Inclusion of Pediatric Samples in an Opt-Out Biorepository Linking DNA to De-Identified Medical Records: Pediatric BioVU

TL McGregor[1,2], SL Van Driest[1], KB Brothers[1,3], EA Bowton[4], LJ Muglia[1,5] and DM Roden[6,7]

The Vanderbilt DNA repository, BioVU, links DNA from leftover clinical blood samples to de-identified electronic medical records (EMRs). 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, more than 11,000 samples from individuals younger than 18 years were included. We compared data from the pediatric BioVU cohort with those from the overall Vanderbilt University Medical Center pediatric population and found similar demographic characteristics; however, the BioVU cohort had higher rates of select diseases, medication exposures, and laboratory testing, demonstrating enriched representation of severe or chronic disease. The fact that the sample accumulation is not balanced may accelerate research in some cohorts while limiting the 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.

In personalized medicine—an important approach in the overall vision of improved patient care—patients are stratified using biomarkers, including genetic markers, to target prevention efforts and individualize treatment. For example, the US Food and Drug Administration now includes pharmacogenomic information (including seven boxed warnings) in drug labels for more than 100 approved medications.[1] To date, the extension of such personalized-medicine approaches to pediatric populations has been limited by the lack of the large pediatric sample sets that are 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 sufficiently large study populations.

Biorepositories represent a potential solution to this problem 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 of these are restricted to samples from adults. The development of pediatric biorepositories would enable validation of biomarkers identified in adults, the discovery of pediatric-specific biomarkers, the study of pediatric pathology, and the 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 described.[5] In brief, BioVU sample collection is based on an opt-out approach. Blood samples drawn as part of routine clinical care and not consumed in clinical analyses are retained. Medical information linked to each sample is generated from a de-identified image of the electronic medical record (EMR). Vanderbilt's de-identified clinical record database, termed the synthetic derivative, 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 (ninth revision of the International Classification of Diseases) and CPT (Current Procedural Terminology) codes, and demographic data. All clinical encounters documented in the EMR, potentially from birth through the individual's current age, are present in the synthetic derivative. De-identification includes irreversible transformation of the medical record number via a "one-way hash" procedure; removal of all identifiers, such as

The first two authors contributed equally as the primary authors of this article.
[1]Department of Pediatrics, Vanderbilt University and the Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, Tennessee, USA; [2]Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, USA; [3]Center for Biomedical Ethics and Society, Vanderbilt University, Nashville, Tennessee, USA; [4]Office of Research, Vanderbilt University, Nashville, Tennessee, USA; [5]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, USA; [6]Department of Medicine, Vanderbilt University, Nashville, Tennessee, USA; [7]Department of Pharmacology, Vanderbilt University, Nashville, Tennessee, USA.
Correspondence: SL Van Driest (sara.van.driest@vanderbilt.edu)

Received 3 August 2012; accepted 9 November 2012; advance online publication 2 January 2013. doi:10.1038/clpt.2012.230

names and locations; and concurrent shifting of all the dates in an individual's medical record.[5] Given that BioVU is the biorepository of DNA extracted from leftover clinical blood samples linked to individuals' synthetic derivative records (**Figure 1**), DNA samples are available for a subset of all the individuals in the synthetic derivative. Furthermore, when a DNA sample is added to the BioVU resource, it becomes linked to the entire synthetic derivative record for that individual.

BioVU initially included samples from adult patients only. In the first three years, the possibility of including pediatric samples was explored and the operational challenges were resolved, as previously reported.[5] Targeted qualitative research conducted with parents of pediatric patients indicated that the model would be well received.[6] Collection of pediatric samples began in March 2010. Inclusion of a pediatric sample in BioVU requires that (i) the patient's parent or caregiver has had the opportunity to view the opt-out form and has chosen not to opt out on behalf of the child, (ii) the patient has had blood drawn for clinical purposes using a phlebotomy tube with EDTA anticoagulant, and (iii) after all clinical laboratory tests were performed, sufficient blood remained in the vial for DNA extraction. Because BioVU is linked to a patient's EMR, de-identified clinical information is available from the person's first encounter through to his or her most recent encounter, regardless of the time point or age at which the blood sample was collected.

Although the methodological approach and ethical issues have been previously described in detail, the outcomes of the methods implemented 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 give a brief overview of the BioVU approach to addressing ethical and technical considerations relating 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, which has been in operation for one year, including the opt-out rates, sample collection rates, and yield of DNA extraction. We then compare the data from the population in the pediatric BioVU cohort with those from the set of all pediatric patients in the synthetic derivative. Finally, we provide data on the potential research utility of this resource by examining population sizes in the synthetic derivative and BioVU subset as they relate to several pediatric conditions, medication exposures, and laboratory-test values.

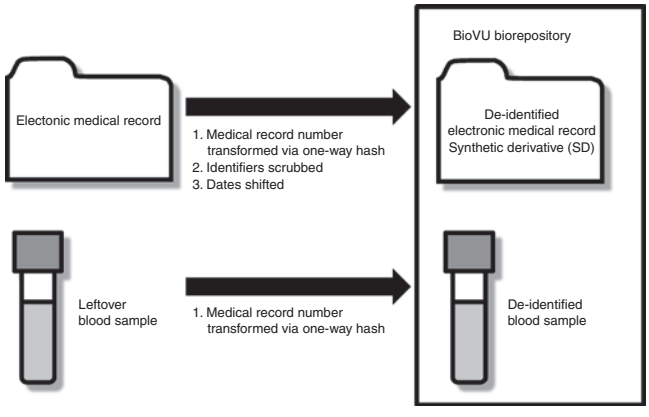

Figure 1 Schematic representation of the relationship between the synthetic derivative (de-identified medical record information) and the BioVU biorepository (de-identified medical record information linked to a de-identified biospecimen).

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] From its inception, however, BioVU has adopted an opt-out approach to ensure that patients (and parents of pediatric patients) can choose not to have their samples included in the repository if they so wish.[5,7,9] This option, included as a module in the Consent for Treatment signed annually in the outpatient setting, was provided so as to further protect the autonomy of patients and to enhance community support for this effort.[7,9–11] We have previously described the process of engagement with patients and their families that was undertaken before starting the collection of pediatric samples 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 rates in pilot studies with adult patients, ~5%.[6] Parents particularly appreciated the fact that inclusion of the sample in the repository would not entail any additional needle sticks.[6,12,13]

With respect to samples from pediatric patients, the opt-out model differs from informed consent in at least two important ways. First, the information provided to patients (or to parents of pediatric patients) focuses on awareness of the program rather than on 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] Although 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 with respect to authority to make decisions. Whereas informed consent to carry out research in children requires the consent of a legally authorized adult, and research in adolescents additionally requires the assent of the adolescent participant, the presumption in the opt-out model is that any adult whom the family has trusted enough to escort the child to a clinic visit may exercise an opt-out decision. Adolescents <18 years of age may also opt out from inclusion, especially when they seek clinical care on their own.[12]

In the first year after samples from pediatric patients began to be included in BioVU, accrual of samples took place only in the outpatient environment because there was no opt-out

mechanism in place for inpatients. More recently, opt-out opportunity has been incorporated into the pediatric inpatient workflow to allow for inclusion of this patient population as well. In the inpatient setting, the opt-out option is presented at the time of discharge because when the child is admitted, families may be under stress or distracted and therefore less likely to carefully consider the opt-out decision.

For BioVU, any opt-out decision is permanent. There is no provision to allow young adults who were opted out as children to opt back in to have their sample included.[14] Because the Consent for Treatment form is presented annually, a patient or parent may choose to opt out at any time, even if they had not chosen to do so previously. Once the patient or parent has opted out, samples from the individual collected previously are excluded from future research, and no additional samples are collected. In addition, because a proportion of patients whose data are otherwise eligible are randomly excluded, even those who do not opt out cannot be certain that their sample is included in BioVU.

Before sample collection rollout, with both outpatients and inpatients, the project plan was presented to the administration and the 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 that nursing staff and medical receptionists were properly trained in the relevant aspects of the BioVU program. BioVU staff perform regular clinic quality checks to ensure that brochures and patient materials are available and that medical receptionists remain appropriately trained to speak to patients about BioVU. In addition, information pertaining to BioVU is regularly included in ongoing staff training sessions.

From 1 March 2010 to 31 March 2011, 102,463 Consent for Treatment forms were signed on behalf of pediatric patients. The opt-out box was checked on 8,940 forms, yielding an opt-out rate of 8.7% (monthly rate: 6–11%). By comparison, only 5.4% of adult patients 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 patients of all ages.

DNA extraction from leftover clinical blood samples

For DNA to be extracted from a leftover blood sample, the sample must have been drawn into a vial with EDTA anticoagulant (purple- or lavender-top tube). Common laboratory tests for which blood samples are collected in these vials include hematology studies (e.g., complete blood count and hemoglobin electrophoresis), blood-typing studies, some biomarkers (e.g., sedimentation rate, B-natriuretic peptide, and troponin-T), genetic testing, and virus testing.

In the first year, 12,378 pediatric samples were processed (monthly rate: 463–1,144, **Figure 2**). The majority of the pediatric blood samples were drawn into standard-sized collection vials, and leftover blood volumes were similar to those in adult samples (>1 ml). From these samples, DNA extraction was conducted using the method previously described for the adult samples in BioVU.[5] However, 3,105 (25% of the total) pediatric blood samples were drawn into small-volume (500-µl) blood tubes. In pediatric patients <2 years of age, the majority of samples were drawn into these small-volume blood tubes. Specialized methods are required to extract DNA from these small-tube samples.

A QIAsymphony automated DNA extraction instrument (Qiagen, Valencia, CA), with the capability of processing smaller vials and lower blood volumes, was used for these samples. The original blood-specimen vial was relabeled and placed directly in the instrument, thereby eliminating sample losses due to transfer. A four-step process was then employed: (i) volume sensing (minimum sample volume necessary for DNA extraction was 200 µl), (ii) DNA extraction by magnetic bead method, (iii) DNA elution into a storage tube, and (iv) two-dimensional barcode labeling of the storage tube. Of the small-volume vials, 801 (26%) contained insufficient volume for DNA extraction. The absolute amounts of DNA extracted from small-volume vials (mean ± SD: 18.3 ± 19.1 µg) were lower than those extracted from pediatric samples from

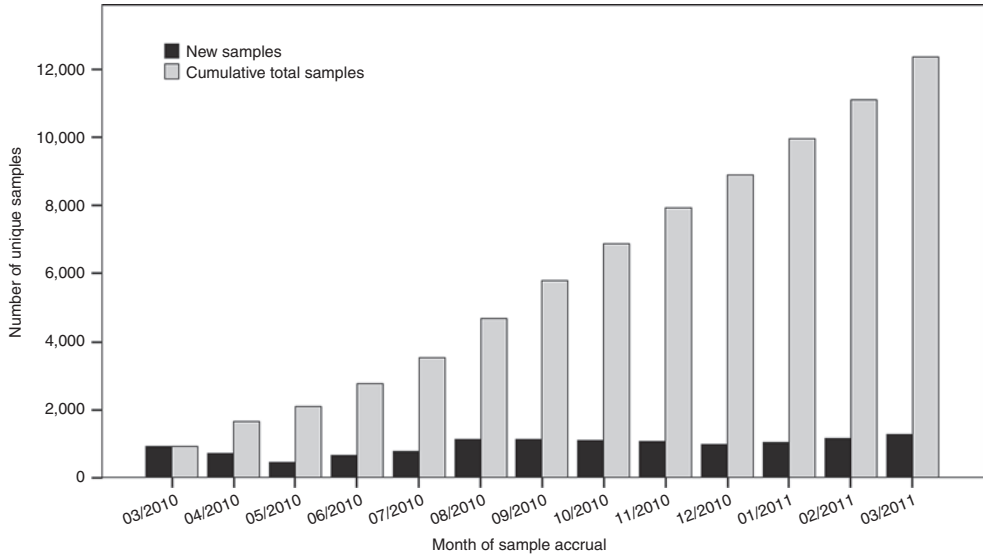

Figure 2 Monthly accrual of new unique samples and cumulative accrual over the first year from individuals <18 years of age at the time of sample collection.

standard collection tubes (89.7 ± 58.4 μg). Both types of pediatric collection tubes yielded less total DNA than adult samples did (114.1 ± 63.1 μg). The fact that the performance rates for pediatric and adult samples in downstream applications were similar was evidence that the DNA extracted from the pediatric samples was of adequate quality. For example, in a large genome-wide association study, 0.4% of 246 pediatric samples were excluded because of genotyping efficiency <98%, as compared with 1.1% of samples from adults (unpublished data).

Pediatric synthetic derivative cohort and BioVU subset

The pediatric synthetic derivative cohort with encounters in 2009–2010 consisted of 184,821 individuals, and DNA samples were extracted and deposited into BioVU from 11,727 (6.3%). The demographics of these populations are similar (Table 1). Despite the BioVU inclusion requirement that a leftover blood sample be available, both the synthetic derivative cohort and the BioVU subset had medians of two inpatient days and three outpatient encounters; however, they exhibited different distribution patterns (Figure 3). With respect to both cohorts, clinical information was available for all age intervals (**Supplementary Figure S1** online), along with longitudinal data for many individuals (Table 1).

To address the characteristics of both cohorts, we selected diagnoses, medication exposures, and laboratory values as described in the Methods section. The data for both cohorts are presented, recognizing that future studies that utilize this resource may not necessarily require a DNA sample. Table 2 summarizes the incidences of selected diagnoses in the synthetic derivative and in the BioVU subset. For some diagnoses, the BioVU representation approximates that of the synthetic derivative 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 synthetic derivative cases in these categories being represented in BioVU. The relative distributions of medication exposures reveal that a higher fraction of patients in the BioVU subset have had exposures to each of the medications listed (1.9–24.7% of individuals in the BioVU subset vs. 0.2–14.9% of patients in the total synthetic derivative) (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 the two cohorts also reveal that more clinical laboratory testing had occurred in the BioVU subset (15.8–71.3% in the BioVU subset vs. 3.7–26.4% in the synthetic derivative) (Table 4). This was consistent with the BioVU inclusion requirement that a residual sample of a clinical specimen be available. For a large proportion of the patients in both cohorts, common laboratory test

Table 1 Demographic parameters of the synthetic derivative cohort and the BioVU subset

	Synthetic derivative	BioVU
Number in cohort	184,821	11,727
Deaths (rate)	1,287 (7.0/1,000)	64 (5.5/1,000)
Current age, years[a]	7 (3–12)	9 (3–14)
Male, *n* (%)	98,133 (53.1%)	5,845 (49.9%)
Race/ethnicity, *n* (%)		
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[a]	2 (1–4)	2 (1–6)
Outpatient clinic encounters[a]	3 (2–5)	3 (3–11)
Span between first and last clinical notes, years[a]	2.5 (0.5–6.0)	3.3 (1.5–7.4)

[a]Median (25–75 percentile).

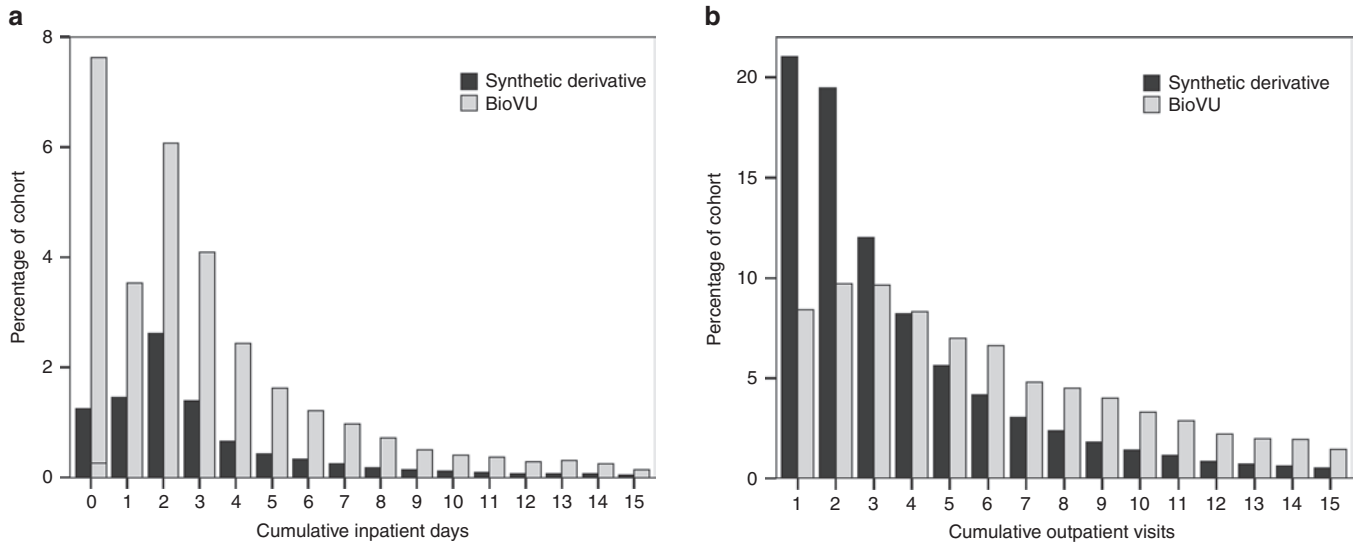

Figure 3 Medical usage data for the full synthetic derivative cohort and for the BioVU subset with respect to (**a**) cumulative inpatient days and (**b**) cumulative outpatient encounters. The data are presented as percentages of the respective cohort because of the large disparity in absolute cohort sizes.

Table 2 Occurrence of selected diagnoses in the synthetic derivative cohort and in 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%
Streptococcal pharyngitis	4,281 (2.3%)	188 (1.6%)	4.4%
Obesity	2,785 (1.5%)	278 (2.4%)	10.0%
Type 1 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%

[a]*n* (Percentage of cohort or subset). [b]Percentage of synthetic derivative cohort included in BioVU subset.

Table 3 Occurrence of selected medication exposures in the synthetic derivative cohort and in the 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* (Percentage of cohort or subset). [b]Percentage of synthetic derivative cohort included in BioVU subset.

results were available, such as a complete blood count (26.4% in the synthetic derivative vs. 71.3% in BioVU). In addition, for a substantial number of individuals in each cohort, less common lab results were available, such as cerebrospinal fluid analyses (6,882 in the synthetic derivative vs. 2,235 in BioVU).

DISCUSSION

Ethical considerations for pediatric biobanking

We have 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

Table 4 Occurrence of selected laboratory data in the synthetic derivative cohort and the BioVU subset

Source	Laboratory	Synthetic derivative[a]	BioVU[a]	BioVU representation[b]
Blood	Complete blood count[c]	48,720 (26.4%)	8,361 (71.3%)	17.2%
	White blood cell differential[d]	22,742 (12.3%)	6,227 (53.1%)	27.4%
	Basic metabolic profile[e]	37,390 (20.2%)	6,672 (56.9%)	17.8%
	Complete metabolic profile[f]	20,364 (11.0%)	4,500 (38.4%)	22.1%
	Thyroid function test[g]	13,937 (7.5%)	2,270 (19.4%)	16.3%
	Coagulation profile[h]	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[i]	19,021 (10.3%)	4,158 (35.5%)	21.9%
	Urine microscopy[j]	8,197 (4.4%)	1,851 (15.8%)	22.6%
CSF	CSF analysis[k]	6,882 (3.7%)	2,235 (19.1%)	32.5%

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BUN, blood urea nitrogen; CSF, cerebrospinal fluid; INR, international normalized ratio; SGOT, serum glutamic oxaloacetic transaminase; SGPT, serum glutamic pyruvic transaminase.

[a]*n* (Percentage of cohort or subset). [b]Percentage of synthetic derivative cohort included in BioVU subset. [c]White blood cell count, red blood cell count, hemoglobin, hematocrit, platelet count, mean corpuscular volume, red cell distribution width. [d]Neutrophils, lymphocytes, monocytes, eosinophils, basophils. [e]Sodium, potassium, carbon dioxide, chloride, BUN, creatinine, calcium, glucose. [f]Basic metabolic profile, albumin, total protein, alkaline phosphatase, ALT/SGPT, AST/SGOT, bilirubin. [g]Thyroid-stimulating hormone, free T4. [h]Prothrombin time/INR, activated partial thromboplastin time. [i]Urine color, specific gravity, pH, nitrites, leukocyte esterase, ketones, glucose, blood, urobilinogen. [j]Urine white blood cells, red blood cells, mucus, bacteria. [k]CSF glucose, protein, nucleated cells, red blood cells.

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 accrual to the BioVU repository as compared with the rate during the first year of sample collection from adult patients. We reported an initial opt-out rate of ~5% in adult patients. However, our most recent data show that opt-outs accumulate over time, and the cumulative opt-out rate has now reached ~15% among all patients seen since 2007. Given that a similar pediatric biobank using an opt-in model found that 82.6% of parents were willing to sign a form allowing their child's leftover blood to be used for a biodepository,[15] the rate of opt-out in BioVU provides reassurance that the public notification and the opt-out procedures used for BioVU are visible and effective. In addition, these rates are consistent with prior studies showing that that 89–94% of the Nashville community approved 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 Institutional Review Board. All investigators must obtain institutional review board approval for individual synthetic derivative 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 with genetic results previously used 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, while enhancing privacy, precludes return of individual results.

Practical considerations for pediatric biobanking

As described earlier, 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 more than 11,000 DNA samples from pediatric patients in the first year. Owing to smaller starting volumes of the leftover samples, DNA yields from pediatric samples are lower than those from adults, but they are adequate for downstream applications such as single-nucleotide-polymorphism genotyping, traditional sequencing, array-based technologies, and possibly even whole-exome or whole-genome sequencing. Of note, when the amount of banked DNA from an individual falls below a threshold value, currently set at 10 μg, BioVU protocols allow for replenishment of stored DNA by additional extractions from subsequent leftover blood samples from the individual. It is important to note that the extraction protocol for replenishment does not permit pooling of leftover whole-blood samples in order to achieve the minimum volume necessary for extraction.

Representation and research utility of pediatric BioVU

Pediatric samples comprise a small subset of BioVU as a whole; as of 31 July 2012, BioVU contained more than 131,000 samples from adult patients. In the first year of collection of samples from adults (2007–2008), 30,000 adult samples accrued to the repository. Apart from the later start, the lower patient volumes and the less frequent blood draws in pediatric practice account for the smaller pediatric cohort. The number and type of patients currently represented 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 the use of data from adults to supplement pediatric studies are two potential 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 synthetic derivative. BioVU includes samples from individuals with inpatient encounters as well as those with outpatient encounters, despite the limitation of accrual only with outpatient encounters during the time frame examined. When compared, notable differences emerged between the synthetic derivative and the BioVU subset with respect to specific diagnoses, medication exposures, and laboratory parameters. Our data indicate that diagnoses of chronic conditions or severe conditions, medication exposures, and laboratory values are more prevalent in the BioVU subset. This could be attributable to the greater frequency of blood draws to monitor disease status or medication levels in these patients, thereby disproportionately increasing the likelihood that their samples will become eligible for inclusion in BioVU. Likewise, given the BioVU-inclusion prerequisite of blood-sample collection in the outpatient setting, qualifying labs are expected to be overrepresented in this population.

The high representation of medication-exposed pediatric patients has facilitated initial pharmacogenomic studies in pediatric cohorts from BioVU. Efforts are under way to validate in pediatric patients two drug–genome interactions that have been well established in adults, namely, those involving warfarin and simvastatin. Additional studies focusing on discovery and/or validation of pharmacogenomic associations for antibiotic, cardiovascular, anticancer, and immunomodulatory agents are in development. These include assessments of both therapeutic benefits and adverse drug events. Performing pharmacogenetic studies in an EMR-based repository poses limitations, particularly with regard to phenotyping.[19] Teams composed 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 synthetic derivative enables the use of longitudinal retrospective data potentially beginning at birth, a limitation of our approach is that the amount and quality of phenotypic information in the synthetic derivative and available for 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 are retained. In addition, 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. Furthermore, radiographic images and scanned reports, including paper forms and records from outside institutions and laboratories, are not yet available in the synthetic derivative, pending validated de-identification methods.

Awareness of these limitations has led to specific strategies for enriching the data available in the synthetic derivative. Efforts are under way to carry out appropriate de-identification of radiologic images; the problem encountered during this process is that names and medical record numbers are not placed in a standard location on radiographic images. Data relating to the administration of inpatient medications may also be incorporated to allow determination of exact dosages, timing, and routes of administration. Demographic information may be enhanced by linking synthetic derivative records to census-tract data; the latter would not be personally identifiable but would provide more granular background information. In the future, biospecimen collections in the repository may extend beyond extracted DNA to include plasma, serum, or urine samples; this would allow measurements of endogenous and exogenous materials, opening a new realm of research possibilities. Before implementing future enhancements, the

identification risk will be assessed and de-identification methods will be validated. Although these measures will improve the overall data content of the synthetic derivative, some will have an effect only 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 preferential inclusion of cohorts with more serious disease conditions and greater medication exposures. Although limitations such as underrepresentation of benign conditions and dependence on EMR data are barriers to some research endeavors, BioVU represents an important resource for advancing child health research.

METHODS

Opt-out rates, sample accrual, and DNA extraction. The opt-out rate is defined as the ratio of unique patients for whom an opt-out was recorded to the total number of unique patients presented with the opportunity to opt out. These data are ascertained for both pediatric and adult samples as part of the BioVU program. New BioVU inclusions are successful DNA extractions from samples taken from patients who meet the criteria and are not already included in BioVU. A "pediatric sample" is defined as one collected from an individual <18 years of age on the day the sample was collected. The quantity of DNA extracted from each sample is routinely assessed before storage.

Pediatric synthetic derivative cohort and BioVU subset. Because the synthetic derivative de-identification process involves a random date-shifting process,[5] we identified a pediatric cohort representing all patients <18 years of age who had a clinical encounter documented in the synthetic derivative as occurring in 2009 or 2010. The BioVU pediatric subset of this cohort includes those for whom an extracted DNA sample was available as of the date of query (17 January 2012). Demographic and clinical data were extracted and summarized both for the synthetic derivative cohort and for the BioVU subset. Data about deaths in the synthetic derivative 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, approximating genetic ancestry.[20] Medical-utilization data were determined using each individual's entire synthetic derivative record, potentially from birth through the date of data availability in the synthetic derivative (30 April 2011). Clinical-visit dates were extracted from the synthetic derivative record and categorized by the age of the patient at the time of the visit.

Characterization of diagnoses, medications, and laboratory test values. 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 10 diagnoses (**Supplementary Table S1** online) were queried for the number of unique individuals <18 years of age for whom there was at least one instance of a qualifying code for each diagnosis; samples of individuals with documentation of multiple diagnoses were also eligible for inclusion. To characterize medication exposures, the authors chose 14 medications representing a range of indications, routes, and frequencies of use in children. This list of drugs was compiled from mentions captured in MedEx, a validated bioinformatics tool developed at our institution for extraction of medication data from clinical narratives.[21] The number of unique individuals with medication mentions was determined, including those with documentation of multiple medications. Ten common tests or panels assayed in samples of blood, urine, or cerebrospinal fluid were also 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 synthetic derivative cohort and BioVU subset, based on their entire de-identified medical records. 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 is linked to the online version of the paper at http://www.nature.com/cpt

ACKNOWLEDGMENTS

The authors acknowledge 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. T.L.M., S.L.V.D., K.B.B., E.A.B., L.J.M., and D.M.R.: Portions of this study were supported by grant 5U01HG004603, which supported the Vanderbilt site in the National Institutes of Health (NIH)/National Human Genome Research Institute Electronic Medical Records and Genomics (eMERGE) network; the Vanderbilt Institute for Clinical and Translational Research; National Center for Research Resources/NIH grant UL1 RR024975; NIH/National Institute of Environmental Health Sciences grant K12 ES015855 (T.L.M.); and NIH/National Institute of General Medical Sciences Clinical Pharmacology Training Program 5T32 GM007569-33 (S.L.V.D.). Funds for the 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").

AUTHOR CONTRIBUTIONS

S.L.V.D., T.L.M., and K.B.B. wrote the manuscript. S.L.V.D., T.L.M., K.B.B., E.A.B., L.J.M., and D.M.R. designed the research. S.L.V.D., T.L.M., K.B.B., and E.A.B. performed the research. S.L.V.D., T.L.M., and K.B.B. analyzed the data.

CONFLICT OF INTEREST

The authors declared no conflict of interest.

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 several factors, including limited accrual and the lack of existing approaches to accommodate specific considerations relating to preserving samples from pediatric patients.

WHAT QUESTION DID THIS STUDY ADDRESS?

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 including samples from pediatric patients in BioVU provides a feasible model for collecting samples and medical record data, with a 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 samples that represent a diverse range of diseases, medication exposures, and laboratory assessments. These will facilitate research in pediatric clinical pharmacology, including investigations of pharmacogenomics, pharmacoepidemiology, and adverse drug events.

1. US Food and Drug Administration. Table of Pharmacogenomic Biomarkers in Drug Labels <http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm> (2012). Accessed 2 October 2012.
2. Ginsburg, G.S., Burke, T.W. & Febbo, P. Centralized biorepositories for genetic and genomic research. *JAMA* **299**, 1359–1361 (2008).
3. McCarty, C.A. *et al.* The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genomics* **4**, 13 (2011).
4. Brisson, A.R., Matsui, D., Rieder, M.J. & Fraser, D.D. Translational research in pediatrics: tissue sampling and biobanking. *Pediatrics* **129**, 153–162 (2012).
5. Roden, D.M. *et al.* Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharmacol. Ther.* **84**, 362–369 (2008).
6. Brothers, K.B. & Clayton, E.W. Parental Perspectives on a Pediatric Human Non-Subjects Biobank. *AJOB Prim. Res.* **3**, 21–29 (2012).
7. Pulley, J., Clayton, E., Bernard, G.R., Roden, D.M. & Masys, D.R. Principles of human subjects protections applied in an opt-out, de-identified biobank. *Clin. Transl. Sci.* **3**, 42–48 (2010).
8. OHRP. *Guidance on Research Involving Coded Private Information or Biological Specimens* (Office of Human Research Protections, Rockville, MD, 2008).
9. Brothers, K.B. & Clayton, E.W. "Human non-subjects research": privacy and compliance. *Am. J. Bioeth.* **10**, 15–17 (2010).
10. Brothers, K.B., Morrison, D.R. & Clayton, E.W. Two large-scale surveys on community attitudes toward an opt-out biobank. *Am. J. Med. Genet. A* **155A**, 2982–2990 (2011).
11. Pulley, J.M., Brace, M.M., Bernard, G.R. & Masys, D.R. Attitudes and perceptions of patients towards methods of establishing a DNA biobank. *Cell Tissue Bank.* **9**, 55–65 (2008).
12. Hens, K., Nys, H., Cassiman, J.J. & Dierickx, K. The storage and use of biological tissue samples from minors for research: a focus group study. *Public Health Genomics* **14**, 68–76 (2011).
13. Jenkins, M.M. *et al.* Maternal attitudes toward DNA collection for gene-environment studies: a qualitative research study. *Am. J. Med. Genet. A* **149A**, 2378–2386 (2009).
14. Brothers, K.B. Biobanking in pediatrics: the human nonsubjects approach. *Per. Med.* **8**, 79 (2011).
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.* **19**, 1115–1118 (2012).
16. Hens, K., Nys, H., Cassiman, J.J. & Dierickx, K. Genetic research on stored tissue samples from minors: a systematic review of the ethical literature. *Am. J. Med. Genet. A* **149A**, 2346–2358 (2009).
17. Hens, K., Nys, H., Cassiman, J.J. & Dierickx, K. The return of individual research findings in paediatric genetic research. *J. Med. Ethics* **37**, 179–183 (2011).
18. Tabor, H.K. *et al.* Parent perspectives on pediatric genetic research and implications for genotype-driven research recruitment. *J. Empir. Res. Hum. Res. Ethics* **6**, 41–52 (2011).
19. Roden, D.M., Xu, H., Denny, J.C. & Wilke, R.A. Electronic medical records as a tool in clinical pharmacology: opportunities and challenges. *Clin. Pharmacol. Ther.* **91**, 1083–1086 (2012).
20. Dumitrescu, L. *et al.* Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. *Genet. Med.* **12**, 648–650 (2010).
21. Xu, H., Stenner, S.P., Doan, S., Johnson, K.B., Waitman, L.R. & Denny, J.C. MedEx: a medication information extraction system for clinical narratives. *J. Am. Med. Inform. Assoc.* **17**, 19–24 (2010).