

The Feasibility of Creating a Population-Based National Twin Registry in the United States

Patricia C. Chulada,¹ Linda A. Corey,² Vani Vannappagari,³ Nedra S. Whitehead,⁴ and Perry J. Blackshear¹

¹ Program in Clinical Research, National Institute of Environmental Health Sciences, North Carolina, United States of America

² Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, United States of America

³ GlaxoSmithKline, Research and Development, North Carolina, United States of America

⁴ Statistics and Epidemiology, Research Triangle Institute, Atlanta, Georgia, United States of America

Between 4 to 6 million twins exist in the US today who offer scientists a valuable potential resource for conducting behavioral and biomedical research. However, unlike many other countries, there is no national system in the US for identifying twins and eliciting their participation in these important research programs. Therefore, the National Institute of Environmental Health Sciences (NIEHS) is conducting a study to determine the feasibility of creating a national, population-based twin registry in the US. The major goal is to estimate the potential size and characteristics of a national twin registry based on the current twin population in the US, our ability to ascertain and enrol them, and their willingness to participate. Existing US twin cohorts are also being examined in this study as well as alternatives for improving US twin resources should a national twin registry be deemed infeasible. The various options will be compared in terms of possible source populations, generalizability and adequacy for statistically powering various types of etiological studies. Two expert advisory panels have been assembled to assist in the conduct of this study. The Scientific Advisory Panel is charged with providing expertise concerning study goals, design and methodology, and evaluating the study's conclusion. A separate Ethics Advisory Panel is charged with providing expertise on the ethical, legal, and social issues that might be encountered if a national twin registry is ultimately pursued. Having a national population-based twin registry in the US would be advantageous to US scientists and those worldwide. It would provide ample numbers of twin pairs to conduct various types of environmental genomic studies currently not possible with existing US twin resources. It would also allow US scientists to select for characteristics (race, ethnicity, environments, and so on) inherent in our own population. Finally and foremost, it would help to meet the worldwide demand for twin resources which is expected to increase over time, as new genomic and analytical tools become available and new hypotheses emerge

concerning the complex interplay between genes, lifestyles and environment.

The classical twin study is considered to be the definitive method for estimating the relative importance of genetic and environmental factors in the determination of variation in complex traits and disease. Now with new genomic and analytical tools such as high throughput genotyping, multivariate analyses and structural equation modeling, twin studies have gone far beyond the classical design and have become powerful tools for identifying causative genes and mapping quantitative trait loci (QTL). Twin studies also have advantages over other study designs for identifying environmental risk factors and resolving the complex interactions between genes and the environment (MacGregor et al., 2000).

While the inclusion of twins enhances the statistical power of a genetic study by exploiting the genetic and environmental sharing between co-twins, large sample sizes of intact twin pairs are still needed. For example, from 200 to several thousand pairs might be required to estimate the degree of influence of a QTL depending on the importance of genetic effects and whether the trait under study is rare, dichotomous or demonstrates genetic nonadditivity (Martin et al., 1997). Large numbers of twin pairs are also needed for multivariate analyses of complex conditions involving environmental risk factors and gene-environment interactions, to control for multiple confounders and stratify for the exposure or other factors. Large population-based twin cohorts are ideal for nonbiased screening and offer the best option for identifying adequate numbers of twin pairs with the condition, even if the condition is rare (Boomsma et al., 2002).

Received 1 August, 2006; accepted 14 September, 2006.

Address for correspondence: Patricia C. Chulada, Program in Clinical Research, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC, USA. E-mail: chulada@niehs.nih.gov

Several large, population-based twin registries have been established in the US; however, few contain the number of twins necessary to support in depth etiological studies of complex traits, especially if potential confounding or disease prevalence dictates stratification based on gender, ethnicity, age, exposures or other factors. Therefore, the National Institute of Environmental Health Sciences (NIEHS) is examining the feasibility of creating a national, population-based twin registry in the US that will be large enough to support studies of this type. Between 4 to 6 million twins exist in the US today who offer scientists a valuable potential resource for conducting behavioral and biomedical research. Yet unlike many European and other nations, there is no national system in the US for identifying twins and eliciting their participation in these important research programs. We expect that as new genomic/analytical tools become available and new etiologic hypotheses are generated, the demand for twin resources will increase. Having a population-based US twin registry would help to meet this demand. It would also allow for selection of characteristics inherent in the US population and the generation of results that are more generalizable to this population.

There are a number of unique challenges that must be overcome if a national, population-based twin registry is to be established in the US. These include the absence of an accessible centralized national database through which twins can be easily identified, heightened concerns over privacy, minority-associated recruitment barriers, and a more conservative attitude of the general population towards research. The major goals of this project are to estimate the potential size and composition of a US national twin registry based on the existing US twin pool, determine the ascertainment and enrolment methods most likely to be successful in reaching the target populations, and assess the willingness of twins to participate in such a registry in the current ethical, legal, and social climate. Other alternatives for improving US twin resources are also being explored and their efficacy will be compared to that of a national twin registry in terms of their potential value to research (statistical power, generalizability) and costs. At the end of this 2-year project, recommendations for improving US twin resources will be made to the National Institutes of Health and other government organizations.

Study Investigators and Expert Panels

The principal investigators of this study are Patricia Chulada (health scientist administrator, Program of Clinical Research, NIEHS), Linda Corey (Professor of Human Genetics, Virginia Commonwealth University), Vani Vanappagari (senior researcher, Alpha Gamma Technologies, Inc), and Nedra Whitehead (senior researcher, Research Triangle Institute, International). Patricia Chulada is a genetic toxicologist with experience in genetic environmental studies and DNA registries, and is currently conducting a twin study to

examine the genetic and environmental risk factors for polycystic ovary syndrome. Linda Corey is a genetic epidemiologist and past co-director of the Mid-Atlantic Twin Registry (MATR). She has extensive expertise in twin studies, the analysis of twin data, and population-based twin registries. Her research foci are seizures and epilepsy, fetal programming regarding syndrome X related phenotypes, prematurity and complications of pregnancy. Vani Vannappagari is a physician and epidemiologist with research experience in international surveys, Human immunodeficiency virus (HIV) behavioral research, and cancer and cardiovascular research. Nedra Whitehead, a genetic epidemiologist, has experience with genetic screening and testing, and sample surveys based on vital statistics data. Other research team members include biostatisticians, survey specialists, a demographer, economist, and others. Perry J. Blackshear, Director of the Program in Clinical Research at the NIEHS, is responsible for initiating the project and will present the findings to the National Institutes of Health and other governmental agencies. Lenn Murrelle, special consultant to the project, is also a past co-director of the MATR with expertise in population-based epidemiologic resources including twin registries. Linda Corey and Lenn Murrelle have played key roles in this project. With experience gained in establishing the Virginia (Corey) and North Carolina (Murrelle) Twin Registries, they transformed the two registries into the MATR which is currently the largest population-based twin registry in the United States (Anderson et al., 2002).

Two advisory panels have been assembled in conjunction with this project. One (Table 1), chaired by Drs. Chulada and Corey, provides expertise concerning study goals, design, and methodology, discusses study results and makes recommendations. The second (Table 2), chaired by Dr. Paul Lombardo, Associate Professor of the School of Medicine and Director of the Program in Law and Medicine at the Center for Biomedical Ethics at the University of Virginia, provides expertise on the ethical, legal, and social issues (ELSI) that might be encountered in establishing a national twin registry. Dr. Lombardo has served on a number of national panels, providing expertise on the proper conduct of research involving human beings as subjects.

Study Objectives and Methods

Objective 1: Estimate the Size and Composition of the Current US Twin Population

To date, there has been no systematic attempt to describe the US twin population with respect to numbers and demographics. The National Center for Health Statistics (NCHS), the United State's principal health statistics agency, collects data on twin births and reports on the trends for multiple birthing, but does not provide any specific information regarding the existing US twin population. Therefore, our first objective was to estimate the current size of the US twin population and stratify

Table 1

Expert Advisory Panel

| Name | Affiliation | Area of Expertise |
|----------------------------|--|--|
| Co-chair, Linda Corey, PhD | Virginia Commonwealth University Mid-Atlantic Twin Registry (MATR) | Twin studies, registry director — epilepsy/seizures, pregnancy complications, blood/hypertension, |
| Lenn Murelle, PhD | Molecular and Statistical Epidemiology Group, Philip Morris | Twin, pharmaco- and molecular genetic studies — alcohol and drug abuse, depression, suicidality |
| Jack Goldberg, PhD, MA | University of Washington Vietnam Era Twin Registry; Black Ederly Twins Registry (BETS) | Twin registry founder, twin studies combat exposure, physical and psychological health of veterans |
| Jennifer Harris, PhD | Director, Norwegian Institute of Public Health; Ethicist Chair, GenomeEUtwin | Twin studies — obesity, asthma, health perceptions and psychological well-being |
| William Page, PhD | Institute of Medicine, National Academy of Sciences; Director, NAS-NRC Twin Registry | Twin studies — veteran's health studies, chronic disease epidemiology, multiple sclerosis |
| Andrew Heath, DPhil | Washington University, St. Louis; Director, Missouri Alcoholism Research Center | Twin studies — Substance abuse and personality disorders, genetic epidemiology, toddlers and adolescents |
| Matt McGue, PhD | Psychology Department, University of Minnesota; Minnesota Twin Family Study | Twin studies — substance use of adolescence to adulthood, genetics of normative aging and mortality |
| Lon Cardon, PhD | University of Oxford; Wellcome Trust Center for Human Genetics | International HapMap Project, statistical genetics, Huntington's Disease, reading disability and endometriosis |
| Paul Lombardo, JD, PhD | Director, Program in Law and Medicine, Center for Biomedical Ethics, University of Virginia | Health law, genetics and the law, legal and ethical regulation of research, bioethics history |
| Stephen Rich, PhD, MS | Department of Public Health Sciences, Wake Forest University School of Medicine | Genetic basis of diabetes and diabetic nephropathy, Type 1 Diabetes Genetics Consortium |
| John Harley, MD, PhD | Oklahoma Medical Research Foundation | Systemic lupus erythematosus, Lupus Multiplex Registry and Repository |
| Dessa Sadovnick, PhD | Departments of Medical Genetics and Medicine, University of British Columbia | Multiple Sclerosis and Alzheimers, Canadian Collaborative Project on Genetic Susceptibility to MS |
| Mary Anne Dooley, MD, MPH | Thurston Arthritis Research Center, University of North Carolina, Chapel Hill | Systemic lupus erythematosus and lupus nephritis, Glomerular Disease Collaborative Network |

them for certain demographics and other twin characteristics. First, we obtained US natality and mortality data from the NCHS for each year from 1920 to 2002. Population mortality rates were derived from the data using period life table techniques (Keyfitz & Caswell, 1985) and then applied to all the twin birth cohorts (twins born in the same year) in the same period. This allowed us to estimate the number of twins alive in the year 2004 partitioned by age. Linked birth/infant mortality data (also from the NCHS) were used in the analysis to account for the higher infant mortality rates for twins. With this approach, we estimated a higher number of US twins (~5.3 million individual twins) compared with a previous estimate of 4 million twins that was published on two twin-related websites (http://www.keepkidshealthy.com/twins/twin_statistics.html and <http://www.twinsfoundation.com/researchupdate/ru-v4n2-1989.htm>). These website estimates are the only US twin population estimates published to date, and neither website provided any

supporting documentation on how the estimates were derived.

Based on the vital statistic data and other data obtained from the NCHS, we further stratified the twins by gender, race/ethnicity, socioeconomic status, geographic location by state and urban living categories. We estimated zygosity by applying monozygotic (MZ) and dizygotic (DZ) twinning rates to the population assuming that the rate of MZ twinning has been consistent over time and across populations, and that the mortality rates between MZ and DZ twins are about equal. Co-twin survival status was estimated from the co-twin mortality rates from other population-based twin registries, but a more in-depth co-twin survival analysis is planned using life table methodology. Detailed information on the current size and composition of the US twin population will be presented in a future publication. This represents the first endeavor ever taken to describe the US twin population with respect to size and demographics.

Table 2

Ethical, Legal and Social Implications Expert Advisory Panel

| Name | Affiliation | Area of Expertise |
|-----------------------------------|--|--|
| Chair, Paul Lombardo, JD, PhD | Director, Program in Law and Medicine, Center for Biomedical Ethics, University of Virginia | Health law, genetics and the law, legal and ethical regulation of research, bioethics history |
| Linda Corey, PhD | Virginia Commonwealth University, Mid Atlantic Twin Registry (MATR) | Twin studies, registry director — epilepsy/seizures, pregnancy complications, blood/hypertension, |
| Jack Goldberg, PhD | University of Washington Vietnam Era Twin Registry | Twin studies — combat exposure, long-term physical and psychological health of veterans |
| Jennifer Harris, PhD | Director, Norwegian Institute of Public Health Ethicist Chair, GenomeEUtwin | Twin studies — obesity, asthma, health perceptions and psychological wellbeing |
| Jeffrey R. Botkin, MD, MPH | Human Genome Center, University of Utah, School of Medicine | Associate VP for Research Integrity, ELSI of genetic technology, cancer susceptibility, |
| Goldie Byrd, PhD | Professor and Chair, Department of Biology, NC A&T State University | Genetics of Alzheimer Disease in African Americans, health disparities in African Americans |
| Mildred K. Cho, PhD | Stanford Center for Biomedical Ethics, Stanford University School of Medicine | Ethical and social impacts of genetic testing and gene therapy, stigmatization of genetic conditions by race and ethnicity |
| Jessica Hobart, MPP | Bloomberg School of Public Health, Johns Hopkins University | Patient advocate and patient rights, mastocytosis research |
| Robert C. Millikan, DVM, MPH, PhD | Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill | Epidemiologist, gene–environment studies on breast and colon cancer, and malignant melanoma |

Objective 2: Evaluate Existing United States Twin Resources

As part of this project, we felt it was important to evaluate already existing US twin resources and assess their adequacy for human genomic environmental research. Some twin registries were referred to us by our research team members or Expert Advisory Panel. Others were discovered by conducting literature and web searches. A total of 19 existing twin registries were examined. These varied in size, composition and research focus; from large population-based twin registries used for behavioral and biomedical research to small cohorts assembled for specific projects.

Data were collected for each registry from their websites, literature searches, or from the directors in a formal survey. The data included (1) registry characteristics (size, length of existence, research focus, enrolment criteria); (2) twin cohort descriptors (size, demographics, zygosity); (3) registry ascertainment and enrolment strategies, and response rates; (4) attrition and retention strategies; (5) procedures being implemented for human subjects protection; (6) data/biological sample collection methods; (7) data security procedures; and (8) policies for sharing data and samples or accessing twins. We also explored the ethical and social issues that these registries have encountered and special strategies used to increase the recruitment of members of minority and other under-represented populations. In addition to permitting an evaluation of these resources for future research studies, this information will also be valuable if the development of a national US twin registry is ultimately pursued. If, however, a national US twin

registry is not deemed to be feasible, this information could then be used to develop alternate strategies for improving US twin resources.

Objective 3: Evaluate the Ethical, Legal and Social Issues Associated With Establishing a National Twin Registry in the United States

The creation of a national US twin registry raises multiple ethical issues. Some of these are common across all research programs involving human subjects, while others are unique to registries or unique to twin research. Common issues include obtaining informed consent, achieving equitable enrolment, and maintaining the privacy and confidentiality of study participants. Registry-related issues include having undefined research goals that complicate the informed consent process, developing policy for sharing data and samples, and providing qualified researchers with access to the registrants for specific research studies. Twin-related issues include strategies for ascertaining intact twin pairs, undue influence for co-twin participation, and the *de facto* disclosure of genetic or disease risk information by a MZ twin.

To explore these and other ELSI and to determine how they might affect establishing a national twin registry in the United States, we assembled an ELSI Expert Advisory Panel (Table 2). The panel and research team met for a day-long workshop to outline relevant ELSI and make recommendations with regard to the best interdisciplinary approaches to address each issue. Also discussed was the need to assemble both ethical and scientific oversight committees with the expertise and flexibility required to

address anticipated and unanticipated ELSI issues, and a participant advocate program to provide a means for twin feedback. The compositions of these committees, reporting structure, and potential areas requiring oversight, for example, research proposals, data monitoring, identity disclosure, privacy and confidentiality, and others, were also addressed. It is clear that regulatory guidelines, policies, and oversight committees must be established early in the twin registry process (but be flexible) in order to maximize the scientific benefit while minimizing harm to the twin participants.

Object 4: Assess Methods for Twin Ascertainment, Enrolment and Retention

Twin ascertainment in the United States poses a difficult challenge due to the lack of an accessible centralized database with personal information through which twins can be identified as such. There are two organizations in the United States that collect and centralize personal information on individuals at the national level. The Social Security Administration (SSA) tracks almost all US citizens for federal benefits and is the closest the United States has to a population register. A previous study showed that it would be possible to identify twins from the SSA data using probabilistic linkage analysis (Kestenbaum, 2004), but the SSA restricts the release of personal information on living persons. The Census Bureau (CB) also collects data on US citizens at the national level through the administration of surveys focused on their social and economic status. These include the decennial census which covers 98% to 99% of the national population and the American Community Survey which covers one in every 40 US households. The addition of questions regarding plurality to either survey and for permission to release personal data to a national twin registry would permit the rapid identification of a high proportion of the twins living in the United States. The census option, however, would require supplementation with other approaches to refresh the cohort of twins moving forward using one or more of the methods discussed below. Piggybacking on a census survey, such as the decennial Census, would be an ideal method for identifying twins in the United States because of its coverage, and also because other data collected as part of the survey could be downloaded into the registry's database if the twins were willing to share these results. However, the next decennial census will be conducted in 2010 and the short amount of time left until then might be a precluding factor. The likelihood of obtaining the necessary approvals to add twin-related questions is also unclear.

Other population-based methods for ascertaining twins include mining data from various public and private databases, electronic medical records, or conducting a random digit dial survey. Each method has its own strengths and weaknesses. The use of electronic medical records, for example, would allow us to ascertain twins and collect data on their health sta-

tuses at the same time. Twins with rare conditions, as well as twins with common conditions, would be identified in the process. The drawbacks of this approach are that electronic medical records are available for only a small fraction of the US population and the systems are not centralized. A random digit dial survey is also considered to be problematic, due to the large number of calls that would be required and a potential bias towards households with land line telephones.

Our review of existing US twin registries found that they use a variety of methods for twin ascertainment that involve either self identification or mining databases such as the states vital records (birth certificates), Departments of Motor Vehicle (DMV) records, Medicare, hospital, military, school records and others. Commercial companies are sometimes employed for tracing twins identified from birth records data, and these can be quite effective especially if the twins' social security numbers are known. Quantitative and qualitative information about these methods and databases were collected from the existing registries and will be reported in a future publication.

The data that have been collected from existing US twin registries and from large federal data collection programs on tracing and enrolment rates are currently being analyzed. However, the success in tracing will depend on the age of the twin at identification and the availability of social security numbers (unique SSA identity numbers). Enrolment rates will be affected by heightened concerns over privacy issues, minority-associated recruitment barriers, and a conservative attitude of the population towards research (issues examined by our ELSI expert panel). Enrolment rates will also depend on respondent burden, the nature of the information collected, our ability to convey the salience of the subject matter to the respondents, and the use of monetary and nonmonetary incentives. Available data indicate that a comprehensive, multifaceted approach would be required for all phases of assembling the twin cohort (ascertainment, tracing and enrolment) in order to achieve a high response rate and population-based sample. Subsequently, good follow-up and a return on the twins' investment in the form of research results and other incentives will be important for minimizing attrition.

Objective 5: Assess Methods for Enrolment of Minorities and Other Underrepresented Subgroups

The United States has long been a 'melting pot' of different races, ethnicities, and cultures. Achieving diversity and minority representation will be critical in this effort, especially if the differential impact of some diseases on various groups is to be addressed. In addition, ethical 'justice' demands that both the benefits and burdens of scientific research be distributed equally among all groups including children, women, all races and ethnicities, and individuals of various socioeconomic conditions, or from different geograph-

ical areas (rural, inner city). In the United States, disparity in research participation is a problem as attested by the literature, and confirmed by a review of the data collected from existing twin registries. Some of this might be due to historical abuses such as the Tuskegee Syphilis Study (Jones, 1993), the Nazi eugenic experiments (Annas & Grodin, 1992), and others. More recent data, however, suggest that the under representation of some groups might be due to little effort on the part of the investigators to recruit them (Wendler et al., 2005).

Therefore, an important aspect of this study was to identify the barriers as well as the incentives for reaching underrepresented research groups. Both were incorporated in our registry surveys and discussed at our ELSI Workshop. Focus group discussions will also be held with representative target populations including mothers of twins, African Americans, Spanish speaking twins, rural twins, and others. Some proactive strategies for achieving diversity might include tailored approaches for disseminating registry information, involvement at the community level, acknowledging past research abuses, and using methods that reflect the groups' culture and languages and make it convenient for them to participate. Our goal of creating a national population based twin registry can be realized only if all groups are represented in the program.

Objective 6: Procedures for Biospecimen Collection

Due to the explosion of genetic research, the collection of biological specimens for future genetic analysis is fast becoming a common component of most cohort research studies. We would be remiss if the opportunity to collect archival tissues from twins upon enrolment into a national twin registry was not taken. The biological samples obtained from twin participants could be used for zygosity testing (depending on the availability of financial resources) and banked for future genetic analyses. As useful as these samples are likely to be, it is clear, however, that the donation of biological specimens for these purposes should not be conditional for enrolment; optional tissue donation will likely increase twin enrolment. As with any tissue repository effort, consent, confidentiality and coding will be key issues that require forethought. Strict policies will also be required for sharing the samples with researchers, as well as for decoding the specimens to ascertain participants for follow-up studies, based on either their phenotype or their genotype. Unique to a twin registry repository, however, policies will be required for the 'de facto' genetic disclosure this creates for MZ twin pairs.

Objective 7: Propose and Evaluate Alternatives to a National Twin Registry

Creating and maintaining a national twin registry in the United States will be no small feat, and might be found to be infeasible as a result of this study. With that in mind, alternatives for improving twin research

resources in place of creating a national registry were examined. For each alternative explored, the technical feasibility of the design was examined, as well as the logistics and costs, and also whether the twin resource was adequate for supporting valid scientific research (population basis, representativeness, and statistical power). Also examined was the possibility of conducting a pilot twin program and the optimal means for doing so.

Various options for ascertaining twins in the United States were discussed above (mining data from birth certificates, department of motor vehicle records and other databases, using electronic medical records, existing national surveys, and random digit dial telephone surveys). One alternative examined was the creation of a smaller twin registry using these same ascertainment methods in combination with a statistical sampling scheme to target a defined probability sample (subset) of US twins. The registry would consist of fewer twins, but would be less costly and involve securing the cooperation of fewer institutions (states or other localities, hospitals, and so on). Another alternative is the possible expansion of an existing twin registry such as the MATR or the creation of a consortium of multiple existing registries. Although this would require obtaining approval from the existing registries for access to their twins, it would reduce the effort needed for ascertaining and tracing potential participants, and collecting baseline data. Use of an additional approach, such as sampling, might be used to supplement the consortium for those twins who are not demographically represented in the existing cohorts.

Objective 8: Assessment of Registry Costs and Infrastructure

The cost of creating a national twin registry or pursuing one of its alternatives for improving twin research resources is an important aspect of this feasibility study. For analysis, cost was broken down into six categories: the costs for case ascertainment, tracing, baseline enrolment, follow-up data collections, and infrastructure and management. Infrastructure costs include those incurred in developing registry policy and governance and establishing steering and oversight committees. An informatics center will be required to collate, manage and distribute core data. Additional costs might be needed for biospecimen collection and tissue archiving.

Considerations regarding these aspects of registry costs encompass both fixed and variable costs, as well as labor costs versus other direct costs that might be associated with the registry. The financial costs for establishing a national twin registry could be great, but some of these can be offset by access fees to investigators.

Discussion

The major goals of the feasibility study outlined above are to estimate the potential size and characteristics of

a national twin registry based on the current US twin population, determine our ability to reach them, and assess their willingness to participate. The ethical and social climate and diverse nature of the US population are critical elements in deriving the estimate. The logistics, infrastructure, and costs needed to pursue US twins at the national level must be factored in as well. The most important consideration, however, is the worthiness of the effort. Before we proceed to create a national twin registry in the United States, these questions must be answered: Will the expected registry population be adequate to support the conduct of valid scientific research? Are there research questions that can best be answered using twin designs? To answer the first question, we must consider the population basis of the registry and whether a registry based on the US population can realistically be achieved. If the registry population is not representative of the US population, generalizing research results could be problematic. We must also consider whether the potential size of the registry population would provide adequate statistical power for conducting various types of studies.

A population-based registry has a defined source population and for our purposes here, the source population is broad in that it includes the overall US population of varying ages, gender, races and ethnicities, income levels and other characteristics. While the US twin population (based on data collected in this study) is representative of the overall US population, deficiencies were identified in the US twin cohorts previously assembled. The deficiencies highlight potential problems that could be encountered in a national registry that will have to be addressed with the appropriate ascertainment and enrolment methods. We expect that a combination of ascertainment methods will be required that include special strategies to target underrepresented groups. An ideal methodology, for example, might encompass the identification of living twins using a large cross-sectional survey with high response rates (US decennial Census or other Census Bureau survey) followed by annual supplementation of the sample with newborn twin pairs identified through selected state vital statistics departments. Tailored community awareness programs, media advertisements or other means could be incorporated to target underrepresented populations. With this ideal strategy and allowing for the response rates of other registries and national surveys, we estimate that 3.2 to 3.4 million twins (60% to 65% of 5.3 million twins) could be enrolled into a national registry if all states participate. Based on demographic data collected as part of this study, this represents about 1.2 million intact twin pairs, with about one third of these being MZ and the other two thirds being DZ. We therefore conclude that a population-based twin cohort is attainable and the research results could be generalized to the US population.

In order for the proposed registry to be scientifically useful, enough twin pairs must be enrolled to ensure adequate statistical power for a variety of research questions and study designs. A hypothetical classical twin study is presented to estimate the heritable contribution of a relatively rare condition (prevalence is 0.002). For this study, let us assume that the concordance rates for the condition are 0.2 and 0.05 for MZ and DZ twins, respectively. Using a one-sided analysis and significance level of .05, we estimate that 72 twin pairs in which one or both co-twins has the trait would be needed for 80% power and 36,000 twin pairs would need to be screened to identify the 72 pairs. In a far more complex multivariate model presented by Martin et al. (1997) involving a pleiotropic QTL, and a phenotype with three correlated continuous traits, between 1000 and 3000 DZ pairs would be needed for 80% power ($\alpha = .001$) depending on the source of the phenotypic variation (QTL, genetic and/or environmental background). Martin showed that by incorporating MZ twins with the same phenotype(s) into the model, power would be substantially increased (or the sample size could be substantially reduced). Regardless of the purpose of the example, both hypothetical studies show that various twin study designs would be sufficiently powered using the US national twin registry twin cohort described above, from a simple classical twin design to a highly complex multivariate genetic model with multiple phenotypes. In fact, the more complex model would not be possible without a large population-based resource such as the one being proposed here.

The above examples also provide an answer to our second question: Are there research questions that can best be answered with twin studies? The classical twin design is considered to be the definitive method for estimating the magnitude of the genetic and environmental contributions to observed variation in complex traits. For more complex studies, the use of multivariate twin models increases the likelihood of detecting QTL, by reducing environmental or genetic noise, and for estimating linkage as a subfraction of the total heritability in comorbidity studies. Twins allow scientists to better address ascertainment bias, the variable onsets of disease, and the interaction of genes with the environment (genetic sensitivity in MZ twins or covariate designs). In parallel with advances being made in analytical methods and genomics, novel twin strategies are continuously emerging. In fact, this is a major goal of the GenomEUtwins project, to develop a framework for facilitating the development and use of novel strategies to capitalize on the unique features of twins (Peltonen, 2003).

Will creating a national US twin registry be feasible? The study described here is not complete, so that question remains yet unanswered. Compared with other countries, the United States will have unique and difficult challenges to overcome, but based on our initial data, we do know that the effort is highly

worthwhile. Well-characterized, large population-based twin cohorts such as GenomEUtwins and the national US population-based twin registry proposed here will only facilitate the development of new strategies using twin designs to detect the genetic and environmental causes for complex disease, and the complex interplay between multiple factors. Although a national US twin registry is quite ambitious, the investment is minor when compared with the significance of the potential research findings and impact these will have on human health.

References

- Anderson, L. S., Beverly, W. T., Corey, L. A., & Murrelle, L. (2002). The Mid-Atlantic Twin Registry. *Twin Research*, 5, 449–455.
- Annas, G. J., & Grodin, M. A. (1992). *The Nazi doctors and the Nuremberg code: Human rights in human experimentation*. New York: Oxford University Press.
- Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics*, 3, 872–882.
- Jones, J. H. (1993). *Bad blood: The Tuskegee syphilis experiment* (new and expanded ed.). New York: Free Press.
- Kestenbaum, B. (2004). Identifying a cohort of US twins using social security administration records. *Twin Research*, 7, 134–138.
- Keyfiz, N., & Caswell, H. (2005). The life table. In *Applied mathematical demography* (3rd ed., pp. 29–46). New York: Springer-Verlag.
- MacGregor, A. J., Snieder, H., Schork, N. J., & Spector, T. D. (2000). Twins. Novel uses to study complex traits and genetic disease. *Trends in Genetics*, 16, 131–134.
- Martin, N., Boomsma, D., & Machin, G. (1997). *Nature Genetics*, 17, 387–392.
- Peltonen, L. (2003). GenomEUtwin: A strategy to identify genetic influences on health and disease. *Twin Research*, 5, 354–360.
- Wendler, D., Kington, R., Madans, J., Van Wye, G., Christ-Schmidt, H., Pratt, L., Brawley, O., Gross, C., & Emanuel, E. (2005). Are racial and ethnic minorities less willing to participate in health research? *PLoS Medicine*, 3, e19 (www.plosmedicine.org).

