Published in final edited form as:

Clin Pharmacol Ther. 2013 February ; 93(2): 204–211. doi:10.1038/clpt.2012.230.

Pediatric sample inclusion in an opt-out biorepository linking DNA to de-identified medical records: Pediatric BioVU

Tracy L. McGregor, MD, MSCI^{*,1,2}, Sara L. Van Driest, MD, PhD^{*,1}, Kyle B. Brothers, MD^{1,3}, Erica A. Bowton, PhD⁴, Louis J. Muglia, MD, PhD^{1,5}, and Dan M. Roden, MD⁶

¹Department of Pediatrics, Vanderbilt University and the Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, TN

²Center for Human Genetics Research, Vanderbilt University, Nashville, TN

³Center for Biomedical Ethics and Society, Vanderbilt University, Nashville, TN

⁴Office of Research, Vanderbilt University, Nashville, TN

⁵Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN

⁶Departments of Medicine and Pharmacology, Vanderbilt University, Nashville, TN

Abstract

The Vanderbilt DNA repository, BioVU, links DNA from leftover clinical blood samples to de-identified electronic medical records. After initiating adult sample collection, pediatric extension required consideration of ethical concerns specific to pediatrics and implementation of specialized DNA extraction methods. In the first year of pediatric sample collection, over 11,000 samples were included from individuals younger than 18 years. We compared the pediatric BioVU cohort to the overall Vanderbilt University Medical Center pediatric population and found similar demographic characteristics; however, the BioVU cohort has higher rates of select diseases, medication exposures, and laboratory testing, demonstrating enriched representation of severe or chronic disease. This unbalanced sample accumulation may accelerate research of some cohorts, but also may limit study of relatively benign conditions and the accrual of unaffected and unbiased control samples. BioVU represents a feasible model for pediatric DNA biobanking but involves both ethical and practical considerations specific to the pediatric population.

Keywords

Pediatrics; Biological Specimen Banks

Corresponding Author: Sara Van Driest, M.D., Ph.D. Department of Pediatrics, Vanderbilt University Medical Center, 8232 DOT, 2200 Children's Way, Nashville, TN 37232; Phone: 615-936-2425; Fax: 615-343-7650 sara.van.driest@vanderbilt.edu.

*These authors contributed equally as the primary authors of this manuscript.

Conflict of Interest/Disclosure: The authors declare no relevant financial interests in this manuscript. The authors declare no other relationships/conditions/circumstances that present a potential conflict of interest that have influenced or give the appearance of potentially influencing the work submitted in this manuscript, other than the funding and research support for the institution and individual authors as described below.

Author Contributions: LJM, TLM, and SVD conceived of the study. All authors were involved in the design of the study. TLM, SLV, KBB, and EAB were involved in the execution of the study, including data acquisition. TLM and SLV analyzed the data for presentation. TLM, SLV, and KBB initially drafted the manuscript. All authors revised the article critically for important intellectual content and approve of the version submitted for publication. TLM and SLV had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Introduction

An appealing vision for improving patient care includes the personalized medicine approach, in which patients are stratified using biomarkers, including genetic markers, to target prevention efforts and individualize treatment. For example, the FDA now includes pharmacogenomic information in drug labels for more than 100 approved medications, including seven boxed warnings.(1) The extension of such personalized medicine approaches to pediatric populations has been limited to date by lack of large pediatric sample sets required for identification and validation of biomarkers such as pharmacogenomic variants. Similarly, research on rare pediatric conditions, including adverse drug reactions, has been hampered by the inability to accumulate study populations.

Biorepositories represent a potential solution by enabling efficient collection of large cohorts.(2-4) These resources prospectively bring together biosamples and medical information into a central repository from which a range of studies can be performed. Over the past decade, several institutions and governments have developed biorepositories; however, a disproportionate number are restricted to adults. The development of pediatric biorepositories enables validation of biomarkers identified in adults, discovery of pediatric-specific biomarkers, study of pediatric pathology, and exploration of disease manifestations across the age spectrum.

BioVU, the Vanderbilt University Medical Center biobank, began collecting samples from adults in 2007, and the design has been previously reported.(5) In brief, BioVU sample collection is based on an opt-out approach that retains blood samples drawn as part of routine clinical care and not consumed by clinical analysis. Medical information linked to these samples is generated from a de-identified image of the electronic medical records (EMR). Vanderbilt's de-identified clinical record database, termed the Synthetic Derivative (SD), contains all de-identifiable elements of the Vanderbilt EMR for more than 2.1 million unique individuals, allowing queries of clinical notes, electronic orders, laboratory values, ICD9 and CPT codes, and demographic data. All clinical encounters documented in the EMR, potentially from birth through the individual's current age, are present in the SD. De-identification includes irreversible transformation of the medical record number via a "one-way hash" procedure, removal of all identifiers, such as names and locations, and concurrent shifting of all the dates in an individual's medical record.(5) BioVU is the biorepository of DNA extracted from leftover clinical blood samples linked to individuals' SD records (Figure 1). Hence, of all the individuals in the SD, a subset has a DNA sample available in BioVU. Furthermore, when a DNA sample is added to the BioVU resource, it becomes linked to the entire SD record for that individual.

BioVU initially accrued samples from adult patients only. During the first three years, the inclusion of pediatric samples was explored and operational challenges were resolved as previously reported.(5) Targeted qualitative research with parents of pediatric patients was conducted, indicating the model would be well-received,(6) and pediatric sample collection began in March 2010. Inclusion of a pediatric sample in BioVU requires that: (1) the patient's parent or caregiver had the opportunity to view the opt-out form and chose not to opt the child out, (2) the patient had blood drawn for clinical purposes using a phlebotomy tube with EDTA anticoagulant, and (3) after all clinical labs are performed, sufficient blood remained in the vial for DNA extraction. Since BioVU is linked to a patient's EMR, de-identified clinical information is available from the person's first encounter through their most recent encounter regardless of the time or age at which the blood sample was collected.

Although the methodological approach and ethical issues have been previously described in detail, the outcome of the implemented methods in the pediatric population have not been

assessed. Information regarding the assembled cohort has relevance for those establishing repositories with similar or disparate models and for interpretation of data derived from such repositories. Here we describe a brief overview of the BioVU approach to address ethical and technical considerations to pediatric biobanking, namely the opt-out model and DNA extraction from small-volume blood tubes. We also report on the characteristics of this unique pediatric biorepository after one year in operation, including the opt-out rate, sample collection rates, and the yield of DNA extraction. We then compare the population in the pediatric BioVU cohort to the set of all pediatric patients in the SD. Finally, we provide data on the potential research utility of this resource by examining population sizes in the SD and BioVU subset for a number of pediatric conditions, medication exposures, and lab values.

Results

The opt-out model and opt-out rate

As previously described, de-identification of the EMR enables the designation of BioVU as a non-human subjects research project under the federal regulations for human subjects research (the “common rule,” 45 CFR 46).(5) Within these regulations, BioVU may collect samples without informed consent.(7, 8) However, since its inception BioVU has adopted an opt-out approach to ensure patients can exercise their preference to not be included.(5, 7, 9) This option, included as a module contained within the Consent for Treatment process signed annually in the outpatient setting, was provided to further protect the autonomy of patients and to enhance community support for this effort.(7, 9-11) We have previously reported a process of patient and family engagement that preceded the start of pediatric sample accrual in BioVU.(6, 10, 11) During pilot use of the opt-out forms in a single pediatric outpatient clinic, parents’ opt-out rate for their children was similar to adult pilot opt-out rates of approximately 5%.(6) Parents particularly favored the ability to have their child participate without additional needlesticks.(6, 12, 13)

Regarding pediatric sample collection, the opt-out model differs from informed consent in at least two important ways. First, information provided to patients focuses on awareness of the program rather than a detailed accounting of risks and benefits. Our research indicates that most patients and parents do not require detailed information to decide whether they will allow their sample or their child’s sample to be included; the choice to opt-out is most likely to be based on a general desire not to have one’s DNA stored and studied.(6) While more detailed information on risks and benefits is available through brochures and phone consultation with program staff, verbal prompts by medical receptionists and the opt-out language included in clinic forms focus on notifying patients that BioVU involves collection of leftover blood samples for DNA research and that the opportunity to opt-out is easily available.

Second, the opt-out model differs from informed consent in the way it addresses authority to make decisions. Informed consent to research in children requires the consent of a legally authorized adult. Research with adolescents additionally requires the assent of the adolescent participant. The presumption in the opt-out model is that any person trusted enough to accompany a child to a clinic visit may exercise an opt-out decision. Adolescents under the age of 18 may also opt themselves out of inclusion, especially when they seek clinical care on their own.(12)

In the first year of pediatric implementation, accrual of samples only took place in the outpatient environment, as an opt-out mechanism was not in place for inpatients. More recently, opt-out language has been included in the pediatric inpatient workflow to allow for inclusion of this patient population as well. In the inpatient setting, the opt-out language is presented to patients at the time of discharge, since hospital admission can be a time when

families may be stressed or distracted and less likely to carefully consider the opt-out language.

The opt-out decision is permanent across BioVU. There is no provision to allow young adults who were opted out as children to opt back in to have their sample included.⁽¹⁴⁾ Since the consent to treat form is presented annually, a patient or parent may choose to opt out at any time, even if they did not choose to do so previously. Once the patient or parent has opted out, any previously collected samples are excluded from future research, and no additional samples are collected. Additionally, a proportion of patients otherwise eligible are randomly excluded from collection, so even those not opting out cannot be certain their sample is included.

Prior to both outpatient and inpatient sample collection roll-out, the project plan was presented to both the Administration and Faculty of the Monroe Carell Jr Children's Hospital at Vanderbilt. A detailed education plan was developed with the Director of Clinical Support Services to ensure both nursing staff and medical receptionists were properly trained on relevant aspects of the BioVU program. BioVU staff perform regular clinic quality checks to ensure brochures and patient materials are available and that medical receptionists remain appropriately trained to speak to patients about BioVU. Additionally, BioVU is regularly included in ongoing staff training sessions.

From March 1, 2010 to March 31, 2011, 102,463 Consent for Treatment forms were signed for pediatric patients. The opt-out box was checked on 8,940 forms, for an opt-out rate of 8.7% (monthly rate: 6% to 11%). By comparison, only 5.4% of adults seen in the first year of collection for BioVU checked the opt-out box. Over time, additional parents or patients have chosen to opt out, resulting in a cumulative opt out rate of 15.4% for all ages.

DNA extraction from leftover clinical blood samples

In order to extract DNA from a leftover blood sample, the sample must have been drawn into a vial with EDTA anticoagulant (purple or lavender top tube). Common testing collected in these vials includes hematology studies (e.g. complete blood count, hemoglobin electrophoresis), blood typing studies, some biomarkers (e.g. sedimentation rate, B-natriuretic peptide, troponin-T), genetic testing, and viral testing.

In the first year, 12,378 pediatric samples were processed (monthly rate: 463 - 1144, Figure 2). The majority of pediatric blood samples were drawn into standard-sized collection vials, and leftover blood volumes were similar to adult samples (> 1 mL). For these vials, DNA extraction was conducted as previously described for the adult samples in BioVU.⁽⁵⁾ However, 3,105 (25%) total pediatric blood samples were drawn using small-volume blood tubes with a capacity of 500 μ L. In the population of children under 2 years, the majority of samples were drawn into the small-volume blood tubes, which require specific handling.

A QIAasympy automated DNA extraction instrument (Qiagen, Valencia, CA), with the ability to process smaller vials and lower blood volumes, was used for these samples. The original blood specimen vial was relabeled and placed directly in the instrument, eliminating sample losses due to transfer. A four step process ensued: 1) volume sensing (minimum volume for extraction is 200 μ L); 2) magnetic bead DNA extraction; 3) DNA elution into a storage tube; and 4) 2-D barcode labeling of the storage tube. Of the small-volume vials, 801 (26%) had insufficient volume for DNA extraction. The absolute amount of DNA extracted from small-volume vials (mean $\pm$ standard deviation: 18.3 $\pm$ 19.1 μ g) was lower than that extracted from pediatric samples in standard collection tubes (89.7 $\pm$ 58.4 μ g). Both types of pediatric collection tubes yielded less total DNA than adult DNA extractions (114.1 $\pm$ 63.1 μ g). Adequate quality of DNA from the pediatric samples is evidenced by similar

performance rates for pediatric and adult samples in downstream applications; for example, on a large genome-wide association study, 0.4% of 246 pediatric samples were excluded due to genotyping efficiency <98%, compared to 1.1% of adult samples (unpublished data).

Pediatric SD cohort and BioVU subset

The pediatric SD cohort with encounters in 2009-2010 consisted of 184,821 individuals, of whom 11,727 (6.3%) had extracted DNA samples deposited into BioVU. Demographic data for these populations are similar (Table 1). Despite the BioVU inclusion requirement of a leftover blood sample, the SD cohort and BioVU subset both had a median of two inpatient days and three outpatient encounters, but exhibited different distributions (Figure 3). Both cohorts have clinical information available from all age intervals (Supplemental Figure 1), with longitudinal data for many individuals (Table 1).

To address the characteristics of both cohorts, we selected diagnoses, medication exposures, and lab values as described in the methods. The data for both cohorts are presented, recognizing that future studies utilizing this resource may or may not require a DNA sample. Table 2 summarizes the incidences of selected diagnoses in the SD and the BioVU subset. For some diagnoses, BioVU representation approximates that of the SD cohort (6%). However, congenital heart disease, seizures, meningitis, and inflammatory bowel disease have higher rates of representation in the BioVU subset, with 19% or more of the SD cases represented in BioVU. The relative distributions of medication exposures reveal that a higher fraction of patients in the BioVU subset have exposures to all medications listed, with 0.2-14.9% of SD patients exposed, compared to 1.9-24.7% of individuals in BioVU (Table 3). This holds true for commonly used medications such as amoxicillin and polyethylene glycol 3350, as well as for more specialized medications such as mesalamine and enoxaparin. The laboratory data in both cohorts also reveal more clinical laboratory testing in the BioVU subset (3.7-26.4% in the SD vs. 15.8-71.3% in BioVU, Table 4), consistent with the requirement of a clinical specimen with residual sample for inclusion in BioVU. A large proportion of patients in both cohorts have common lab results available, such as a complete blood count (26.4% in SD vs. 71.3% in BioVU). Additionally, a substantial number of individuals in each cohort have less common lab results available, such as CSF analyses (6,882 in SD vs. 2,235 in BioVU).

Discussion

Ethical considerations for pediatric biobanking

We previously presented a detailed discussion of the ethical considerations relevant to the use of the BioVU model in a pediatric setting.⁽¹⁴⁾ Importantly, this approach aims to ensure voluntary participation in research through the use of an opt-out paradigm rather than an informed consent process. To assess the effectiveness of this approach, we conducted a number of quantitative and qualitative assessments of patient perspectives and awareness.^(6, 10) The opt-out rate for pediatric patients was slightly higher in the first year of pediatric accrual compared with the first year of sample collection of adult samples in BioVU. Although we reported an initial ~5% opt-out rate in adults, our most recent data show that opt-outs accumulate over time, with the cumulative opt-out rate now reaching ~15% among all patients seen since 2007. Given that a similar pediatric biobank using an *opt-in* model found that 82.6% of parents will sign a form allowing their child's leftover blood to be collected⁽¹⁵⁾, this number provides reassurance that public notification and opt-out procedures used for BioVU are visible and effective. Additionally, these rates are consistent with prior studies demonstrating that that 89-94% of the Nashville community approve of the BioVU model.^(10, 11)

The design of BioVU emphasizes protection of patient privacy, and records are de-identified using state-of-the-art methods.⁽⁵⁾ As previously described, oversight includes input from the Medical Center Ethics community, External Ethics Advisory Board, External Community Advisory Board, Operational Oversight Board, and Scientific Review Committee. In addition, audits are performed by program staff, and the BioVU program is reviewed annually by the Vanderbilt IRB. All investigators must obtain IRB approval for individual SD and BioVU research projects. In addition to this oversight, investigators are required to sign a Data Use Agreement, stating that no attempts will be made to re-identify records by recognition of the clinical presentation, comparison to genetic results previously performed for research, or by any other means. A consequence of this design is that return of results to individual patients is not possible. Significant controversy surrounds the return of individual research results;^(14, 16-18) we have elected to implement an approach that enhances privacy, but consequently precludes return of individual results.

Practical considerations for pediatric biobanking

As discussed, inclusion of pediatric samples in BioVU required development of pediatric-specific opt-out and DNA extraction approaches. These protocols have been successfully implemented, as evidenced by accrual of over 11,000 DNA samples from pediatric patients in the first year. Owing to lower starting volumes, DNA yield from pediatric samples is lower than adults but remains adequate for downstream applications such as SNP genotyping, traditional sequencing, array-based technologies, and possibly even whole exome or genome sequencing. Importantly, BioVU protocols allow for replenishment of stored DNA by additional extractions from subsequent leftover blood samples from individuals already included in BioVU when banked DNA falls below a threshold value, currently set at 10 µg. Although DNA samples can be replenished by subsequent extractions from additional whole blood samples, the extraction protocol does not allow accumulation of leftover whole blood samples in order to achieve the minimum volume for extraction.

Representation and research utility of pediatric BioVU

Pediatric samples comprise a small subset of BioVU as a whole; as of July, 31, 2012, BioVU contained over 131,000 adult samples. In the first year of adult sample collection (2007-2008), 30,000 adult samples were accrued. In addition to the later start, the lower patient volumes and less frequent blood draws in pediatric practice account for the smaller pediatric cohort. The number and type of patients currently in BioVU may present an obstacle for pediatric research using techniques such as genome wide association, which require relatively large cohorts. Collaboration with other institutions and use of adult data to supplement pediatric studies are two strategies for overcoming this hurdle, each with its own limitations.

The pediatric BioVU subset is demographically similar to the pediatric patient population encountered at our institution as represented by all patients in the SD. BioVU includes individuals with inpatient and outpatient encounters, despite the limitation of accrual during only outpatient encounters only during the time frame examined. Notable differences emerge when specific diagnoses, medication exposures, or laboratory parameters are compared in the SD and the BioVU subset. Our data indicate that chronic or severe diagnoses, medication exposures, and laboratory values are more prevalent in the BioVU subset. This may be due to the frequency of blood draws to monitor disease status or medication levels, thus disproportionately increasing the patient's likelihood for BioVU eligibility. Likewise, because of the outpatient lab draw prerequisite for entering BioVU, qualifying labs are expected to be overrepresented in this population.

The high representation of medication-exposed children has facilitated initial pharmacogenomic studies in pediatric cohorts from BioVU. Efforts to validate two drug-genome interactions well established in adults affecting warfarin and simvastatin are underway. Additional studies focused on discovery and/or validation of pharmacogenomic associations for antibiotic, cardiovascular, anticancer, and immunomodulatory agents are in development. These include assessment of both therapeutic benefit and adverse drug events. Performing pharmacogenetic studies in an EMR based repository does pose limitations, particularly with regard to phenotyping.⁽¹⁹⁾ Teams comprised of individuals from multiple areas of expertise – including bioinformatics, clinical pharmacology, genetic epidemiology, and clinical content experts – are collaborating to address these limitations.

Although linking BioVU to the SD enables use of longitudinal retrospective data potentially beginning at birth, a limitation of our approach is that the amount and quality of phenotypic information available in the SD or to BioVU-related research are limited to what is entered into the patient's EMR. Furthermore, the EMR de-identification process removes some data that may be of interest for research purposes. For example, records and samples of related patients are not linked; only family history data explicitly stated in an individual patient's medical record is retained. Additionally, because of random date shifting,⁽⁵⁾ studies related to seasonality (e.g. effectiveness of a vaccine at different time points during a specific seasonal outbreak) cannot be pursued. Further, radiographic images and scanned reports, including paper forms and records from outside institutions or laboratories, are not yet available in the SD pending validated de-identification methods.

Awareness of the above limitations has led to specific strategies for enriching the data available in the SD. Efforts are underway to appropriately de-identify radiologic images; the issue is that names and medical record numbers have not been placed in a standard position on radiographic images. Inpatient medication administration data may also be incorporated to allow determination of exact dosing, timing, and routes of inpatient medications. Demographic information may be enhanced by linking SD records to census tract data which would not be personally identifiable, but would provide more granular background information. Future biospecimen collection may extend beyond extracted DNA to include plasma, serum, or urine samples, allowing measurements of endogenous and exogenous materials, opening a new realm of research possibilities. For future enhancements, the identification risk will be assessed and de-identification methods validated prior to implementation. Although these measures will improve the overall data content of the SD, some will only have an effect after implementation and cannot be applied retroactively.

Biobanking of pediatric samples requires consideration of ethical and practical issues unique to this patient population. BioVU represents one feasible model to address these matters. The 11,000 pediatric samples accrued in the first year of pediatric inclusion into BioVU indicated accelerated collection of disease and exposure cohorts. While limitations such as underrepresentation of benign conditions and dependence on EMR data are barriers to some research endeavors, BioVU represents an important resource to advance child health research.

Methods

Opt-out rates, sample accrual, DNA extraction

The opt-out rate is defined as the ratio of unique patients for whom an opt-out was recorded to total number of unique patients presented with the opportunity to opt-out. These data are ascertained as part of the BioVU program for both pediatric and adult samples. New BioVU samples are successful DNA extractions for individuals meeting the above criteria who are not already included in BioVU. Pediatric samples are those collected from individuals less

than 18 years of age on the day the blood sample is drawn. The quantity of DNA extracted is routinely assessed for each sample prior to DNA storage.

Pediatric SD cohort and BioVU subset

Because the SD de-identification process involves a random date-shifting process,(5) we identified a pediatric cohort representing all patients under the age of 18 with a clinical encounter documented in the SD as occurring in 2009 or 2010. The BioVU pediatric subset of this cohort includes those individuals with an extracted DNA sample available as of the date of query (January 17, 2012). Demographic and clinical data were extracted and summarized for both the SD cohort and the BioVU subset. Death data in the SD reflect information from the Social Security Death Index and/or a “deceased” indicator in the EMR. Race and ethnicity are administratively assigned in the EMR, which approximates genetic ancestry.(20) Medical utilization data were determined using each individual’s entire SD record, potentially from birth through the date of data availability in the SD (April 30, 2011). Clinical visits were extracted from the SD record and categorized by the age of the patient at the time of the visit.

Diagnosis, medication, and laboratory value characterization

The authors selected pediatric diagnoses, medications, and laboratory values with the goal of representing a spectrum of disease prevalence and severity, as well as a variety of pediatric subspecialties. Qualifying ICD9 codes for each of the ten diagnoses (Supplemental Table 1) were queried for the number of unique individuals under 18 years of age with at least one instance of a qualifying code for each diagnosis; individuals could have documentation of multiple diagnoses. To characterize medication exposures, the authors chose fourteen medications representing a range of indications, routes, and frequencies of use in children. Mentions of these drugs were captured by MedEx, a validated bioinformatic tool developed at our institution for extraction of medication data from clinical narratives.(21) The number of unique individuals with medication mentions was determined; individuals could have documentation of multiple medications. Ten common tests or panels assayed in blood, urine, or cerebral spinal fluid were again chosen by the authors, and the number of unique individuals with at least one laboratory value for each of the tests or panels was determined for the SD cohort and BioVU subset based on their entire de-identified medical record. When disparate numbers of individuals were found for individual components of lab panels (e.g. white blood cell count vs. platelet count), the lowest number for that panel is reported.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the contributions of Melissa Basford, Gordon Bernard, Ellen Wright Clayton, Miguel Herrera, Jennifer Madison, Dan Masys, Jill Pulley, Cara Sutcliffe, and Xiaoming Wang for their significant contribution to this work. Portions of this study were supported by 5U01HG004603, supporting the Vanderbilt site in NIH-NHGRI’s Electronic Medical Records and Genomics (eMERGE) network; the Vanderbilt Institute for Clinical and Translational Research (VICTR), NCRR/NIH grant UL1 RR024975; NIH/NIEHS grant K12 ES015855 (TLM); and NIH/NIGMS Clinical Pharmacology Training Program 5T32 GM007569-33 (SLV). Funds for purchase of the QIASymphony instrument were obtained through the NIH’s Shared Instrumentation Grant Program (S10 RR027764: “Automated DNA Extraction for Small Volume Samples Enabling Pediatric Biobanking”).

Funding/Research Support TLM, SLV, KBB, EAB, LJM, DMR: Portions of this study were supported by 5U01HG004603, supporting the Vanderbilt site in NIH-NHGRI’s Electronic Medical Records and Genomics (eMERGE) network, and NCRR/NIH grant UL1 RR024975, supporting the Vanderbilt Institute for Clinical and Translational Research (VICTR). Purchase of the QIASymphony supported by the NIH’s Shared Instrumentation

Grant Program (S10 RR027764: “Automated DNA Extraction for Small Volume Samples Enabling Pediatric Biobanking”).

TLM: Portions of this study supported by NIH/NIEHS grant K12 ES015855.

SLV: Portions of this study supported by NIH/NIGMS Clinical Pharmacology Training Program 5T32 GM007569-33.

Abbreviations

EMR	Electronic Medical Record
SD	Synthetic Derivative

References

- (1). U.S. Food and Drug Administration. [Accessed 02-10-2012] Table of Pharmacogenomic Biomarkers in Drug Labels. 2012. <<http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm>>
- (2). Ginsburg GS, Burke TW, Febbo P. Centralized biorepositories for genetic and genomic research. JAMA- J Am Med Assoc. 2008; 299:1359–61.
- (3). McCarty CA, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics. 2011; 4:13. [PubMed: 21269473]
- (4). Brisson AR, Matsui D, Rieder MJ, Fraser DD. Translational Research in Pediatrics: Tissue Sampling and Biobanking. Pediatrics. 2012; 129:153–62. [PubMed: 22144705]
- (5). Roden DM, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. 2008; 84:362–9. [PubMed: 18500243]
- (6). Brothers KB, Clayton EW. Parental Perspectives on a Pediatric Human Non-Subjects Biobank. Am J Bioethics: Primary Research. 2012; 3:21–9.
- (7). Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects protections applied in an opt-out, de-identified biobank. Clin Transl Sci. 2010; 3:42–8. [PubMed: 20443953]
- (8). OHRP. Guidance on Research Involving Coded Private Information or Biological Specimens. Office of Human Research Protections; Rockville, MD: 2008.
- (9). Brothers KB, Clayton EW. “Human non-subjects research”: privacy and compliance. Am J Bioethics. 2010; 10:15–7.
- (10). Brothers KB, Morrison DR, Clayton EW. Two large-scale surveys on community attitudes toward an opt-out biobank. Am J Med Genet A. 2011; 155A:2982–90. [PubMed: 22065592]
- (11). Pulley JM, Brace MM, Bernard GR, Masys DR. Attitudes and perceptions of patients towards methods of establishing a DNA biobank. Cell Tissue Bank. 2008; 9:55–65. [PubMed: 17960495]
- (12). Hens K, Nys H, Cassiman JJ, Dierickx K. The storage and use of biological tissue samples from minors for research: a focus group study. Public Health Genomi. 2011; 14:68–76.
- (13). Jenkins MM, et al. Maternal attitudes toward DNA collection for gene-environment studies: a qualitative research study. Am J Med Genet A. 2009; 149A:2378–86. [PubMed: 19839045]
- (14). Brothers KB. Biobanking in pediatrics: the human nonsubjects approach. Per Med. 2011; 8:79. [PubMed: 21691419]
- (15). Marsolo K, et al. Challenges in creating an opt-in biobank with a registrar-based consent process and a commercial EHR. J Am Med Inform Assoc. 2012
- (16). Hens K, Nys H, Cassiman JJ, Dierickx K. Genetic research on stored tissue samples from minors: a systematic review of the ethical literature. Am J Med Genet A. 2009; 149:2346–58. [PubMed: 19764035]
- (17). Hens K, Nys H, Cassiman JJ, Dierickx K. The return of individual research findings in paediatric genetic research. J Med Ethics. 2011; 37:179. [PubMed: 21059631]
- (18). Tabor HK, et al. Parent perspectives on pediatric genetic research and implications for genotype-driven research recruitment. J Empir Res Hum Res. 2011; 6:41–52.

- (19). Roden DM, Xu H, Denny JC, Wilke RA. Electronic Medical Records as a Tool in Clinical Pharmacology: Opportunities and Challenges. *Clin Pharmacol Ther.* 2012; 91:1083–6. [PubMed: 22534870]
- (20). Dumitrescu L, et al. Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. *Genet Med.* 2010; 12:648–50. [PubMed: 20733501]
- (21). Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. *J Am Med Inform Assoc.* 2010; 17:19–24. [PubMed: 20064797]

Study Highlights

What is the current knowledge on the topic?

Although prospective pediatric biorepositories are an important resource for child health research, efforts to develop such collections have been hampered by multiple factors, including limited accrual and lack of accommodations for pediatric considerations.

What question this study addressed?

At this institution, pediatric samples are now included in a large, prospective, opt-out biorepository (BioVU) linking DNA to de-identified electronic medical records. We sought to characterize the pediatric cohorts represented in the de-identified data bank and the DNA repository.

What this study adds to our knowledge?

Our approach to pediatric inclusion in BioVU provides a feasible example for collection of samples and medical record data, with bias toward collection of samples from patients with medication exposures, laboratory testing, and select diagnoses.

How this might change clinical pharmacology and therapeutics?

This biorepository includes a diverse range of diseases, medication exposures, and laboratory assessments to facilitate pediatric research in clinical pharmacology, including investigations of pharmacogenomics, pharmacoepidemiology, and adverse drug events.

Figure 1. Schematic representation of the relationship between the Synthetic Derivative (de-identified medical record information) and the BioVU biorepository (de-identified medical record information linked to a de-identified biospecimen).

Figure 2. Monthly accrual for new unique samples and cumulative accrual over the first year from individuals <18 years of age at the time of sample collection.

Figure 3. Medical usage data for both the full SD cohort and the BioVU subset for (A) cumulative inpatient days and (B) unique outpatient encounters. The data are presented as percentages of the respective cohort because of the large disparity in absolute cohort sizes.

Table 1

Demographic parameters for the Synthetic Derivative cohort and the BioVU subset

	Synthetic Derivative	BioVU
Number in Cohort	184,821	11,727
Deaths (rate)	1,287 (7.0/1000)	64 (5.5/1000)
Current Age, y [*]	7 (3-12)	9 (3-14)
Male, No (%)	98,133 (53.1%)	5,845 (49.9%)
Race/Ethnicity, No (%)		
White	84,740 (45.9%)	6,348 (54.1%)
Hispanic	10,314 (5.6%)	692 (5.9%)
Black or African American	26,445 (14.3%)	1,710 (14.6%)
Other	6,276 (3.4%)	469 (4.0%)
Unknown	57,046 (30.9%)	2,508 (21.4%)
Utilization Data		
Days of inpatient stay [*]	2 (1-4)	2 (1-6)
Outpatient clinic encounters [*]	3 (2-5)	3 (3-11)
Span between first and last clinical notes, y [*]	2.5 (0.5-6.0)	3.3 (1.5-7.4)

^{*} median (25-75% ile)

Table 2

Occurrence of selected diagnoses in the Synthetic Derivative cohort and the BioVU subset

Diagnosis	Synthetic Derivative ^a	BioVU ^a	BioVU Representation ^b
Well Child Check	17,723 (9.6%)	1,964 (16.8%)	11.1%
Acute or Chronic Otitis Media	10,321 (5.6%)	808 (6.9%)	7.8%
Asthma	10,108 (5.5%)	969 (8.3%)	9.6%
Strep Pharyngitis	4,281 (2.3%)	188 (1.6%)	4.4%
Obesity	2,785 (1.5%)	278 (2.4%)	10.0%
Type I Diabetes	2,288 (1.2%)	240 (2.1%)	10.5%
Seizures	3,634 (2.0%)	710 (6.1%)	19.5%
Meningitis	277 (0.2%)	67 (0.6%)	24.2%
Congenital Heart Disease	3,808 (2.1%)	739 (6.3%)	19.4%
Inflammatory Bowel Disease	406 (0.2%)	269 (2.3%)	66.3%

^aN (% of cohort or subset)^bPercentage of SD cohort included in BioVU subset

Table 3

Occurrence of selected medication exposures in the Synthetic Derivative cohort and BioVU subset

Medication	Synthetic Derivative ^a	BioVU ^a	BioVU Representation ^b
Amoxicillin	27,560 (14.9%)	2,444 (20.8%)	8.9%
Ceftriaxone	9,513 (5.1%)	1,678 (14.3%)	17.6%
Vancomycin	5,582 (3.0%)	1,362 (11.6%)	24.4%
Albuterol	20,866 (11.3%)	1,959 (16.7%)	9.4%
Montelukast	11,306 (6.1%)	949 (8.1%)	8.4%
Morphine	15,963 (8.6%)	2,900 (24.7%)	18.2%
Polyethylene glycol 3350	12,174 (6.6%)	2,293 (19.6%)	18.8%
Mesalamine	390 (0.2%)	219 (1.9%)	56.2%
Furosemide	4,930 (2.7%)	1,026 (8.7%)	20.8%
Enalapril	1,225 (0.7%)	358 (3.1%)	29.2%
Levetiracetam	3,425 (1.9%)	618 (5.3%)	18.0%
Levothyroxine	1,543 (0.8%)	234 (2.0%)	15.2%
Enoxaparin	772 (0.4%)	233 (2.0%)	30.2%
Mercaptopurine	453 (0.2%)	320 (2.7%)	70.6%

^a N (% of cohort or subset)^b Percentage of SD cohort included in BioVU subset

Table 4

Occurrence of selected laboratory data available in the Synthetic Derivative Cohort and the BioVU subset

Source	Laboratory	Synthetic Derivative ^a	BioVU ^a	BioVU Representation ^b
Blood	Complete Blood Count ¹	48,720 (26.4%)	8,361 (71.3%)	17.2%
	White Blood Cell Differential ²	22,742 (12.3%)	6,227 (53.1%)	27.4%
	Basic Metabolic Profile ³	37,390 (20.2%)	6,672 (56.9%)	17.8%
	Complete Metabolic Profile ⁴	20,364 (11.0%)	4,500 (38.4%)	22.1%
	Thyroid Function Tests ⁵	13,937 (7.5%)	2,270 (19.4%)	16.3%
	Coagulation Profile ⁶	8,611 (4.7%)	2,142 (18.3%)	24.9%
	C-reactive peptide	13,331 (7.2%)	2,823 (24.1%)	21.2%
Urine	Urinalysis ⁷	19,021 (10.3%)	4,158 (35.5%)	21.9%
	Urine microscopy ⁸	8,197 (4.4%)	1,851 (15.8%)	22.6%
CSF	CSF analysis ⁹	6,882 (3.7%)	2,235 (19.1%)	32.5%

^a N (% of cohort or subset)^b Percentage of SD cohort included in BioVU subset¹ white blood cell count, red blood cell count, hemoglobin, hematocrit, platelet count, mean corpuscular volume, red cell distribution width.² neutrophils, lymphocytes, monocytes, eosinophils, and basophils.³ sodium, potassium, carbon dioxide, chloride, BUN, creatinine, calcium, glucose.⁴ basic metabolic profile, albumin, total protein, alkaline phosphatase, ALT/SGPT, AST/SGOT, bilirubin.⁵ thyroid stimulating hormone, free T4.⁶ prothrombin time/INR, activated partial thromboplastin time.⁷ urine color, specific gravity, pH, nitrites, leukocyte esterase, ketones, glucose, blood, urobilinogen.⁸ urine white blood cells, red blood cells, mucous, bacteria.⁹ CSF glucose, protein, nucleated cells, red blood cells.