



Published in final edited form as:

Clin Transl Sci. 2010 February ; 3(1): 42–48. doi:10.1111/j.1752-8062.2010.00175.x.

Principles of human subjects protections applied in an opt-out, de-identified biobank

Jill Pulley, MBA,

Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Ellen Clayton, MD, JD,

Center for Biomedical Ethics and Society, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Gordon R. Bernard, MD,

Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Dan M. Roden, MD, and

Office of Personalized Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Daniel R. Masys, MD

Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Introduction

There are a growing number of DNA biobanks across the country and around the world. These generally rely on sample collection and information capture protocols that fall within the scope of human subjects research as defined by U.S. regulations and by the local ethics committee or Institutional Review Boards (IRB). BioVU, the Vanderbilt DNA Databank, represents one of few biobanks with an operational protocol that makes it non-human subjects research as determined by the local IRB and the federal Office of Human Research Protections (OHRP). The Code of Federal Regulations, 45 CFR 46.102 (f), defines a “Human Subject” as a living individual about whom an Investigator conducting research obtains either data through intervention or interaction with the individual, or identifiable private information, even in the absence of intervention or interaction, or an individual who is or becomes a participant in research. BioVU accrues DNA samples extracted from blood remaining from routine clinical testing after samples have been retained for 3 days and are scheduled to be discarded. Thus there is no intervention or interaction with individuals for research purposes. The resource is linked to a de-identified version of data extracted from an Electronic Medical Record (EMR) system, called the Synthetic Derivative (SD), in which all personal identifiers have been removed. Thus, there is no identifiable private information attached to the records. A full description of this resource has been published elsewhere¹.

Corresponding Author: Jill Pulley, MBA, Asst. Professor, Medical Education and Administration, Director, Research Support Services, Vanderbilt University Medical Center, 2525 West End Avenue 6th Floor, Nashville, TN 37203, 615-343-2842 (phone), Jill.pulley@vanderbilt.edu.

The authors report no conflicts of interest related to the content of this paper.

Academic medical institutions have long performed research that involves data and/or tissue derived from humans, but is not human subjects research as defined by the regulations. We and others have observed that some members of the lay public are unaware of the federal definitions and believe that existing tissues (for example, stored surgical pathology specimens) are not (and should not) be used for research without explicit consent². This belief contradicts the current actual practice of research permitted by the Code of Federal Regulations Title 45, Part 46, (the 'Common Rule') Subpart A 46.101(b)(4), which provides an exemption from the regulations for human subjects research studies that involve '...the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.' Although not a use of 'existing tissues' the BioVU model relies on subsequent OHRP guidance on "Research Involving Coded Private Information or Biological Specimens"³, that states: "This guidance applies to existing private information and specimens, as well as to private information and specimens to be collected in the future for purposes other than the currently proposed research. The following are examples of private information or specimens that will be collected in the future for purposes other than the currently proposed research: (1) medical records; and (2) ongoing collection of specimens for a tissue repository." Despite the fact that BioVU falls outside the coverage of the Common Rule, Vanderbilt has chosen to publicize its use of residual samples and to permit patients to opt out of such use, a practice for which there was effectively no precedent.

The Belmont Report, published in 1979, is a foundational document for modern research ethics. It enumerates the importance of the boundary between practice and research, and three principles: Respect for Persons, Beneficence, and Justice, which constitute the essential ethical framework by which IRBs and ethics committees judge the risks and benefits of research involving human subjects. In the Belmont Report, Respect for Persons 'incorporates at least two ethical convictions: first, that individuals should be treated as autonomous agents, and second, that persons with diminished autonomy are entitled to protection.' Beneficence is embodied in the rules '(1) do not harm and (2) maximize possible benefits and minimize possible harms.'. Justice is the principle of the fairness of the distribution of the benefits of research, relative to those who bear its burdens'.

BioVU was developed by designing and implementing new procedures and components that are consistent with the principles of the Belmont Report but for which there were no established methods. These included special oversight and governance, new informatics technologies, provisions to accommodate patients' preferences not to be included in the research program, as well as an extensive public education and communications component. We expect that the BioVU model would not be generalizable or transportable to many sites and situations. Further, other models might be more appropriate in different settings or among different population groups. The intent of this paper is not to make comparisons. Rather, consideration of core principles and protections in the practical implementation of BioVU is the focus of this paper.

Methods

In the formation of BioVU, the program designers received input from the Medical Center Ethics Committee, the Medical Records Committee, an institutional operations oversight committee established specifically for BioVU, and a Community Advisory Board also established specifically for the project, as well as initial review and on-going oversight of the resource and its use by the IRB. The principles and their application described here were discussed at length by all groups.

Boundaries between Practice and Research

One foundational principle of the Belmont report is preserving the boundary between clinical care and research. This distinction is particularly relevant for BioVU, given that notification for the program is provided in patient registration forms that are part of the healthcare practice environment. A common, central registration form was used so as to not disrupt logistics or other patient flow procedures by introducing a separate research informational form. Text in this Consent to Treatment/ Agreement to Pay form - that every patient at Vanderbilt University Medical Center (VUMC) signs annually - was altered and expanded to explain that research on tissues might be conducted. Patients are explicitly notified of the DNA Databank in these forms under the “Research” sub-paragraph:

I understand and agree that any specimens or tissues normally removed from my body by VUMC in the course of any diagnostic procedures, surgery, or medical treatment that would otherwise be disposed of may be retained, used for educational purposes or research, including research on the genetic material (DNA) or other information contained in those tissues or specimens.

The text serves to distinguish health care from research related practices and functions that may involve data or tissues. The intention of preserving this boundary between practice and research was also supported by the systematic removal of all identifying information from the records used in BioVU for research. Both project staff and ethical oversight committees understood that the information contained in the Consent to Treatment form is not always read in its entirety by patients, and sometimes not read at all. For this reason, the text present in this document is supplemented by other methods of communication.

Respect for Persons—The principle of Respect for Persons was an essential component of the design for BioVU. While the most common application of this principle (for example, a separate research informed consent document and consent process) is not employed in this program, respect for persons, including autonomy, and voluntariness were clear objectives in implementing the provisions listed below:

1. Despite the IRB’s determination (and OHRP’s confirmation) of the project did not involve human subjects as defined by federal regulation such that efforts either to inform about the project or to seek their permission were legally required, there was consensus by the IRB and BioVU planners that the individuals from whom DNA and records could be derived have the right to be notified of the program. As implemented, this notification is provided primarily via registration forms wherein documentation of patient notification is obtained via signature. These notifications are provided at least annually. The forms are provided both in English and Spanish, languages spoken by an overwhelming majority of the patients in the Vanderbilt catchment area.
2. Patients have been given the right not to participate. Specifically, individuals can opt out of having their DNA included in the biobank at any time by checking a box placed conspicuously on the Consent to Treatment form described above (see image 1, opt out). As shown, the box to opt out is directly above the signature line. If an individual chooses to opt-out, the opted-out status is permanent, and is kept confidential as part of the demographic data stored in the operational EMR. Individuals can opt out at any time, preventing future usage of their tissue and tissue-derived data.
3. To enhance the number of individuals who are aware of the BioVU program, educational materials have been implemented through various print materials and distribution channels. These include explanatory full color brochures (see Image 2), posters, and newspaper advertising and articles. One purpose of the newspaper

advertising is to provide some opportunity for individuals to be exposed to the program prior to coming in for their visit. The brochure is specifically intended to represent key information components in an understandable format, including: a) program description and rationale, b) operational protocols for obtaining and handling DNA, c) restrictions on usage, d) privacy, e) the right to choose not to be included, f) research-clinical boundary, g) oversight, and h) handling common misconceptions. The program brochure has written readability of grade level 3.9 according to the Flesch-Kincaid Grade system level, and is distributed along with the treatment consent form annually as well as made available on counters in waiting areas in outpatient clinics. A toll free number, which is also included on the brochure, can be called if a patient has additional questions or prefers to speak with program staff personally. A printed community newsletter is also planned. An empiric observation is that many patients read the brochure while waiting for outpatient appointments.

In addition to these communications efforts, an ongoing series of research efforts involving assessment of patient attitudes inform program development and operations. These have included patient surveys, exit interviews, and focus groups. There have been 16 separate IRB approved protocols conducted to guide the development, design, and utility of BioVU. Data collected from these efforts have been published⁴ and are reviewed by the advisory bodies routinely for incorporation into project planning. Significantly, we have consistently found that when asked explicitly, an overwhelming majority of patients choose to have their records and samples included in the program. Similar acceptance of widespread favorability for DNA biobanking has been reported by others^{5,6} as well as support for opt out models⁷ (personal communication, D Kaufman, manuscript in preparation).

Beneficence—The principle of beneficence has been honored by minimizing the risk of harm to individuals while maximizing benefit to the broader society. One form of potential harm is the inadvertent or intentional re-identification of the identity of individuals whose samples and/or data are contained in the research resource. BioVU includes the linkage of DNA samples to a de-identified version of the EMR called the Synthetic Derivative (SD). To minimize the risk of harm, the SD provides a de-identified copy of data from Vanderbilt Medical Center's inpatient and outpatient EMR systems, reformatted into a separate database structure for secondary research purposes. The current SD user interface contains a search mechanism that allows researchers to extract sets of records of interest based on various structured data fields (e.g. reimbursement codes and laboratory values) and unstructured data fields (e.g. the words contained in clinical narratives and discharge summaries, in a fashion similar to Internet search engines such as Google). The current SD search relies on a simple, "cross-sectional" search through the chronologically ordered clinical data contained within the SD data repository. The retrieval and review functions provide real-time results for simple and complex queries. To maximize benefit (i.e., its utility for a broad variety of research) the data resource preserves the sequence of clinical events but employs date shifting, so that dates contained in the data are never the actual dates of a clinical event, but the chronological order and date intervals are retained. Algorithms shift the dates by a time period that is consistent within each record, but differs across records (for example, if the date in a particular record is April 1, 2005 and the randomly generated shift is 45 days in the past, then the date in the SD is February 15, 2005, while another individual's record would be shifted systematically by a different number of days). Importantly, as new entries are accumulated in the EMR system, the scrubbing process described below is applied to new additions and the SD is thereby updated.

The SD data resource facilitates research with and without the associated biobank samples. Researchers contemplating the feasibility of a clinical trial can easily determine the number

of cases of a particular disorder that exist, what patterns of therapy have been, and the occurrence of intended and unintended effects of treatment. As of this writing, more than 315 searches of the resource by 32 users have occurred. The creation and maintenance of the associated DNA biobank is supported as an institutional research resource that is availability to all faculty who have IRB-reviewed and approved studies (see below). In addition, the terms of use of the resource require of redeposit of genetic information generated in a research study back into a genomic database; this will also help to maximize benefit. All protocols are reviewed by a committee of investigators and other experts to ensure the scientific quality of the research projects undertake, with the goal to promote the resource in terms of improvements to generalizable knowledge, and to facilitate these enhancements to knowledge in an efficient way, with samples and data that have already been collected.

To date the resource has supported genome-wide association studies of approximately 15 different disease conditions and physiologic traits such as variability of cardiac conduction^{8,9}. Other work underway is identifying adverse drug effects as documented in the de-identified clinical data, and correlating susceptibility to those adverse effects with genome scans. The long term benefit of these efforts will be improvement in healthcare services, expansion of the knowledge base of clinically important human genetic variation for a coming era of ‘personalized medicine’, and reduction of avoidable suffering caused by adverse effects of therapy across a broad range of human diseases. In this context benefits of the resource will accrue for society, though not directly or immediately to the individuals from whom the materials and information are derived.

Several approaches are employed to further minimize harm to individuals:

1. There is extensive oversight for the program. Although not generally mandated for studies that do not involve human subjects as defined by the Common Rule, the Vanderbilt IRB determined that “given the seminal nature of the proposed work and its potential broad impact upon the local community the Committee agreed that additional safeguards are warranted...” and, that “The IRB will ...maintain oversight for the specimen repository.” This includes oversight of all changes to the protocol, including the Data Use Agreement signed by users of BioVU, communication materials, exclusions and other procedural elements. The IRB, with its institutional jurisdiction to review and approve research, will also review and approve protocols seeking to access BioVU; verification of IRB approval will be made by program staff before access will be granted to an approved investigator. The BioVU project itself is on an annual continuing review cycle. In addition to the IRB, the following other groups, as shown in Figure 1, provide guidance to BioVU: a Community Advisory Board, an Operations Oversight Board (OOB), an external Ethics Advisory Board, and the Medical Center Ethics Committee. The OOB reviews all studies seeking access to the resource in addition to the review conducted by the IRB. An external Scientific Advisory Board will convene in the upcoming year.
2. A key characteristic of BioVU (and essential to the IRB’s approval of it) is the effective de-identification of the medical data. The Synthetic Derivative is designed to be structurally compliant with the ‘Limited Data Set’ provisions of the HIPAA Privacy Rule (Sec. 164.514) by removal or permutation of 17 data elements and any other unique identifying number, characteristic or code, specified as person-identifiable by the Rule. Because the resource contains the full text of clinical documents, there is a possibility of ‘inferential re-identification’, particularly of individuals who are celebrities or for whom there is a widely-publicized set of characteristics that might be recorded in the EMR (e.g., “This 82 year old retired astronaut...”) The IRB, which also serves as Vanderbilt’s Privacy Board, reviewed

the effectiveness of the computer system that removes identifiers and determined that the de-identification process is sufficient to ensure that the resource does not involve “human subjects” as defined by the regulations.. In addition, users must submit a project proposal that undergoes IRB review, and must sign a Data Use Agreement, which requires that “in the event the Data Recipient becomes aware of any personally identifiable health information unintentionally missed by the de-identification process, such that the personally identifiable health information is retained in the Data Set received by Data Recipient, [the user agrees] to report all such occurrences only to BioVU program staff and/or the privacy office for quality assurance purposes, and not to any third party.” A simple online mechanism has been developed to fulfill this reporting requirement. However, we are well aware that the removal of identifiers does not completely remove re-identification potential. Quantitative risk assessment models are under development to provide a statistical basis for likelihood of re-identification potential when data are being sent to public databases such as dbGaP^{10,11}.

3. To support security equal to or greater than that which is required by federal regulations and the OHRP guidance Research Involving Coded Private Information or Biological Specimens, the samples are coded by a Research Unique Identifier, (RUI)]¹. The RUI is a 128-character code derived from the Medical Record Number (MRN) generated by the Secure Hash Algorithm (SHA-512, National Security Administration). SHA-512 produces a string of characters (the Research Unique Identifier, RUI), that is unique to a particular input and has the property that is not possible to infer or compute the MRN that generated it. No link between the Research Unique Identifier and the Medical Record Number is maintained, unlike “Trusted Broker” systems, such as that served by a formal data management center which maintains links to data but prohibit the release of the key to the investigators, as permissible by OHRP guidelines¹².
4. A percentage of samples are randomly excluded and discarded. The rationale for random exclusion is to make it impossible to infer which individuals’ DNA samples might be in the bank by knowing who has opted out. Because of this exclusion, it will never be known with certainty that any particular individual has a sample included in the Databank.
5. Upon IRB acceptance, users may interact with the database query system, to select records of interest. To reduce residual re-identification potential, the system returns a clinical dataset to an approved investigator only if the numbers of matching records exceed a specified minimum set. This is normally 5, consistent with CMMS health data guidelines [ref], but number returned is parameter driven and can be modified on the basis of quantitative real-time assessments of risk of re-identification. All search outputs are tracked and audited. The methodology for audit is currently to match search output (which is stored in a system log file) to the stated phenotype description approved by the IRB. A random selection of investigator-initiated searches is validated by audit trail analysis. Search criteria that do not relate to the study description are flagged by staff for follow up with the PI. At a minimum, one search for each study and one search per investigator are audited. User access accounts are project-specific, rather than person-specific, enabling a basis for comparison of the search statements and records returned with the project approval granted by the IRB.
6. A Data Use Agreement is signed by all users which details the accepted and prohibited uses of the data. For example, users agree “not to identify the information contained in the Limited Data Set by any means, including using Star Panel [the Operational EMR system] or other information together with the Data

Set, or to contact any individual whose information is contained in the Limited Data Set". Institutional penalties similar to those employed for intentional breeches of confidentiality of clinical data will be enforced and users are made aware of these provisions, including the Sanctions for Privacy and Information Security Violations.

Justice—Another key principle considered was justice, which addresses who receives the benefits of research and bear its burdens, in the sense of "fairness in distribution" or "what is deserved". In this regard, it is known that a traditional consent or opt in model may exclude large segments of the population^{13,14,15,16}. This is unfortunate from both an ethical and scientific standpoint. In part for this reason, some have advocated an opt-out approach for many "low-risk" studies.¹⁷ With BioVU, sample acquisition is inherently unbiased, other than by the availability of a leftover blood sample after clinically-indicated testing. All patients age ≥ 18 years with an outpatient laboratory draw and who have a signed consent to treatment form and who have not made a formal indication to opt-out are potential inclusions in BioVU. Specifically, all samples from individuals who have diagnostic testing requiring a blood draw have an equal likelihood of being included. There are no health or demographic related exclusions. In addition, the large sample size enabled by this approach allows new types of research previously not feasible, including those on less common diseases and conditions or those focusing on minority demographic populations. The implication of this BioVU feature from the perspective of justice is that the diseases are amenable to research in direct proportion to the proportion of individuals who are receiving healthcare for those diseases and conditions.

Results

As of December 7, 2009, there are over 73,000 samples collected to date, with an average DNA yield of ~ 120 ugs per sample. The cost per sample including all processing, collection, storage, labor, and program management is less than \$25/sample. At this pace of accrual, and assuming pediatric sample collection is initiated in early 2010, the resource will reach 200,000 samples by 2013 (see Figure 2). There have been 34,444 individuals who have opted out, representing approximately 5% of patients seen since BioVU began. These numbers are consistent with expectations from prior patient attitude surveys². Complaints have been relatively few, and these issues have been addressed via an 800 number hotline, which is staffed by patient services representatives with training in the design and operation of the biobank. Callers may also speak with senior institutional representatives associated with BioVU. This ability to discuss concerns with patients who call with complaints was sufficient; further follow up was not required in any case to date. The biobank is currently comprised of a sample that is representative of our overall patient population (see table 1 for demographic composition of BioVU records). Many relatively difficult to research demographic populations are represented, for example, samples from 542 individuals over the age of 89 are included in the repository. The average record size for BioVU records is 145kb, which is approximately 12 pages of text. There are 14,537 unique ICD-9 codes represented in the database, and an average of 25 medications in each record. Select rare disease counts are provided in Table 2 for illustrative purposes. Over 250,000 brochures have been distributed to date. In the creation of the Synthetic Derivative, 1.7 million patient records have been de-identified and as discussed above, 72,000 of these have associated DNA samples to date.

Discussion

The BioVU model is still a work in progress. Some advantages, however, are already apparent. First, it is capable of generating a larger number of biobank samples at

considerably lower cost than traditional prospectively consented cohort models. This allows cost savings to be passed on to investigators and to funding agencies. As noted above, the scale of the resource will also allow research that could not otherwise be conducted on rare diseases and some minority demographic populations. The resource uses byproducts of routine care -- blood that would otherwise be *discarded* -- to obtain genetic samples. By also using the information byproducts of the healthcare system, the SD is cost effective to generate and contains information that is rich in phenotypic attributes. The chronology of events and longitudinal nature of the data are preserved by the methodologies employed.

By design, one corollary benefit of de-identifying patient records for the purpose of creating a biobank that does not include human subjects as defined by the Common Rule is that the frequency with which identified individual patient charts are viewed by researchers under a more traditional IRB Waiver of Consent is substantially reduced, therefore increasing privacy protections. The existence of such a resource may reduce the frequency with which the IRB will be asked to grant exempt status for retrospective medical record review research using identified patient charts. The existence of the resource provides the IRB with an alternative data resource to direct researchers to whenever identifiable data is not necessary for a proposed research protocol.

Disadvantages of the model include the implementation time required to assess the acceptability and feasibility of such a model and obtain the ethical and regulatory approvals. In addition, the design explicitly precludes re-contact with any individual, and no ability to obtain any information other than that contained within EMR. For example, environmental exposures of many types are not represented. Also, family structures are generally poorly represented in the EMR and so are not available for this work. Paradoxically, an unanticipated benefit of the program has been to highlight such deficiencies, and to put in place mechanisms to improve the information content of the clinical EMR. Many individuals receive care outside of the local institution's health system, and this additional clinical information is not available for research purposes. Further, because dates have been shifted, it is impossible to conduct research that relies on accurate relationships to known dates (e.g., affective disorders associated with a natural disaster such as a hurricane or a seasonal event such as a holiday).

Conclusion

There are numerous DNA biobanks across the country and internationally that rely on a traditional human subjects approach to sample collection and information capture procedures. BioVU represents one of few biobanks with an operational protocol that does not involve human subjects as defined by regulation. BioVU accrues DNA samples extracted from blood remaining from routine clinical testing. By design, BioVU is based upon the ethical principles described in the Belmont Report, but not the traditional *application* of those principles. BioVU is a resource that reflects the diversity of the population, thereby promoting a just distribution of the benefits of research while minimizing the risks to individuals. This paper is intended to inform other institutions considering initiating a biobank regarding procedures that can be employed to help protect the rights and welfare of patients, but preserve the ability of such a resource to efficiently propel genomic research. Some authors have asserted that people receiving the benefits of healthcare have an obligation to participate in research¹⁸. Consistent with this, studies have shown that 'altruism' and a desire to play a part in scientific advancement is often a motivational factor for participating in research^{19,20}. We suggest that if individuals have such an obligation to participate in research, then donating tissues and information to a de-identified data and tissue resource such as BioVU at the exact time that the individual is receiving benefits of healthcare -- satisfies this obligation in a societal sense.

Acknowledgments

Funding:

This work was supported in part by Vanderbilt CTSA grant 1 UL1 RR024975 from NCRR/NIH, and in part by Vanderbilt eMERGE network grant 5U01HG004603-02 from the National Human Genome Research Institute

The authors would like to acknowledge the input of the Vanderbilt University Institutional Review Board, the Medical Center Ethics Committee, the External Ethics Advisory Committee, the

References

1. Roden DM, Pulley JM, Basford MA, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther.* 2008; 84:362–369. [PubMed: 18500243]
2. Pulley JM, Brace MM, Bernard GR, Masys DR. Attitudes and perceptions of patients towards methods of establishing a DNA biobank. *Cell Tissue Bank.* 2007
3. <http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.pdf>
4. Pulley JM, Brace M, Bernard GR, Masys D. Evaluation of the effectiveness of posters to provide information to patients about a DNA database and their opportunity to opt out. *Cell Tissue Bank.* 2007; 8:233–241. [PubMed: 17273897]
5. Kaufman D, Murphy J, Erby L, Hudson K, Scott J. Veterans' attitudes regarding a database for genomic research. *Genet Med.* 2009 May; 11(5):329–37. [PubMed: 19346960]
6. Chen DT, Rosenstein DL, Muthappan P, Hilsenbeck SG, Miller FG, Emanuel EJ, Wendler D. Research with stored biological samples: what do research participants want? *Arch Intern Med.* 2005 Mar 28; 165(6):652–5. [PubMed: 15795341]
7. Wendler D, Emanuel E. The debate over research on stored biological samples: what do sources think? *Arch Intern Med.* 2002 Jul 8; 162(13):1457–62. [PubMed: 12090881]
8. Ritchie, MD.; Denny, JC.; Crawford, DC.; Havens, AK.; Weiner, JB.; Pulley, JM.; Basford, MA.; Brown-Gentry, K.; Balser, JR.; Masys, DR.; Haines, JL.; Roden, DR. Robust replication of genotype-phenotype associations across multiple diseases in an Electronic Medical Record. (manuscript in submission)
9. www.gwas.net
10. Malin B, Karp D, Scheuermann R. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. *Journal of Investigative Medicine.* Forthcoming.
11. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. *Journal of the American Medical Informatics Association.* 2010 Forthcoming.
12. <http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.pdf>
13. Huang N, Shih SF, Chang HY, Chou YJ. Record linkage research and informed consent: who consents? *BMC Health Services Research.* 2007; 7:18. [PubMed: 17291357]
14. Buckley B, Murphy AW, Byrne M, Glynn L. Selection bias resulting from the requirement for prior consent in observational research: a community cohort of people with ischaemic heart disease. *Heart.* 2007; 93:1120. [PubMed: 17699175]
15. Hewison J, Haines A, Hewison J, Haines A. Overcoming barriers to recruitment in health research. *BMJ.* 2006 Aug 5; 333(7562):300–2. [PubMed: 16888308] *Br Med J.* 2006; 333:300–302. [PubMed: 16888308]
16. Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. *BMJ.* 2009 Mar 12; 338:b866. Review. 10.1136/bmj.b866 [PubMed: 19282440]
17. Junghans C, Feder G, Hemingway H, Timmis A, Jones M. Recruiting patients to medical research: double blind randomised trial of “opt-in” versus “opt-out” strategies. *Br Med J.* 2005; 331:940. [PubMed: 16157604]
18. Schaefer GO, Emanuel EJ, Wertheimer A. The obligation to participate in biomedical research. *JAMA.* 2009 Jul 1; 302(1):67–72. [PubMed: 19567441]

19. Jenkins MM, Reed-Gross E, Rasmussen SA, Barfield WD, Prue CE, Gallagher ML, Honein MA. Maternal attitudes toward DNA collection for gene-environment studies: A qualitative research study. *American Journal of Medical Genetics Part A*. 2009 Nov; 149A(11):2378–86. [PubMed: 19839045]
20. Morgan AT, Reilly S, Eadie P, Msppath AW, Simpson C. Parental consent for neuroimaging in paediatric research. *Child: Care, Health and Development*. 2009 Jul 23. [Epub ahead of print].

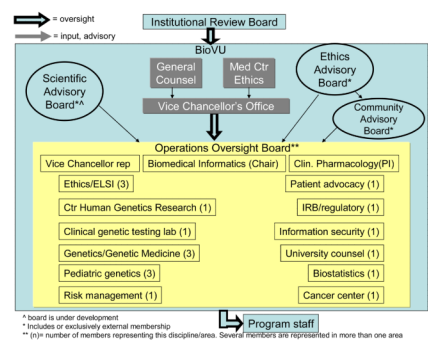


Figure 1.
BioVU oversight structure

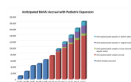


Figure 2.
Accrual Projections

VI. USE, RETENTION AND DISPOSAL OF TISSUE AND BLOOD

I understand and agree that any specimens or tissues normally removed from my body by VUMC in the course of any diagnostic procedures, surgery, or medical treatment that would otherwise be disposed of may be retained, used for educational purposes or research, including research on the genetic material (DNA) or other information contained in those tissues or specimens.

I acknowledge that such research by VUMC may result in new inventions that may have commercial value and I understand that there are no plans to compensate me should this occur, regardless of the value of any such invention.

I understand that any research using these leftover specimens or tissues will be done in a way that will not identify me or my medical information.

I also understand that if I do not want DNA research to be done using my leftover blood, I need to check the box shown below. If you have questions, please call 1-866-436-4710.

☐

Do not use my leftover blood for the DNA Databank

PLEASE READ THIS ENTIRE AUTHORIZATION PRIOR TO SIGNING.

Patient/
Legal Representative _____ Date _____ Time _____ A.M. P.M.
(Relationship to Patient) _____

Image 1.

Opt out portion of Consent to Treatment/Agreement to Pay form

THE VANDERBILT BIOVU QUESTIONS AND ANSWERS

1. What is the Vanderbilt BioVU Program?

We are collecting DNA from adults and children who are patients at Vanderbilt to learn how the differences we inherit can affect health. Learning more about this could help us to improve health care for everyone.

2. What is DNA?

DNA is how people store information we inherit from our parents that directs how we grow and develop. The information is contained in pieces called genes.

3. How does Vanderbilt get DNA from patients?

When blood is drawn from a patient, it is used for the tests that were ordered. DNA can be obtained from the blood that is left over after the tests are done. This blood would normally be thrown away.

4. What will Vanderbilt do with the DNA?

The first thing Vanderbilt will do is to put a code on the DNA and then enter it into a specially protected computer. This computer can then match the DNA with information about the person's health that is contained in the medical records here. Researchers will not be able to tell who the information applies to, but they will be able to study how differences in the genes affect health. Researchers may sometimes send the DNA to another Vanderbilt-approved institution or facility for research related services, such as doing a laboratory test to find particular genes in the DNA. De-identified information about genes or a person's health may also be shared with other researchers.

5. Will anyone be able to tell whose DNA is whose?

NO. No one will know who the DNA belongs to. Vanderbilt researchers, employers, insurance companies and law enforcement agencies will not be able to get information about a specific person's DNA. The purpose of the coding is to make sure that each person's privacy is protected.

6. How does Vanderbilt tell me about this program?

Patients treated at Vanderbilt sign the Vanderbilt Consent for Treatment and Agreement to Pay form. The DNA project is addressed in that form and in brochures and posters located in patient areas. If you have questions, you can call (866) 436-4710.

7. What if I do not want my DNA used this way?

All you have to do is check the box on the form called Consent to Treatment and Agreement to Pay or call (866) 436-4710 to opt out. We would like to collect all leftover blood but recognize that some people will not want their blood used. If you feel this way, you can simply opt out. Your doctors and nurses will *not* know if you opt out, and you will get the same care either way. After you choose to opt out, our system will remember this preference and will prevent your blood samples from being used for research.

8. Will someone let me know if something different is found in my DNA?

NO. Because we will not know whose DNA is whose, it will be impossible to give any person results about his or her DNA. On the other hand, we hope that learning more about how genes work will help make everyone's health care better, including yours.

9. What other concerns do people sometimes have?

Some people have worried that DNA could be used for cloning. This project does not involve cloning. Some people have wondered if it requires taking more blood. It does not require the drawing of any additional blood. Others have wondered if there is an extra cost. You will not have to pay anything extra.

10. Who will make sure that Vanderbilt is using the DNA the way they are supposed to?

This project will be watched over by three groups: 1) our research review board; 2) a separate group of scientists, ethicists, and people who manage the medical center; and 3) a group of people from the community.

Image 2 .
Part of BioVU brochure

Table 1

Demographic characteristics of BioVU

Demographic characteristic	% in BioVU*
Gender	
Male	42%
Female	58%
Race/ethnicity	
African American	11%
Asian	1%
Caucasian	85%
Hispanic/Latino	1%
Other	1%
Age	
<20	1%
20–29	11%
30–39	13%
40–49	16%
50–59	21%
60–69	20%
70–79	12%
80–89	5%
>89	1%

* percentages represent the percentage of all banked samples (73,014 as of December 7, 2009) with this information present in the record

Table 2

Select examples of rare disease counts

Example Rare Disease	#in BioVU
Bell's Palsy	141
Cushing's	129
Spina Bifida	77
Acromegaly and Gigantism	57
Peyronies Disease	57
Meningioma	52
Wilson's Disease	36
Narcolepsy without Cataplexy	17
Ehlers-Danlos Syndrome	9
Pica	9
Tourette Syndrome	9
Pick's Disease	7
Microcephalus	6
Stiff-Man Syndrome	5