SOCIOLOGICAL METHODOLOGY
AMERICAN SOCIOLOGICAL ASSOCIATION

1

DNA COLLECTION IN A RANDOMIZED SOCIAL SCIENCE STUDY OF COLLEGE PEER EFFECTS

Guang Guo[*,†]
Jessica Halliday Hardie[*]
Craig Owen[*]
Jonathan K. Daw[*]
Yilan Fu[*]
Hedwig Lee[*]
Amy Lucas[*]
Emily McKendry-Smith[*]
Greg Duncan[‡]

This data collection effort was funded by a seed grant to Guang Guo from the Odum Institute at the University of North Carolina–Chapel Hill and a major grant to Guang Guo from the William T. Grant Foundation. Special acknowledgment is due to the BioSpecimen Processing Facility at the University of North Carolina–Chapel Hill for extracting the DNA from the saliva samples and for preparing the DNA plates for genotyping and to UNC Mammalian Genotyping Center for genotyping. We would also like to thank the Chapel Hill Sociology Department—Tianji Cai, Yi Li, Alice Weng, Allison Matthews, Bethany Remington, Brandon Wagner, Cindy Karol, Crista Cuccaro, Eric Amidon, Jacob Rosch, Jennie Smith, Kayo Suzuki, Marybelle Guo, Reine Duffy, Richard Silc, Rick Bradley, Shawn Bauldry, Stefanie Knauer, Teresa Edwards, Tim Cupery, and Tim Wander—for their contribution to the project. Direct correspondence to Guang Guo, CB#3210, University of North Carolina, Chapel Hill, NC 27599-3210; e-mail: guang_guo@unc.edu.

*University of North Carolina at Chapel Hill
†Carolina Center for Genome Sciences
‡University of California, Irvine

We describe the DNA collection processes of an initial pilot and full study, which is designed to investigate joint peer and genetic effects on health behaviors and attitudes in a college campus setting. In the main study, 2664 (79.5%) students completed a Web survey and 2080 (78.7% of the survey completers after adjusting for the ineligible) provided a saliva DNA sample. The response rate for completing both the survey and the DNA portion of the study is 62.5%. Our DNA yields are of high quality. Overall, our experiences and results demonstrate that genetic data can be successfully collected as a part of traditional social science survey research projects. To aid others in doing so, we provide extensive details of our data collection experiences and offer recommendations to future researchers seeking to do or evaluate similar work.

1. INTRODUCTION

Social scientists today increasingly acknowledge and investigate the potential role of biological factors in relevant social processes (Committee on Population 2001; Hernandez and Blazer 2006). A large number of population-based social surveys have collected biological markers other than DNA, including the Los Angeles Family and Neighborhood Survey (L.A.FANS); Midlife in the United States (MIDUS); National Social Life, Health, and Aging Project (NSHAP); and the Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan. A growing number of studies have collected or plan to collect DNA data for genetic analysis, including Add Health; the Danish 1905-Cohort Study; the Fragile Families Study; the Health and Retirement Study (HRS); the MacArthur Research Network on Successful Aging Community Study; and the Wisconsin Longitudinal Study (WLS).

Add Health pioneered DNA data collection among a subset (about 2600) of its participants at Wave III in 2002 (Harris et al. 2003). The DNA data collection was designed by Add Health investigators and implemented by professional data collection organizations. These data are currently a primary source of molecular genetic information used in social science studies. There is one major difference in the mode of DNA collection between Add Health and our effort: While Add Health locates and visits a respondent's home, due to budget constraints we can only ask respondents to visit our DNA collection offices on campus. Recently, WLS carried out a massive DNA data collection, asking more

than 8000 WLS graduates to donate a saliva sample[1] via an Oragene kit and return it by mail (Hansen et al. 2007). This report focuses on the details of DNA collection procedures, processes, and experiences in our pilot and full studies.

As DNA collection, genotyping, and related 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, in 2004, 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 and Blazer 2006) identified specific approaches, recommending that all three disciplines (social sciences, behavioral sciences, and genetics) consider the others' work in their own research, develop more rigorous gene-environment interaction models, and collect new data sets with the information needed to pursue these aims.

Advances in molecular genetics over the past few decades have made it possible to begin deciphering the linkages between genetic factors and multiple human social and health outcomes. Many recent studies have yielded discoveries concerning the genomics of complex traits (e.g., Frayling et al. 2007; Scott et al. 2007; Sladek et al. 2007; Steinthorsdottir et al. 2007; Zeggini et al. 2007). These studies identified genetic variants associated with acute lymphoblastic leukemia, obesity, type 2 diabetes mellitus, prostate cancer, breast cancer, and coronary heart disease. These recent discoveries were made via sophisticated methodological approaches, based on at least several thousand individuals, applying state-of-the-art techniques to control for potential population admixture, addressing multiple testing, and replicating at least a number of large and independent samples.

Studies incorporating genetic measurement into the study design hold the potential to separate the roles of social and biological factors in human behaviors of interest to sociologists. In particular, the

[1] Briefly, a saliva sample is collected by spitting into a container while a sample of buccal cells is collected by rolling a swab firmly inside the cheek. The saliva method has become more popular in recent years.

incorporation of molecular genetic information 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. Articles that take advantage of molecular genetic measures have started appearing in leading sociological journals (e.g., Guo, Roettger, and Cai 2008). A special issue of the *American Journal of Sociology* (Vol. 114, No. S1) devoted to social structure and genetics was published in late 2008.

The main purpose of this article is to provide a detailed description of the DNA collection processes of the initial pilot and main studies, which were designed to investigate joint social and genetic effects on health behaviors and attitudes in a college campus setting. The article will also describe an essential component of the study—a "natural experiment" of randomly assigning roommates that avoids the confounding effects of residential choice. Ours will be one of the first instances in which a gene by environment interaction is tested using a randomly assigned "environment." Our overview of the data collection process at both phases of the studies will help to inform future research in this growing area of social science. We hope this review will prove valuable for researchers who seek to collect biological markers including DNA data, to use such data collected by others, or to evaluate research using this type of data.

Section 2 begins with a brief discussion of the overall research design of the project. Section 3 situates DNA collection within the historical context of previous genetic research. Section 4 discusses the steps leading up to DNA collection in the pilot study and the main study, and Section 5 briefly describes the Web study carried out with the DNA collection. The process of DNA collection itself is described in Section 6, followed by DNA processing and genotyping in Section 7. Sections 8 and 9 discuss the experiences of participants in the DNA collection and draw some preliminary conclusions. Section 10 summarizes recommendations for social science researchers interested in collecting DNA.

2. THE OVERALL RESEARCH DESIGN

Our study, "Peer Impacts on Attitudes and Drinking Behavior" (PIADB), is designed to examine the role of peer influences on

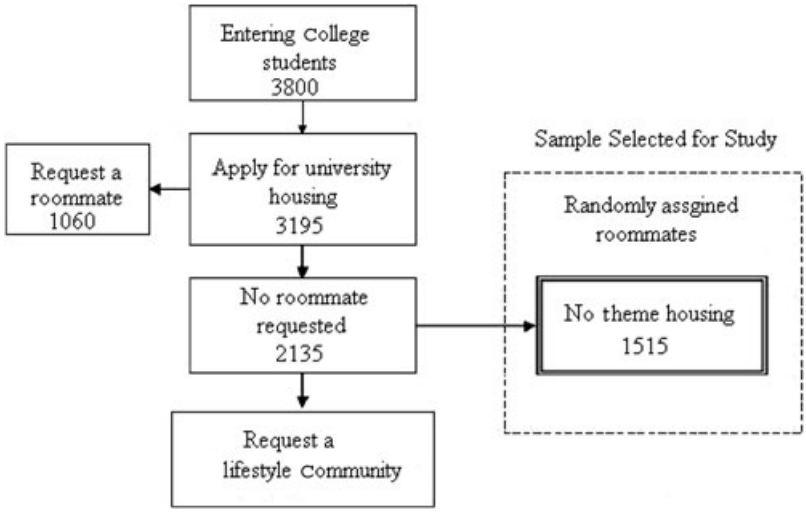

FIGURE 1. Example of university roommate assignment process. Note that the numbers included in the figure are from a single year preceding our sample and are included to represent the typical assignment process in place during the three years of roommate assignments that were used in this study.

health-relevant behaviors such as binge drinking by replicating and extending a study of roommate effects conducted at a large Midwestern university (Boisjoly et al. 2006; Duncan et al. 2005). A key feature of PIADB and its predecessor study is to sample pairs of roommates from a population of randomly assigned dormitory residents at a large public university. In most studies of peer influences, peers themselves are selected through some nonrandom process. Our study removes threats to bias arising from selection processes by studying roommates who were randomly assigned to one another.

The PIADB drew its sample from a university that is much more racially and economically diverse than the original study university. We worked with the university's housing department to verify that their roommate assignment criteria did indeed pair individuals randomly. We illustrate the sampling framework using data observed from a year that precedes the years of our study (Figure 1). In each academic year, a large majority of incoming freshmen and 10% of returning students complete an online housing application. This application asks a variety of questions about an individual's housing preferences as well as their

gender. Of the 3800 freshman students applying for university housing in the 2004–2006 academic years, 3195 applied by the May application deadline. Of these, 2135 did not request a specific roommate. Of the 2135 students, 1515 did not request to participate in a themed housing program (e.g., substance free, foreign languages, health sciences, and global business).

To randomly assign roommates to the residual pool of 1515 students, the Housing office placed data from applications into a large Excel file, which was loaded into the RMS software. Every student was then assigned randomly a unique RMS-ID number. After the first student was placed in a room, the RMS program assigned his or her roommate as the next student in the chronological RMS-ID order who had compatible gender and other preferences. In the procedure, a roommate was essentially randomly assigned to each student within each gender/smoking/type of room cell. Each year, about 400 returning students were also assigned a roommate randomly, although we did not use such students in our study.

By including both roommates from this type of randomly assigned roommate pairing in our sample, a survey instrument can be used to measure peer influence processes on health-relevant behaviors without the usually attendant threats to causal inference.[2] The key innovation of the 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 is the first time gene-environment interactions have been investigated using randomly assigned environmental influences. This approach should yield far more reliable estimates of gene-environment interactions than purely observational study designs.

[2] Random assignment took place within combinations of room-type and dormitory preferences expressed by applicants. To ensure that analyses of peer effects are based solely on variation induced by random assignment fixed-effect controls for preference, "cells" will be introduced into our regression models. It should also be noted that limiting our data collection to the subset of students opting for random assignment limits the generality of our study, although it does not compromise the random-assignment variation we exploit in our analyses.

3. HISTORICAL CONTEXT OF GENETIC DATA COLLECTION AND ITS POTENTIAL IMPACT

3.1. *Concerns over Collecting Genetic Data*

The contemporary efforts in social science research examining biological influences must be viewed in light of the historical legacy in the United States of the eugenics movement—Nazi Germany's racial "hygiene" campaign, and more recently, ethically questionable medical research in the United States. 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 (Halbert et al. 2006; Kevles 1985). 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 inherited. 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.

3.2. *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 place 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 probable ethnically structured participation rates may be related to the troubling history of biological explanations in social sciences whose effects have disproportionately harmed 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 evaluation of racial and ethnic participation in DNA collection should, therefore, provide crucial information on the validity of social science studies with genetic components.

4. PROTECTION OF HUMAN SUBJECTS

4.1. *Institutional Review Board*

Our applications to the Institutional Review Board (IRB) for the pilot and full studies included a standard written application incorporating a description of the study objectives, procedures, and potential risks of participation; brochures describing our commercial DNA collection kits; and our study budget. For the pilot application, 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 describe the DNA component after students had completed a Web survey, the board requested that we more fully describe our data collection plan at the time of the initial e-mail contact. Based on these recommendations, we included a full description of the study in our correspondence with participants in both the pilot and main studies. 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.

In both projects, the IRB was most concerned with the maintenance of respondent privacy at all stages of the project. The incorporation of genetic data into a social science research study is unfamiliar to many within the academic community. Obtaining IRB approval for this study required careful explanation of security procedures from collection and storage of raw samples to the final disposition of coded data.

DNA collection techniques are a constantly evolving field with changes to collection procedures or changes in public perceptions of safety procedures often precipitating changes to related IRB procedures. As an example, the pilot study was able to make use of in-dorm visits to respondents who were unable to visit the DNA collection office. However, although the full study used similar collection methods, it occurred after the 2007 Virginia Tech shootings and was therefore unable to make use of this procedure as rules for campus safety had changed during the interim period.

4.2. *Certificate of Confidentiality*

Study participants in research studies such as ours may have feared 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 would not occur. To avoid the possibility of compulsory disclosure as the result of civil, administrative, legal, or legislative proceedings, we obtained 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) for the certificate, which required that we submit our IRB application materials and the DNA and Web survey consent forms.

5. THE WEB SURVEY

In addition to DNA collection, both the pilot study and the main study included a Web survey of attitudes and health behaviors. The housing department at the university provided us with a list of 200 students for the pilot study and 3500 students for the full study, all of whom had lived in the university dorms during their first year with a randomly assigned roommate. We initially contacted the students by letter (along with a cash incentive of two one-dollar bills), alerting them they had been selected for participation in the study and that we would contact them subsequently by e-mail. We launched our main study in late January 2008 with an e-mail inviting participation in the Web survey.

Our first e-mail 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. The first page online included a consent letter and a link for students to click on if they agreed to participate. If students consented, they were guided through the Web survey, which took 15 to 20 minutes, on average, for them to complete.

In the ensuing two months, a total of 16 e-mail reminders were sent primarily on a Sunday through Thursday schedule to students who had not completed the Web survey. We made a point of sending an e-mail reminder by Sunday afternoon in hopes that the Web survey might be completed during the remainder of Sunday. Figure 2 shows

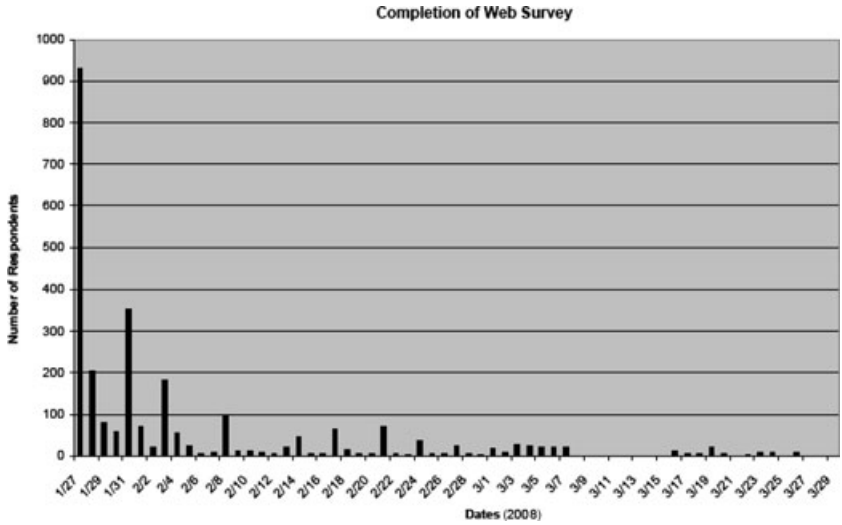

FIGURE 2. Number of Web survey completions by date.

the daily number of Web survey completions for the main study. The actual dates in a year are labeled rather than the number of days since the beginning of the study because the dates may provide additional insights into response rates in a study of health behaviors, including drinking in a college setting. About 30% (930) of the potential participants completed the survey within the first day and the majority of completions (63.3%) for the Web survey occurred within the first week. Survey completion was highly responsive to e-mail reminders with spikes in survey completions occurring on the same day of a survey reminder, although the number of students who responded to the reminder declined considerably over time. We stopped sending e-mail reminders during spring break (from March 8 to March 16). The final response rate for the Web survey was 78% for the pilot and 79.5% for the full study.

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 that Web surveys are 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.

Our cash incentive includes $2 cash enclosed in the initial contact letter sent to all eligible students for the study on our sampling, $5 for completion of the Web survey, and $15 for a DNA sample donation, for a maximum $20 compensation beyond the initial $2 primer. With approval from IRB, we increased the incentive from $20 to $30 in the last 10 days of saliva collection.

6. DNA COLLECTION

6.1. *Saliva Versus Blood as Means of Collection*

Researchers wishing to collect genetic information from their study participants currently have two main means by which to do so: collection of blood or saliva samples. Traditionally human genetic studies collected peripheral blood samples, but new technologies have been developed that collect human DNA from saliva or 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. This conclusion is further supported by a recent genome-wide genetic association study (Hunt et al. 2008). In addition to including saliva DNA from a subset of the individuals in the genome-wide study, Hunt and colleagues genotyped for 1025 SNPs in four subjects using both blood and Oragene saliva and found 99.85% concordance between saliva DNA GoldenGate genotypes and blood DNA Infinium genotypes.

Other studies (Hansen et al. 2007; Philibert et al. 2008) reported higher genotyping success rates for blood DNA than saliva or buccal cell DNA. The success of genotyping using saliva or buccal cell DNA is highly sensitive to quality and concentration of DNA samples. Nevertheless, Hansen and colleagues found that participants were much less likely to provide a blood sample (31%) than a saliva (72%) or buccal cell (80%). Since the nonblood strategy yields a far higher response rate, involves far fewer serious IRB issues, is far less expensive, and does not require medical personnel to collect and handle the samples, we judge this far less invasive means especially suitable for large-scale social science research.

6.2. *Preparation for DNA Collection*

Until recently, buccal cell DNA was collected directly from a subject's cheek using a cotton swab or mouth rinse (Freeman et al. 1997; Freeman et al. 2003; Meulenbelt et al. 1995). However, our study utilized a simpler collection technique, the Oragene DNA self-collection kit. The Oragene kit has been used successfully in previous studies (Ahituv et al. 2006; McCready et al. 2005) and was selected for use in the Add Health Wave IV full sample DNA collection effort slated for 2008–2009. The kits contain small collection vials containing 2 milliliters of a preservation liquid, come with collection instructions, and are similar in appearance to a single, large contact lens case. Our DNA collection process using the Oragene kits is as follows. If participants 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. During our pilot study we found that many research participants found it difficult to produce enough saliva to fill the collection vial to the marked line. Accordingly, for the full study 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 nonlatex 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. During the pilot study, staff observed that some study participants appeared very uncomfortable supplying the required level of

saliva. In the pilot, the participants provided samples in full view of the staff. Study staff reported that participants would turn their backs to the staff or report feeling uncomfortable donating saliva in front of another person. Based on our knowledge from the pilot study, we decided to maximize participant privacy in the full study. Our collection office was located in an office containing two rows of desks with each desk separated by partitions that allowed ten respondents to provide saliva samples simultaneously in privacy. Outside the office were several chairs set up to provide a waiting area. Study staff would remain nearby to answer any questions during this process, but they kept the appearance of being busy working at a computer console, so that participants did not feel as if the staff were hovering while they completed the process. These arrangements had the advantage of increasing the sense of privacy that respondents felt, and this increase in their comfort levels helped speed the collection process.

6.3. *DNA Collection Procedures*

In the pilot and full study, respondents automatically received an e-mail after completing the online survey. The e-mail informed students of the location and hours of our office where they could contribute a DNA sample and/or collect their compensation. We elected to collect participant DNA in a central campus study office near a location where students regularly attend classes and a secondary location closer to student dormitories for maximum convenience and privacy.

Potential study participants were more reluctant to contribute a saliva sample than respond to a Web survey in part because of the extra time it took for students to come to the collection office in person. To maximize participation rates in saliva collection, we again depended on e-mail communication. Study staff varied the text of each e-mail to avoid redundancy. The length of the e-mails varied as well; when e-mails were sent in close succession to one another, the later e-mails were designed as quick refresher notes. Each e-mail included the contact information and office hours for the study, and it 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.

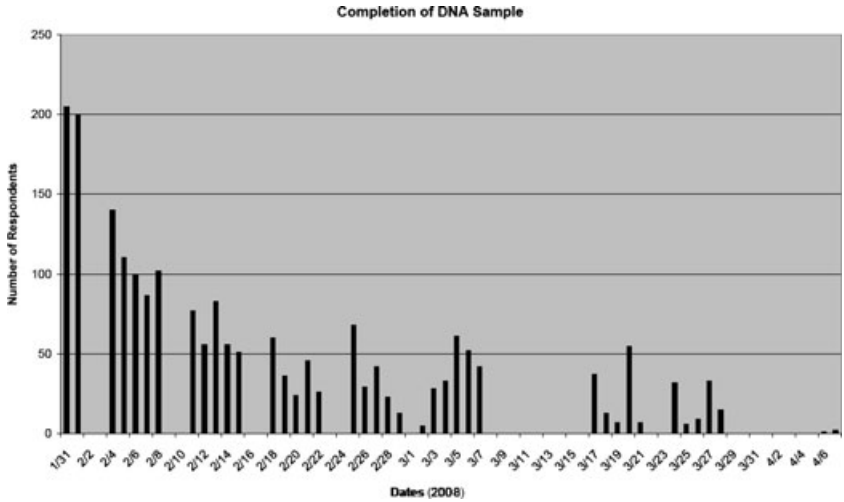

FIGURE 3. Number of DNA sample completions by date.

Figure 3 plots the number of students who contributed a DNA sample by date. To ease initial crowding, we did not send e-mail reminders for DNA collection to all survey respondents at the same time. Instead, only 300 students were reminded of DNA collection in each of the five days during the first week of our full study. Out of the 2080 students who contributed a DNA sample, 205 and 200 came in the first and second days, respectively; about 1000 made the contribution during the first two weeks. As in the Web survey, potential study participants proved responsive to our e-mail reminders, with most spikes in Figure 3 corresponding to e-mail reminders that had been sent the night before. In the final two weeks of the study, after the spring break, we raised the incentive from $20 to $30 for those who would complete the Web survey and donate a DNA sample. We ran out of our DNA collection supplies on the Friday afternoon of March 28, but at this point we had already reached our target response rate determined from the experiences of the pilot study and our budget constraints.

Upon entering our office, students were directed to an area where they could sit at a desk equipped with partitions to provide privacy. Respondents were then handed a 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 between two options concerning long-term DNA storage. After agreeing

to participate in this part of the study, respondents were provided with a saliva kit and paper towel to place the kit on and use for cleaning if necessary. Study staff then gave a quick description of how to provide a saliva sample and left the students to complete the process in privacy.

During the pilot study, staff recorded the length of time a respondent took to complete the entire process. 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. These estimates include a short 2-minute debriefing survey that asked respondents about their experiences with saliva collection.

We found during the pilot and full study a portion of respondents expressed concerns regarding DNA collection, use, and privacy. To assuage the concerns, the research team drafted an official response to these concerns for the pilot and main studies on why the study sought to collect respondents' DNA. The study contact person sent this e-mail as the first response to any respondent who personally contacted her or him with fears or questions about this phase of the study.

In the pilot study, 83.2% of those who had completed the survey provided a saliva sample. The response rate for DNA collection in the main study is consistent with that in the pilot study, with 78.7% of those who had completed the survey providing a saliva sample. The overall response rate of completing both the Web survey and donating a saliva sample is 62.5%. According to the online article entitled "Biomarkers in Wave III of the Add Health Study," Wave III of Add Health collected about 2600 saliva samples from 3787 respondents who were identified to be full siblings or twins at earlier waves and who were asked to provide a saliva sample for DNA analysis with a response rate of 70.2%. Locating and visiting respondents' homes on the part of Add Health may explain its greater success.

7. DNA PROCESSING AND GENOTYPING

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

A genotyping facility on the university campus then carried out the test genotyping.[3] We genotyped one SNP[4] 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 four blind duplicated controls. Five ng DNA were used for each SNP. Both SNPs were successfully detected and the call rate (the rate at which a SNP can be reliably decoded) 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.

In the main study, DNA was extracted according to the manufacturer's instructions from 2 ml of saliva (containing buccal epithelial and white blood cells) collected from participants in an Oragene disc DNA collection kit (DNA Genotek; Ottawa, Ontario, Canada). Kits were collected over a period of several months and remained at room temperature until batch processing began. DNA was quantitated using the Quant-it Picogreen DNA quantization kit (Invitrogen, Carlsbad, CA). DNA was plated for Illumina genotyping at 30 ul at >50 ng/ul. Our median yield was 27.33 ug, the min was 0 (from six samples), and the max was 71.32.

We designed an Illumina GoldenGate assay for 384 candidate SNPs including about 180 ancestral informative markers (Hodgkinson et al. 2008) and tagSNPs as well as putatively functional SNPs. A total of 352 SNP assays have succeeded—that is, 352 SNPs out of 384 have been successfully genotyped. Preliminary Hardy-Weinberg equilibrium tests were performed on each SNP within each ethnicity. Fewer than

[3] The genotyping was carried out using the Applied Biosystems TaqMan genotyping technology.

[4] Single-nucleotide polymorphism, commonly referred to as 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 that may result in differently functioning genetic processes—variation that may in some cases provide statistical traction on outcomes of interest to social scientists.

1% of the SNPs yield a P-value smaller than 0.00001 or 0.001. Nineteen of the SNPs have a MAF smaller than 0.01.

For the purpose of quality control, the DNA samples for genotyping included 92 parent-parent-child trios and 46 duplicated samples. In the latter case, 46 of our study participants were randomly selected to be genotyped twice. For each pair of duplicates, one sample was labeled with the original ID and the other was label as blind with all related information wiped out. The trios allow a check on Mendelian inheritance and the duplicates test if the genotyping could be reproduced. Our genotyping has passed both tests.

To identify excessively similar pairs of individuals (due to a person taking part in the study more than once or unsuspected identical twins), we calculated identity-by-state (IBS) for all pairs of individuals in the data. This amounts to $N(N-1)/2$ unique pairs of individuals for a sample of N. The calculation identified three pairs of identical twins who share the same genotype, the same permanent address, the same last name, and the same birthday. The calculation also identified two pairs of problematic individuals who have exactly the same genotype but a different birthday and name. Apparently, these two individuals each contributed saliva twice.

The genotyping results also helped us correct two typos. The genotyped gender differed from the self-reported gender for two participants. We went back and checked the University Housing recorded gender, which is consistent with the genotyped gender. We concluded that the self-reported gender must have resulted from a typo.

8. OVERALL PARTICIPATION RATES AND DIFFERENTIAL PARTICIPATION RATES

8.1. *Overall Participation Rates*

Participation rates for both the pilot and main studies are reasonably high. In the pilot study, we received an initial sample of 200 from the University Housing Department. Of these, 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 e-mail contact information listed with the Housing Department. Table 1 displays overall participation rates at each stage of the pilot. Of the eligible sample of 191, 149 or 78% completed the Web survey. Of those who completed the

TABLE 1
Participation Rates for the Pilot and Main Studies

	Pilot Study		Main Study	
	Number	Participation Rate	Number	Participation Rate
Students completed survey	149	78%	2664	79.5%
Students provided DNA	124	83.2% of survey respondents 64.9% of eligible	2080	78.7% of survey respondents 62.5% of eligible

online Web survey, 83.2% also completed the DNA collection phase, representing an overall participation rate of 64.9%. Of the remaining eligible sample members who did not complete the DNA stage of the pilot, we encountered only two explicit refusals to participate.

In the main study, 2664 students completed the Web survey. Adjusting for those who did not live on campus, who were too young to be included in the alcohol study, and who were in a foreign country in a study-abroad program for the semester, 79.5% of the eligible students finished the online survey. The adjustment includes excluding individuals in the study-abroad program from both the denominator (114 individuals) and the numerator (15 individuals). In the main study, 2080 students provided a saliva sample. Excluding the ineligible (e.g., individuals in the study-abroad program), the response rate for completing both the online and DNA parts of the study is 62.5%.

8.2. *Participation by Race and Gender*

In Table 2 we compare the ethnic and gender composition among both survey completers and DNA providers against those in the study population on the university campus. The gender distribution in our studies, especially in the main study, is quite close to that in the student population. The female-male split in the 2005 and 2006 freshman enrolling classes at the university is 60.3–39.7%. In the main study, 60.6–39.4% of the survey respondents are females and males, respectively; the corresponding gender distribution for DNA providers is 60.1–39.9%.

Our study deliberately oversampled roommate pairs in which one or more students are African American or Hispanic. This oversampling is reflected in the ethnic/racial distribution in Table 2, where

TABLE 2
Percent Distribution of Student Demographic Characteristics in the Study Population

		Pilot Study		Main Study	
	Student Population*	Web Survey Participants	DNA Participants	Web Survey Participants	DNA Participants
Sex					
Male	39.7	37.58	37.90	39.4	39.9
Female	60.3	62.42	62.10	60.6	60.1
Race/ethnicity					
Black	11.7	14.77	16.13	15.3	16.6
White	75.3	67.79	65.32	67.8	66.6
All non-White or Black	13.0	17.45	18.45	16.9	16.8
Asian/ Pacific Islander	N/A	N/A	N/A	8.5	8.0
Latino	N/A	N/A	N/A	6.9	7.4
Other	N/A	N/A	N/A	1.5	1.5

*Data for freshman enrolling classes, 2005 and 2006, provided by admissions department at the study university.

African American students and other minority students are overrepresented among both survey completers and DNA providers. For example, 15.3% of the survey completers and 16.6% of the DNA providers in the main study are African American versus 11.7% in the general study population.

As a further check of possible racial and gender shifts between the Web survey and DNA collection components, we compare the DNA collection rates among survey completers for each race-gender subgroup. Black students who completed the survey exhibit the highest levels of conditional DNA participation in both studies, at 90.9% for the pilot and 83.8% for the main study. Black students may be more responsive to the monetary incentive than their Caucasian counterparts. White students evinced the lowest completion rates, at 80.2% in the pilot and 76.2% in the main study. This information is presented in Table 3. The marginal distribution of DNA completion given survey completion by gender reveals a slightly higher completion rate for males. In sum, these results reveal slight gender discrepancies in DNA completion,

TABLE 3
Percentage of Web Survey Completers Who Provided Saliva Samples

	Pilot Study			Main Study		
Race	Male	Female	Total	Male	Female	Total
Black	100.0	87.5	90.9	85.1	83.2	83.8
White	79.0	81.0	80.2	77.6	75.3	76.2
All non-White or Black	91.7	85.7	88.5	78.0	76.5	77.2
Asian/Pacific Islander	N/A	N/A	N/A	74.4	72.6	73.3
Latino	N/A	N/A	N/A	84.1	81.6	82.7
Other	N/A	N/A	N/A	68.4	77.3	73.2
Total	83.9	82.8	83.2 (N = 149)	78.6	76.9	78.7

and a racial pattern of DNA participation opposite to that predicted by prior medical DNA participation studies.

8.3. *Participant Experiences of DNA Collection*

Any study of socially sensitive topics such as 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 in the pilot study. 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 discreetly observed participants during the DNA collection process and made notes on their behavior, demeanor, and speech. These were later converted to a dummy variable indicating whether participants expressed or demonstrated discomfort 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. We did not collect specific measures of comfort in the main study. However, we once again asked respondents to give their consent to long-term DNA storage. We report the differences in consent by gender and racial/ethnic background here.

Table 4 reports the percentage of participants willing to contribute DNA again, expressing concern about the process, and agreeing

TABLE 4
Indicators of Comfort by Sex and Race/Ethnicity

	Pilot Study				Main Study
	Contribute DNA Again	Concern About Process	Agreed to Long-term Storage	Reported Comfort Level†	Agreed to Long-term Storage
Sex					
Male	97.8%	8.9%	63.8%	4.5	73.2%
Female	98.7%	12.5%	63.2%	4.3	73.5%
P-value*	—	.55	.94	.26	.91
Race/ethnicity					
Black	95.5%	11.1%	60.0%	4.2	63.7%
White	98.9%	11.7%	70.0%	4.4	77.1%
All non-White or Black	100%	9.9%	43.5%	4.4	67.9%
P-value*	—	.94	.06	.06	.00
Asian/Pacific Islander	N/A	N/A	N/A	N/A	63.6%
Latino	N/A	N/A	N/A	N/A	71.7%
Other	N/A	N/A	N/A	N/A	72.4%
P-value*	N/A	N/A	N/A	N/A	.00
Total	98.4% (N = 123)	11.1% (N = 123)	63.4% (N = 123)	4.4 (N = 123)	73.4% (N = 2080)

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

†Average reported comfort level from a 5-point scale where 1 means "not at all comfortable" and 5 means "very comfortable with giving a saliva sample."

to long-term storage, as well as the mean level of reported comfort by gender and racial/ethnic group. Strikingly, we observe very high rates of willingness to participate again in a hypothetical, identical study. Males and females were about equally likely to say that they would contribute DNA again for a similar study, and only small racial/ethnic differences were observed; Black participants were slightly less likely to say they would participate again. We did not observe any significant difference in participants expressing concern about the project by gender or race/ethnicity. 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).

In both the pilot and main study, males and females agreed to seven-year storage of their genetic material at nearly identical rates.

However, large racial differences were observed for this measure of comfort. White participants were the most likely to agree to long-term storage in both studies, at 70% in the pilot and 77% in the main study. In both studies, Black participants had similar rates of consent, at about 60–64%. Consent rates for participants from an "other" racial/ethnic background differed between studies; 43.5% of this group consented in the pilot, compared to 68% in the main study. Using detailed racial/ethnic classifications in the main study, it appears that Asian/Pacific Islanders were less likely to consent to long-term storage than Latinos.

8.4. *Specific Participant Concerns*

Finally, we recorded specific participant concerns in the pilot study, in order to inform our planning of the main study. When asked, a small number of sample members expressed concerns about the collection of their genetic information. Additionally, Table 4 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. Other participants frequently asked for clarification regarding the purpose of the genetic portion of the project.

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 health 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. This response also emphasized the confidentiality and privacy measures included in the research design.

A number of participants were apprehensive 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 address sample members' apprehensions, both to minimize psychological costs to participants and to maximize participation rates. Data on participant concerns were not collected in the main study. However, during the data collection process for the main study, we made use of the oral and written responses developed during the pilot study.

8.5. *Sampling Roommate Pairs*

One of the key design features for PIADB is the sampling of roommate pairs. Our analyses of peer influences on health-related behaviors required obtaining data from both individuals in a roommate pairing. Therefore, we took certain measures in an attempt to maximize the paired response rates for both portions of the study.

First, to be eligible for inclusion in our sample, students must have been residents in a university dormitory during their first year on campus with a randomly assigned roommate. However, our sample was not limited to first-year students. Because some students may have left the university between their first year and when the sample was drawn, we selected only roommate pairs where both students were still enrolled in the university.

Our ultimate goal was to get as many paired survey and saliva samples as possible. However, at a minimum we needed paired survey data for tests of peer influences and paired survey data with DNA samples from at least one of the pair for genetic tests of susceptibility to peer influences. Therefore, during the course of the full study, we organized respondents into groupings based upon their pair completion status. Two groupings were targeted as priority samples: (1) pair groups where one student had completed nothing while his or her roommate had completed both the Web survey and provided a saliva sample and (2) pair groups where one student had completed the survey and

his or her roommate had completed both the survey and provided a saliva sample. About 61% of the study participants were in pairs in which both individuals completed the Web survey, and 57.7% of the study participants were in pairs in which both individuals completed the Web survey and in which at least one individual provided saliva DNA.

9. DISCUSSION AND CONCLUSION

Our PIADB study successfully paired traditional social science survey research with DNA collection by using a Web survey instrument and Oragene DNA self-collection kits. In both the pilot and main study, no major demographic differences in participation rates were observed by race/ethnicity or gender. Racial/ethnic minority participation is, if anything, higher than average, which runs counter to what history and previous research would suggest. In both the pilot and the full studies, members of all racial/ethnic minority groups agreed to long-term storage of their DNA samples at significantly lower rates than Whites. Large numbers of students completed the survey and DNA portions of the study for modest compensation and without undue pressure. The structure of our data promises methodologically innovative and substantively interesting analyses to come.

We emphasize that studies of joint influences of social environment and genome on human behavior constitute an important and underutilized frontier in research that other social scientists should—and, our results show, *can*—pursue. Our experiences in these studies 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.

Add Health pioneered DNA data collection among a subset of about 2600 of its participants at Wave III in 2002 (Harris et al. 2003). The DNA data collection was designed by Add Health investigators and implemented by professional data collection organizations.

We acknowledge that our sample is unrepresentative of the U.S. population, as our participants hailed from a university setting. College students are by definition better educated than the population at large and more likely to have some familiarity with research. While we believe that most of our experiences of collecting genetic data are generalizable

to a wider population, participation may be more modest in other settings.

10. 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 for requesting genetic samples at the outset, taking all possible steps to preserve participants' anonymity and confidentiality, and taking time to reduce participants' legitimate apprehensions, respondents will be more comfortable with the process, costs of participation will be minimized, and trust will be established with potential respondents. In the interests of consistency and efficacy, we recommend developing standard written and oral responses to typical concerns and considering incorporating them into your correspondence with sample members.
- *Provide monetary compensation.* Recall the finding of Crider et al. (2006) 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 nonparticipation 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 that we then sought to alleviate; presumably many such persons would not have come by the office were it not for the compensation. Perhaps most importantly, compensating sample members for their genetic donations demonstrates recognition of the 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 data sets become publicly available. DNA collection operations should be placed in discreet 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 securing 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 participants 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 about 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 likely increase participation rates.

REFERENCES

Ahituv, N., N. Kavaslar, W. Schackwitz, A. Ustaszewska, J. M. Collier, S. Hebert et al. 2006. "A PYY Q62P variant Linked to Human Obesity." *Human Molecular Genetics* 15:387–91.

Bhatti, P., A. J. Sigurdson, S. S. Wang, J. B. Chen, N. Rothman, P. Hartge et al. 2005. "Genetic Variation and Willingness to Participate in Epidemiologic Research: Data from Three Studies." *Cancer, Epidemiology, Biomarkers and Prevention* 14:2449–53.

Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." *American Economic Review* 96:1890–905.

Committee on Population. 2001. *Cells and Surveys: Should Biological Measures be Included in Social Science Research?* Washington, DC: 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." *American Journal of Epidemiology* 164:805–12.

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* 33:375–85.

Frayling, T. M., N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren et al. 2007. "A Common Variant in the FTO Gene is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity." *Science* 316:889–94.

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* 27:251–57.

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." *Behavior Genetics* 33:67–72.

Guo, G., E. M. Roettger, and T. Cai. 2008. "The Integration of Genetic Propensities into Social Control Models of Delinquency and Violence Among Male Youths." *American Sociological Review* 73:543–68.

Halbert, C. H., O. H. Gandy, A. Collier, and L. Shaker. 2006. "Intentions to Participate in Genetics Research Among African American Smokers." *Cancer Epidemiology Biomarkers and Prevention* 15:150–53.

Hansen, T. V. O., M. K. Simonsen, F. C. Nielsen, and Y. A. Hundrup. 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 Biomarkers and Prevention* 16:2072–76.

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," vol. 2005. Retrieved February 26, 2008 (http://www.cpc.unc.edu/projects/addhealth/design).

Hernandez, L. M., and D. G. Blazer, eds. 2006. *Genes, Behavior, and the Social Environment: Moving Beyond the Nature/Nurture*. Washington, DC: National Academies Press.

Hodgkinson, C. A, K. Xu, Q. Yuan, P. H. Shen, E. Heinz, E. A. Lobos et al. 2008. "Addictions Biology: Haplotype-Based Analysis for 130 Candidate Genes on a Single Array." *Alcohol and Alcoholism* 43(5):505–15.

Hunt, K. A., A. Zhernakova, G. Turner, G. A. R. Heap, L. Franke, M. Bruinenberg et al. 2008. "Newly Identified Genetic Risk Variants for Celiac Disease Related to the Immune Response." *Nature Genetics* 40:395–402.

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* 117:285–87.

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* 57:1252–54.

Moorman, P. G., C. S. Skinner, J. P. Evans, B. Newman, J. R. Sorenson, B. Calingaert et al. 2004. "Racial Differences in Enrolment in a Cancer Genetics Registry." *Cancer, Epidemiology, Biomarkers, and Prevention* 13:1349–54.

Philibert, R. A., O. Zadorozhnyaya, S. R. H. Beach, and G. H. Brody. 2008. "Comparison of the Genotyping Results Using DNA Obtained from Blood and Saliva." *Psychiatric Genetics* 18:275–81.

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

Scott, L. J., K. L. Mohlke, L. L. Bonnycastle, C. J. Willer, Y. Li, W. L. Duren et al. 2007. "A Genome-wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants." *Science* 316:1341–45.

Sladek, R., G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre et al. 2007. "A Genome-Wide Association Study Identifies Novel Risk Loci for Type 2 Diabetes." *Nature* 445:881–85.

Steinthorsdottir, V., G. Thorleifsson, I. Reynisdottir, R. Benediktsson, T. Jonsdottir, G. B. Walters et al. 2007. "A Variant in CDKAL1 Influences Insulin Response and Risk of Type 2 Diabetes." *Nature Genetics* 39:770–75.

Tapper, M. 1999. *In the Blood: Sickle Cell Anemia and the Politics of Race. Critical Histories*. Philadelphia: University of Pennsylvania Press.

Zeggini, E., M. N. Weedon, C. M. Lindgren, T. M. Frayling, K. S. Elliott, H. Lango et al. 2007. "Replication of Genome-wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes." *Science* 316:1336–41.