic studies have involved medical, not social, outcomes. Typically, these studies report lower participation rates for African Americans and Hispanics. Using a sample of women who had previously participated in a population-based case-control study of breast cancer in North Carolina, Moorman et al. (2004) found that African American women were less likely to enroll in the cancer genetics registry than White women. Crider et al. (2006) assessed the determinants of DNA sample completion from the Atlanta, Georgia study site of the National Birth Defects Prevention Study that used mailed buccal-cell collection kits (cheek swab) following telephone interviews. Of the interviewed members of the sample (71.9%), 47.6%

submitted buccal-cell DNA: 61.9% of non-Hispanic Whites, 34.9% of non-Hispanic African Americans, and 39.1% of Hispanics. Among non-Hispanic Whites, higher education, intention to become pregnant, and having a child with a birth defect positively correlated with completion rates. Among non-Hispanic African American and Hispanic participants, those who received the redesigned packet and a $20 incentive were more likely to participate. In addition, Hispanic mothers who were interviewed in English or were more highly educated were also more likely to submit genetic samples. Even when accounting for these other factors, racial and ethnic differences in buccal-cell completion rates remained substantial.

Although these racial/ethnic differences in participation rates all took places in medical research settings, prior to data collection we had anticipated that demographic variation in participation rates would be reproduced, and perhaps amplified, in social science research contexts. The rationale for genetic data collection is not immediately apparent to participants in a social science study. This ambiguity may put any potential research subject on guard. In particular, the likely ethnically-structured participation rates may be related to the troubling history of biological explanations in social sciences whose effects have disproportionately hurt members of racial minorities. The systematic differential participation rates cannot be ignored because of the potential threat to the basis for population inference. Additionally, a paucity of research involving minority subjects may limit our ability to understand unique genetic-environmental interactions that affect these populations. Careful evaluations of racial and ethnic participation in DNA collection should, therefore, provide crucial information on the validity of social science studies with genetic components.

1.4. Study of Peer Impacts on Attitudes and Drinking Behavior

Our full study, "Peer Impacts on Attitudes and Drinking Behavior" (PIADB), is designed to examine the role of peer influences on health-relevant behaviors such as binge drinking by replicating and extending a study of roommate effects conducted at a large Midwestern university (Duncan *et al.* 2005; Boisjoly *et al.* 2006). Like this previous study, PIADB will sample pairs of roommates from the population of randomly-assigned dorm residents and analyze the effects of roommate influence on health-relevant behaviors at a large southern public university. Since the present university is much more racially and economically diverse than the site of the original study, this replication should contribute additional knowledge of roommate influences on behavior. The key innovation incorporated into PIADB, though, involves the collection of genetic information from study participants, allowing the estimation of gene-environment interactions – how individuals' susceptibility to roommate influences may be structured genetically (or, alternatively, how the effects of these genes on health behaviors may be structured by one's social environment). This will be the first time gene-environment interactions are investigated using randomly assigned environmental influences. This approach should yield far more reliable estimates of gene-environment interactions than purely observational study designs. This article, however, focuses on the DNA sample collection.

2. Preparing for DNA Collection

2.1 Institutional Review Board

Our initial application to the Institutional Review Board (IRB) included a standard written application form incorporating basic study descriptions, explanations, and potential risks of participation; brochures describing our commercial DNA collection kits; and our study budget. The board subsequently expressed a number of concerns. We chose not to inform study participants that this study was a pilot study until after the completion of the data collection

effort; IRB members questioned the need for this "deceit". They also expressed concerns about the effects of asking participants about their risky behaviors. We were able to allay these concerns by noting that our DNA collection procedures could only be effectively evaluated using a full "dress-rehearsal" approach from the participants' perspectives and agreeing to include brochures on drinking and drug use from the university's student health department in our debriefing letter.

Furthermore, the IRB inquired why we did not inform participants of the genetic portion of the study from the very beginning, and requested additional justification for the collection of genotypic data in the pilot. Although we planned to mention the DNA component in our initial contact with students, the board requested that we more fully describe our data collection plan at the time of the initial email contact. When making these changes, we were careful in all correspondence to use terms such as "DNA" and "saliva sample" rather than "genetic(s)," as study staff felt that the latter word might be more tightly linked in participants' minds with previous, harmful research on the subject. Finally, we explained to the IRB that part of the purpose of this pilot was to ensure that the Oragene kits (a description of these kits is provided below) and collection team could effectively collect genotypic data; this clarification assuaged a number of the board's concerns.

2.2 Certificate of Confidentiality

Study participants in research studies like ours may fear that their genotypic and survey information would be used for purposes beyond the scope of the immediate research project. Our consent forms assured participants that this scenario would not occur. To avoid the possibility of compulsory disclosure as the result of civil, administrative, legal, or legislative proceedings, we applied for and received a National Institutes of Health (NIH) Certificate of

Confidentiality, which ensures maximum protection from involuntary disclosure. Due to the nature of our project, we applied to NIH's National Institute on Alcohol Abuse and Alcoholism (NIAAA) department, which required that we submit our IRB application materials and the DNA and web survey consent forms.

3. The Web Survey

In addition to DNA collection, the pilot study for PIADB included a web survey of health behaviors and attitudes administered to a random sample of dorm residents. The housing department at the university provided us with a list of 200 first- and second-year dorm residents who lived in the university dorms in their first year. Departing from the design of the full study, we did not sample roommate pairs; instead, we sampled dorm residents 18 years old or above. We initially contacted the students by letter (along with two dollars cash) alerting them they had been selected for participation in the study and that we would contact them subsequently by email. In our subsequent communications with students, all conducted via email, we linked additional prorated compensation to completion of the two stages of data collection - $5 for completion of the web survey and $15 for DNA sample donation, for a maximum $20 compensation beyond the initial $2 primer. Additional email reminders were sent periodically over the course of the study to students who had not completed the web survey.

Although web surveys are often criticized for failing to provide representative samples of the general population, contemporary U.S. college students possess high rates of both computer and web access and competence, suggesting web surveys as an ideal subpopulation-specific instrument. At this particular university all entering students are required to own laptop computers at the outset of their studies, and the university provides additional computer access to all students through the library system. Additionally, our central campus study office included a

private computer available for participant use to complete the survey if they preferred or required it. Because email is a heavily used means of communication in university settings and we allowed 3-4 weeks for completion of the web survey, we anticipated that all sampled students received our emails, save those whose email addresses were listed incorrectly in the campus directory or who employed filters that might have categorized our emails as 'spam.'

Our first email communication to study participants described the scope of the study and included a link to the web survey, along with a confidential PIN number that participants used to log in to the survey site. Online, the first page included a consent letter and a link for students to click if they agreed to participate. If students consented, they were guided through several web pages to complete the survey.

4. DNA Collection

4.1 Saliva vs. Blood as Means of Collection

Researchers wishing to collect genetic information from their study participants have two main means by which to do so: by collecting blood or saliva samples. Traditionally human genetic studies collected peripheral blood samples, but new technologies have been developed that collect human DNA from buccal cells. To compare the two methods, Bhatti et al. (2005) performed the same genotyping in both blood samples (n=554) and buccal samples (n=209), and found no differences in the quality of the two DNA collection strategies. Although the quantity of genomic DNA collected from buccal cell methods is typically much lower than that of blood samples, it is generally sufficient for all but the most demanding (e.g., full genome) genotyping projects. Buccal cell methods routinely obtain 30-40 μg (micrograms) of DNA per individual, and these methods were used in the Add Health Wave III DNA collection effort, carried out by the Research Triangle Institute (Harris et al. 2003). Since the buccal cell strategy is less

expensive, does not require medical personnel to handle the samples, and is a far less invasive means of collecting genetic information, we judge it especially suitable for large scale social science research efforts studying the effects of relatively small numbers of genetic loci.

4.2 Supplies for DNA Collection

Until recently, buccal cell DNA was collected directly from a subject's cheek using a cotton swab or mouth rinse (Meulenbelt *et al.* 1995; Freeman *et al.* 1997; Freeman *et al.* 2003). However, our study utilized a simpler collection technique, the Oragene DNA self-collection kit. The Oragene kit has been used successfully in previous studies (McCready *et al.* 2005; Ahituv *et al.* 2006) and was selected for use in the Add Health Wave IV full sample DNA collection effort slated for 2008-9. The kits contain a small collection vial containing two milliliters of a preservation liquid, come with collection instructions, and are similar in appearance to a single, large contacts lens case. Our DNA collection process using the Oragene kits is as follows. If participants in our pilot had eaten in the prior 15 minutes, they were asked to rinse with water to avoid sample contamination, whereupon they filled the collection vial with their saliva up to a marked line. This intuitive process simplifies the DNA collection effort, as researchers can very quickly instruct participants by indicating the point to which participants should fill the vial. Furthermore, the Oragene kit, once filled with saliva, requires no special storage over the short term. This allows researchers to select a site convenient to study participants, rather than a site that is chosen only because it is close to a biospecimen processing lab.

In addition to the Oragene kits, collecting participant DNA required some supplementary materials. Many research participants found it difficult to produce enough saliva to fill the collection vial to the marked line. Accordingly we kept a box of white sugar cubes on hand, which stimulates saliva production, but does not contaminate the DNA samples. In addition, we

kept a ready supply of paper towels, wet wipes, and tissues handy for cleaning, along with non-latex gloves for the study staff to handle the collection vials in order to maintain sanitary conditions and prevent contamination of the sample with the staff's DNA. Finally, we kept bottled water on hand for participants to use for rinsing if they had eaten recently. Together these supplementary materials facilitated the proper collection of the DNA samples, kept matters sanitary, and helped to preserve the comfort of study participants as much as possible.

In addition to sanitation, we found privacy to be an important factor. Some study participants apparently found it difficult or otherwise uncomfortable to supply the level of saliva required with study staff in view. Therefore study stuff remained nearby to answer any questions while participants completed their saliva donations, but busied themselves with apparently unrelated activities in order to provide maximum comfort for participants and to speed the process along. Staff also offered to leave the room to provide participants with additional privacy.

4.3 Logistics of DNA Collection

Respondents automatically received an email after completing the on-line survey. The email informed students of the location and availability hours of our office where they could contribute a DNA sample and/or collect their compensation. We elected to collect participant DNA in our central campus study office for maximum convenience and privacy. In the event that this e-mail did not prompt respondents to come and give a DNA sample, the research team sent one or more e-mail reminders for the study. The most e-mail reminders a respondent could have received was ten. The first and last emails were sent in the first and seventh weeks of the study, respectively.

The first e-mail reminder focused upon how the study would safeguard and protect a respondent's privacy. Other emails were either timed to coincide with specific events, such as fall break, or provided new information such as the projected end date of the study to provide a justifiable context for contacting respondents. Study staff varied the text of each e-mail to avoid redundancy, tie in directly to the concurrent phase of the study, and, hopefully, maximize participation rates. The length of the emails varied as well; when emails were sent in close succession to one another, the later emails were designed as quick, refresher notes. Each e-mail included the contact information for the study, the office hours for the study, and emphasized the monetary incentive that respondents would receive for their participation. These measures required a minimum of effort on the part of the researchers and helped to increase the overall participation rate for the study.

In addition to sending reminder emails about the DNA phase of the study, researchers also had email contact with respondents that focused solely on assuaging concerns that respondents had toward DNA collection. The research team drafted an official response on why the study sought to collect respondent's DNA, and the study contact person sent this email as the first response to any respondent who personally contacted her with fears or questions about this phase of the study. The main text of this email is enclosed as Appendix A.

Upon entering our office, students were handed a separate, detailed consent form for the DNA collection portion of the study which, in addition to obtaining independent consent for their DNA submission, allowed them to choose two options concerning long-term DNA storage and availability. Specifically, they could choose to have their DNA destroyed immediately after the completion of the current study, or else allow study staff to make use of their (anonymous) genetic information for similar purposes for an additional seven years. Regardless of the option

selected, the document reiterated that their genetic information would never be given to anyone outside of the study staff. After completing their DNA sample donation, students completed an orally administered survey of their study participation experiences. This survey asked for their comments, their comfort with the DNA collection process, their previous DNA collection experiences, and their amenability to a hypothetical, identical participation experience in the future.

As coded by study staff, the study participants took between 3 and 22 minutes to fill the Oragene vials to the indicated line, with a mean completion time of 9 minutes and a standard deviation of 3.5.

5. DNA Processing and Genotyping

A BioSpecimen Processing Facilitating Center processed the DNA samples. The saliva DNA collection from all 124 participants was completed successfully. The average yield of DNA is 144.5 micrograms per participant with a range of 15.6 to 684.8 micrograms and a standard deviation of 125.4. Even the minimum yield is more than sufficient for all genotyping proposed in the main study.

A genotype analysis facility then carried out the test genotyping[1]. We genotyped one SNP[2] in the dopamine D2 receptor gene (*DRD2* TaqIA, dbSNP reference: rs1125394, LocusID: 1813) and a second SNP in the Catechol O-methyltransferase (COMT) val met SNP (dbSNP reference: rs4680; LocusID: 1312). The two SNPs were typed for 128 DNA samples including 4 blind duplicated controls. Five ng DNA was used for each SNP. Both SNPs were successfully

[1] The genotyping was carried out using the Applied Biosystems TaqMan® genotyping technology.
[2] SNPs are a standard means by which to characterize variations in the genome within species. Briefly, DNA is comprised of two strings of matching base pairs of nucleotides (adenine (A), thymine (T), guamine (G), and cytosine (C)). The nucleotide in one string invariantly predicts the other, as A always pairs with T, and G with C. The vast majority of the human genome is identical between individuals. When the nucleotide found at any particular locus varies within the human species, this constitutes a SNP, and indicates the presence of multiple alleles at that location which may result in differently functioning genetic processes – variation which may in some cases provide statistical traction on outcomes of interest to social scientists.

detected and the call rate was 99.2% and 98.4% for rs1125394 and rs4680 respectively. Negative controls were tested and no signal was detected. These genotyping tests indicate that the DNA collected is of excellent quality.

6. Differential Participation in DNA Collection

6.1. Overall Participation Rates

From the initial sample of 200 provided by the University Housing Department, 191 students were deemed eligible to participate; three were excluded because they were younger than 18, four did not live on campus, and two did not have email contact information listed with the housing department. Figure 1 displays overall participation rates at each stage of the pilot. 78% (149 out of 191) of the eligible sample completed the web survey, and 54% of this group completed the survey in the first three days between the initial email and the first reminder. Of the 149 study participants who completed the online web survey, 83.2% (124 out of 149) also completed the DNA collection phase, representing an overall participation rate of 64.9% (124 out of 191). Of the remaining eligible sample members who did not complete the DNA stage of the pilot, we encountered only two explicit refusals to participate.

Figure 1 about here

6.2 Participation by Race and Gender

In Table 1 we compare the ethnic and gender makeup of the previous two entering freshman classes at the university under study with the makeup of study participants who completed the web survey and submitted DNA in our pilot. We observe some minor gender differences between the population and web survey gender distributions, but no real differences between the latter and DNA participant distributions. The racial distribution did shift slightly between each stage from target population to web survey completion to DNA submission, but

not in the expected direction – the proportion of non-Whites who completed our web survey was slightly higher than that of the target population, and the percentage of non-Whites in our DNA stage completion tally was slightly, albeit insignificantly, higher than that. We observe this pattern for both African American students and students of 'other' races/ethnicities (excluding non-Hispanic Whites).

Table 1 about here

As a further check of possible racial and gender shifts in the pilot study makeup between the web survey and DNA collection components, we compare the DNA collection rates among survey completers for each race-gender subgroup. Although the subsample sizes are small, African American students who completed the survey exhibit the highest levels of conditional DNA participation, 90.91%, followed by students of other races/ethnicities, 88.46%, and White students at 80.2% (Table 2). The marginal distribution of DNA completion given survey completion by gender reveals no noteworthy differences. Looking to the inner cells of the table, we find no substantial intra-racial completion differences between males and females, and the racial distribution of DNA completion by gender preserves the marginal patterns. In sum, these results reveal no evidence of gender discrepancies in DNA completion, and a racial pattern of DNA participation opposite to that predicted by prior medical DNA participation studies.

Table 2 about here

6.3 Participant Experiences of DNA Collection

Any study of socially sensitive topics like genetic effects on behavior should pay close attention to the comfort and experiences of study participants. To address this potential concern, we measured a number of indicators of participant comfort. First, in the post-DNA interview, study staff asked participants to rank their comfort level with the DNA collection process on a

Likert scale. In the same interview we asked respondents whether they would be willing to participate in a hypothetical future study with identical procedures and incentives. Additionally, study staff discretely observed participants during the DNA collection process and made notes concerning their behavior, demeanor, and speech, which were later converted to a dummy variable indicating whether participants expressed or demonstrated concern with the process during their time in the study office. Finally, we interpret respondents' long term storage preferences for their DNA selected on the DNA consent form as an indicator of their comfort with the genetic research.

Table 3 reports mean levels for each of these variables - participant comfort, willingness to participate in future studies, and behavior - by gender and racial/ethnic group. Strikingly, we observe very high rates of willingness to participate again in a hypothetical, identical study. In fact, only two students, one Black female and one White male, responded 'no' to this question. Study staff were more likely to determine via behavior that females displayed concern about the DNA collection process than men, by a small margin (8.9% for males versus 12.5% for females), and little racial variation was observed; Whites were the most likely to appear concerned about the process. Chi squared tests for independence of variables confirm that these characteristics are unrelated bivariately.

Table 3 about here

Males and females agreed to seven-year storage of their genetic material at nearly identical rates (63.8% for males and 63.2% for females, chi squared p-value .94), whereas large racial differences were observed for this variable. Whites were the most likely to agree to long term storage (70%), followed by Blacks (60%), and respondents of other races/ethnicities (43.5%), providing some evidence that these two traits are not statistically independent (chi

squared p-value .063). Finally, males reported slightly higher Likert scale comfort (4.5) than did females (4.3), while Blacks reported slightly lower mean comfort (4.2) than non-Black participants (4.4).

Further examination of the distributions in Table 4 of these reported levels of comfort by race reveal that Black respondents were more likely to report low levels of comfort and Whites were more likely to report high levels of comfort with the process than expected under statistical independence. No clear pattern is discernable for members of other racial/ethnic groups. These are not large effects, but they are consistent with our understanding of the history of race and biological social science in the twentieth century.

Table 4 about here

Together, Tables 3 and 4 illustrate that the largest racial discrepancies in apparent comfort with the process of DNA collection manifest in the long-term storage and self-reported comfort variables. One possible mechanism by which these patterns could result is through the race dynamics of the DNA collection itself – if the race of the study participant and study staff member present differ, this difference may prime the salience of race for the respondent and increase discomfort with the process.

We created a variable indicating a racial mismatch between the participant and the staff member collecting DNA and compared the distributions of self-reported comfort and long-term storage decisions between these two groups, reported in Table 5. The results in the top half examine the relationship between participant-staff race dynamics and self-reported comfort levels, where no clear pattern emerges. Given these results the explanation for the small observed ethnic discrepancies in comfort levels most likely originates outside our study office. However, the cross-tabulations in the bottom half of the table show that participants who were

members of a different ethnic group than the staff member that collected their saliva sample were statistically significantly less likely to agree to the longer storage term for their genetic materials. This effect is large, as the observed frequency agreeing to the longer-term storage is 16% lower than that expected under statistical independence. We did not vary the race of the staff member and race of the study participant randomly for this test. Without knowing when each participant would arrive at the office, there was no way to do so. Given this, and the small size of the sample, these results are suggestive, but not conclusive.

Table 5 about here

6.4 Specific Participant Concerns

When asked, a small minority of sample members expressed concerns about the collection of their genetic information. Table 3 indicates that 11.1% of those who entered our DNA collection office appeared to show some concern about the process or its implications. Study staff also took down specific comments participants made about the DNA collection process. Participant inquiries ranged from the source of study funding to the issue of cloning. One sample member informed study staff by email that he was unwilling to participate in the DNA collection stage of the project, writing, "I am sorry, but in these genetic-laden days I am really wary of giving a saliva sample. I don't understand how it is relevant to roommate relations" (email correspondence).

Others who did participate frequently asked for clarifications on the purpose of the genetic portion of the project. One White female respondent asked a research assistant while reading the DNA consent form, "You are not going to be able to clone me or anything right?" When asked for his comments after submitting a saliva sample, one Black male participant responded, "I don't understand why you are taking my saliva. I don't get the roommate,

genetics, saliva thing, but it is way over my head and I don't think you could ever explain it to me."

Upon encountering the first of these participant inquiries, research staff developed standard oral and written explanations for the genetic portion of the project. The oral explanation focused on the theory behind the genetic component, explaining that we were researching whether DNA or social contacts better predicted their drinking behaviors. We found this to be a parsimonious explanation that most sample members readily understood and accepted. The written response to such inquiries explained at greater length that because genes had been linked to drinking behavior, we were investigating whether roommate influences on behavior were conditional on genetic makeup, then proceeded to emphasize the confidentiality and privacy measures included in the research design.

A number of participants also worried about giving out their genetic information. In these cases study staff referred them to the privacy assurances contained in the consent form and our Certificate of Confidentiality, emphasizing that the data would be maintained only in confidential form and used only for the purposes listed in the consent form. The majority of respondents appeared to be reassured by these responses. No participants who entered our office voicing concerns thereafter declined to participate.

These comments, while representing a small minority of sample members, highlight the potential concerns the general public may hold regarding genetic research. Great care should be taken in DNA collection studies such as this one to anticipate and assuage sample member's apprehensions, both to minimize psychological costs to participants and to maximize participation rates.

7. Discussion and Conclusion

Our results indicate that an excellent response rate may be expected from a DNA collection effort embedded in a traditionally social science study. Our pilot has already obtained fine response rates (64.9% of those eligible for the web survey and 83.2% of those who had completed the survey). These response rates were obtained with limited resources. We set up one single DNA collection office at the central campus, which was open (manned) for four hours a day, four days a week for about four weeks. No telephone calls were made. We have used our experiences in this data collection effort to prepare for the PIADB study. We plan a number of strategies to increase participation and allay sample members' potential concerns with the genetics portion of the study. First, we intend to more completely explain the theoretical motivations for collecting genetic information at the onset of the study. Most sample members who asked about this seemed to not understand our reasons for doing so and accepted our explanations when given. We anticipate that this step will help to streamline the collection effort, as some students who felt unsure of our motivations likely elected not to participate in the project.

Second, in the interests of participant convenience we plan to setup discrete DNA collection sites closer to university dormitories and open for DNA collection during the evening. Although our study office for this pilot was located in the middle of campus, close to many classroom and student services buildings, most dormitories at this university lay some distance from this area. We believe that this step will assist those who work or have busy course loads during the day in completing this crucial stage of the project, and encourage those unsure whether they wish to participate to do so.

In this vein, we also intend to make telephone calls to sample members to encourage their participation and assuage any concerns they may have about the project. We received direct

responses from only a tiny number of nonparticipants in the pilot, but presumably some of those who did not participate or respond directly harbored similar concerns. Direct phone contacts with these persons could help to clarify our intent, help to decrease psychological costs of participation, and increase participation rates.

In summary, we consider our DNA and survey data collection efforts a success. Gender differences in participation and comfort with DNA collection were minor or nonexistent; and race differences in these measures appear small but potentially important. The most relevant statistic to keep in mind, however, is 83.22% - the percentage of survey participants who subsequently donated DNA samples. Given the unfamiliarity of much of the public with this type of research, we find this number to be a very encouraging one.

Finally, we emphasize that genetic studies of behavior constitute an important and underutilized frontier in social science research. Our experiences in this pilot study demonstrate that high response rates and high quality genetic samples can be collected and analyzed by traditional social science research teams in tandem with widely available genetics laboratories.

8. Coda: Recommendations

We close this article with a summary discussion of recommendations to future researchers incorporating similar DNA collection efforts into a traditional social science research study:

Recognize and address respondent concerns in donating their DNA. Few data are more psychologically sensitive to would-be participants than their genetic makeup, and many respondents will not immediately recognize the relevance of these data to social scientific inquiries. By discussing the reasons you request their genetic samples from the beginning, by taking all possible steps to preserve participants' anonymity and confidentiality, and by taking

time to reduce participants' legitimate apprehensions, you will increase their comfort with the process, minimize the costs of participation, and build trust with your study sample members. In the interests of consistency and efficacy, we recommend you develop standard written and oral responses to typical concerns and consider incorporating them into your correspondence with sample members.

Provide monetary compensation. Recall the Crider et al. (2006) finding that the promise of $20 compensation increased the response rate of mail-in buccal cell collection kit participation among African American and Hispanic sample members, two groups at heightened risk for non-participation in many genetic studies. Although our project did not assign incentives experimentally, we discovered some anecdotal evidence that our own $20 compensation increased participation levels. Some sample members entered our study office harboring reservations on the use of their genetic data in social science research which we then sought to assuage; presumably many such persons would not have come by the office were it not for the promise of compensation. Perhaps most importantly, compensating sample members for their genetic donations demonstrates recognition of time commitment and trust they invest in your project.

Ensure participant privacy and confidentiality. This recommendation extends beyond the typical privacy measures implemented in social research, such as anonymity and confidentiality protections. All steps possible should be taken to protect knowledge of an individual's participation from inadvertent public disclosure. Otherwise deductive disclosure could present a very real threat to participants' privacy should the datasets become publicly available. Therefore DNA collection operations should be placed in discrete locations, in areas respondents might frequent without attracting notice. Steps should be taken to minimize the

likelihood that participants, who know the purpose of the study, will recognize one another in the study setting. Furthermore we strongly recommend you take steps to secure a Certificate of Confidentiality from the relevant government agency to protect your participants from involuntary disclosure of your data. Finally, we recommend in the interests of efficiency and dignity that private saliva donation stations be provided to all participants.

Acquire supplementary materials to maximize participant comfort and preserve sanitary conditions for donation. We discussed the utility of these materials above, and simply list them here again for convenience: paper towels, wet wipes, sanitary gloves, sugar cubes, and a supply of water. We recommend handing out paper towels to each participant along with the kit, so that they do not need to ask specifically for this item.

Provide clear DNA submission instructions for participants. Even after our standard instructions, many participants asked questions, frequently inquiring on the best way to fill the vial, and whether a certain amount of saliva was "enough." In order to answer these practical sorts of questions, we strongly recommend that all research assistants for this type of project spend some time practicing the DNA submission method themselves. Having experienced the process themselves will instill in study staff a level of empathy with participants and knowledge of the collection process that will help to minimize the costs of participation for respondents and streamline the DNA collection process.

Make study staff contact information available to sample members. In our pilot a number of sample members inquired concerning the purposes of the genetic component of our project, and expressed concerns on the applications to which their data would be put to use. By providing contact information for members of your research team, you facilitate these inquiries,

giving you the opportunity to address concerns, build trust with study participants, and increase participation rates.

Establish a waiting area nearby your DNA collection location. Many DNA participants in our pilot came by the collection office with company in tow. Furthermore, certain times in the day (i.e., lunch time) tended to be much busier than others, and on occasion even with our small sample we had to request that study participants wait at a distance while others completed their saliva donations. Luckily our study office was located directly next to a lounge with a number of chairs and tables available in which these persons could wait comfortably.

Hiring a diverse study staff may prove helpful. Given the apparent relationship of respondent-study staff race/ethnicity differences with respondent approval of long-term storage of their DNA, it is possible that the demographic makeup of your research team could beneficially influence respondent comfort with the process and increase participation rates. Although our evidence is insufficient to establish a causal claim, interviewer effects are documented in the survey methodology literature (Johannes *et al.* 1997; Andersen and Olsen 2002).

As with other recommendations we have outlined in this coda, being sensitive to the ethnic makeup of a research team serves both practical and ethical interests – you stand to maximize participation rates of all demographic groups, thereby increasing your sample size and ensuring the representativeness of your research sample, while simultaneously recognizing and addressing the potential sources of participation distress in your sample.

Figure 1: Response rates for the web survey and DNA collection

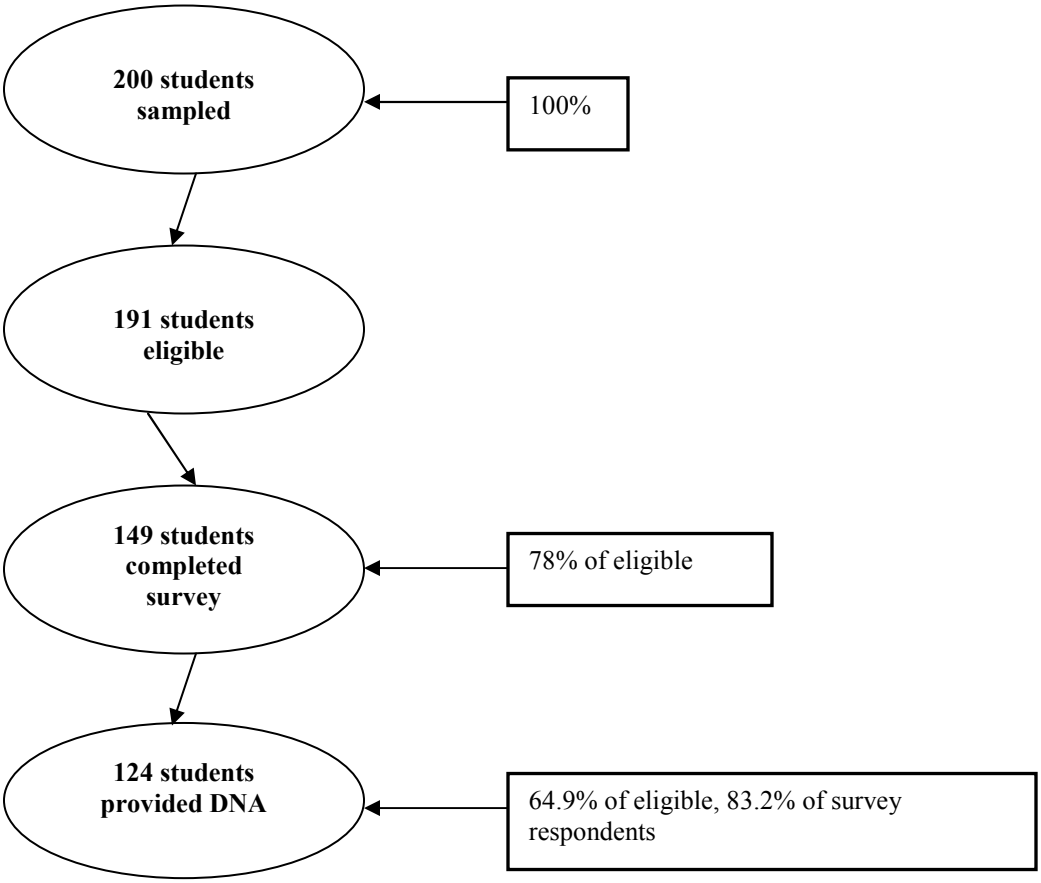

Table 1: Percent distribution of student demographic characteristics in the school population, in web survey participants, and in DNA participants.

Gender	Student Pop.*	Web Survey Participants	DNA Participants
Male	39.74%	37.58%	**37.90%**
Female	60.26%	62.42%	**62.10%**
Race/Ethnicity			
Black	11.71%	14.77%	**16.13%**
White	75.25%	67.79%	**65.32%**
Other	13.04%	17.45%	**18.45%**

*Freshman enrolling classes, 2005 and 2006, admissions department at the study university.

Table 2: Percent of web survey completers who contributed saliva samples

Race	Male	Female	Total
Black	100.00% (N=6)	87.50% (N=16)	**90.91% (N=22)**
White	78.95% (N=38)	80.95% (N=63)	**80.20% (N=101)**
Other	91.67% (N=12)	85.71% (N=14)	**88.46% (N=26)**
Total	**83.93% (N=56)**	**82.80% (N=93)**	**83.22% (N=149)**

Table 3: Mean respondent levels of willingness to contribute DNA again, staff-coded indicators of concern with the process, agreement to long-term storage of DNA, and Likert-scale comfort level with the DNA collection process, by gender and race/ethnicity

	Contribute DNA Again	Concern about Process	Agreed to long-term storage	Reported Comfort Level
Gender				
Male	97.8%	8.9%	63.8%	**4.5**
Female	98.7%	12.5%	63.2%	**4.3**
Pvalue*	---	*.545*	*.940*	***.262***
Race				
Black	95.5%	11.1%	60.0%	**4.2**
White	98.9%	11.7%	70.0%	**4.4**
Other	100%	9.9%	43.5%	**4.4**
p-value*	---	*.943*	*.063*	***.058***
Total	**98.4%**	**11.1%**	**63.4%**	**4.4**

*P-values result from chi square tests of independence. Values for "Contribute DNA Again" are omitted due to insufficient variation.

Table 4: Frequency distribution of comfort levels by race, with expected frequencies (in parentheses) under independence

Comfort	White	Black	Other	Total
2	0	2	0	**2**
	(1.3)	*(0.3)*	*(0.4)*	
3	7	2	1	**10**
	(6.5)	*(1.6)*	*(1.9)*	
4	31	7	12	**50**
	(32.5)	*(8.1)*	*(9.3)*	
5	42	9	10	**61**
	(39.7)	*(9.9)*	*(11.4)*	
Total	**80**	**20**	**23**	**123**

Table 5: Participant comfort levels by race difference status with research staff.

	Participant-Staff Race Dynamics	
Comfort (p-value = .158)	Same	**Different**
2	1	**1**
	(1.2)	**(0.8)**
3	5	**5**
	(6.0)	**(4.0)**
4	29	**21**
	(30.1)	**(19.9)**
5	39	**22**
	(36.7)	**(24.3)**
Long-Term Storage? (p-value = .052)		
No	22	**23**
	(27.1)	**(17.9)**
Yes	52	**26**
	(46.9)	**(31.1)**

*Expected distribution under independence of race and comfort level are shown below the observed frequencies.

Appendix A. An official response on why the study sought to collect respondent's DNA

The purpose of this study is to understand roommate influences on undergraduate student health behaviors. Health behaviors such as binge drinking have been linked to biological factors including genetic influences by past research. Our purpose in collecting saliva samples for this project is to help understand whether roommate influences are conditional on participants' biological factors.

Your privacy and confidentiality will be strictly guarded during this project. The process is supervised by the [Name of Institution] Institutional review Board (IRB), which was formed for the sole purpose of protecting participants in scientific studies. For example, our study will never ever identify individuals. We are only interested in group averages and comparison of group averages. The location of the study in the research report will be referred to as a "Southern public university"; the name of the University will never be mentioned. Both the survey data and biological data will be used by our team alone and will never be transferred to anyone else. Also, we will only use the survey and biological data for the very limited purposes approved by you and the university IRB.

The References

Ahituv, N., N. Kavaslar, W. Schackwitz, A. Ustaszewska, J. M. Collier, S. Hebert, H. Doelle, R. Dent, L. A. Pennacchio, and R. McPherson. 2006. "A pyy q62p variant linked to human obesity." *Human Molecular Genetics* 15:387-391.

Andersen, A. M. N. and J. Olsen. 2002. "Do interviewers' health beliefs and habits modify responses to sensitive questions? A study using data collected from pregnant women by means of computer-assisted telephone interviews." *American Journal of Epidemiology* 155:95-100.

Bhatti, P., A. J. Sigurdson, S. S. Wang, J. B. Chen, N. Rothman, P. Hartge, A. W. Bergen, and M. T. Landi. 2005. "Genetic variation and willingness to participate in epidemiologic research: Data from three studies." 14:2449-2453.

Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles. 2006. "Empathy or antipathy? The impact of diversity." *American Economic Review* 96:1890-1905.

Committee on Population. 2001. *Cells and surveys: Should biological measures be included in social science research?* Washington D.C.: National Academy Press.

Crider, K. S., J. Reefhuis, A. Woomert, and M. A. Honein. 2006. "Racial and ethnic disparity in participation in DNA collection at the atlanta site of the national birth defects prevention study." 164:805-812.

Duncan, G. J., J. Boisjoly, M. Kremer, D. M. Levy, and J. Eccles. 2005. "Peer effects in drug use and sex among college students." *Journal of Abnormal Child Psychology* 33:375-385.

Freeman, B., J. Powell, D. Ball, L. Hill, I. Craig, and R. Plomin. 1997. "DNA by mail: An inexpensive and noninvasive method for collecting DNA samples from widely dispersed populations." *Behavior Genetics* 27:251-257.

Freeman, B., N. Smith, C. Curtis, L. Huckett, J. Mill, and I. Craig. 2003. "DNA from buccal swabs recruited by mail: Evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping." *Behav Genet* 33:67-72.

Halbert, C. H., O. H. Gandy, A. Collier, and L. Shaker. 2006. "Intentions to participate in genetics research among african american smokers." 15:150-153.

Harris, K.M., F. Florey, J. Tabor, P.S. Bearman, J. Jones, and J.R. Udry. 2003. "The national longitudinal study of adolescent health: Research design." Available online at: Http://www.Cpc.Unc.Edu/projects/addhealth/design." vol. 2005.

HERNANDEZ, LYLA M., DAN G. BLAZER, and (EDS.). 2006. *Genes, behavior, and the social environment: Moving beyond the nature/nurture debate*. Washington, DC:: National Academies Press.

Johannes, C. B., S. L. Crawford, and J. B. McKinlay. 1997. "Interviewer effects in a cohort study - results from the massachusetts women's health study." *American Journal of Epidemiology* 146:429-438.

Kevles, Daniel. 1985. *In the name of eugenics: Genetics and the uses of human heredity* New York: Knopf.

McCready, M. E., A. Grimsey, T. Styer, S. M. Nikkel, and D. E. Bulman. 2005. "A century later farabee has his mutation." *Human Genetics* 117:285-287.

Meulenbelt, I., S. Droog, G. J. M. Trommelen, D. I. Boomsma, and P. E. Slagboom. 1995. "High-yield noninvasive human genomic DNA isolation method for genetic-studies in geographically dispersed families and populations." *American Journal of Human Genetics* 57:1252-1254.

Moorman, P. G., C. S. Skinner, J. P. Evans, B. Newman, J. R. Sorenson, B. Calingaert, L. Susswein, T. S. Crankshaw, C. Hoyo, and J. M. Schildkraut. 2004. "Racial differences in enrolment in a cancer genetics registry." 13:1349-1354.

Proctor, Robert. 1988. *Racial hygiene: Medicine under the nazis* Cambridge, MA: Harvard University Press.

Tapper, Melbourne 1999. *In the blood: Sickle cell anemia and the politics of race. Critical histories*. Philadelphia: University of Pennsylvania Press.