Ethical and practical challenges to studying patients who opt out of large-scale biorepository research

S Trent Rosenbloom,1,2,3 Jennifer L Madison,4 Kyle B Brothers,5,6 Erica A Bowton,4 Ellen Wright Clayton,2,6 Bradley A Malin,1,6,7 Dan M Roden,3,4,8 Jill Pulley4,8

ABSTRACT

Large-scale biorepositories that couple biologic specimens with electronic health records containing documentation of phenotypic expression can accelerate scientific research and discovery. However, differences between those subjects who participate in biorepository-based research and the population from which they are drawn may influence research validity. While an opt-out approach to biorepository-based research enhances inclusiveness, empirical research evaluating voluntariness, risk, and the feasibility of an opt-out approach is sparse, and factors influencing patients’ decisions to opt out are understudied. Determining why patients choose to opt out may help to improve voluntariness, however there may be ethical and logistical challenges to studying those who opt out. In this perspective paper, the authors explore what is known about research based on the opt-out model, describe a large-scale biorepository that leverages the opt-out model, and review specific ethical and logistical challenges to bridging the research gaps that remain.

BACKGROUND

Leveraging knowledge about the influence that genetic variants can have on disease prevention, diagnosis, and response to medications has emerged as a promising approach to improving and ‘personalizing’ the medical treatments that patients receive.1–3 The rapid evolution of technologies capable of acquiring and analyzing high-dimensionality data now means that this approach can be extended, making it a centerpiece of an era of personalized medicine.4 5 Identifying meaningful relationships among high dimensional data sources requires very large sample sets for discovery and validation.6–10 Biorepositories meet this need by linking large-scale collections of DNA and other biologic specimens with related electronic health records (EHR) containing documentation of phenotypic expression. Biorepositories make it possible to reuse large sample sets repeatedly to answer multiple research questions, with the overall utility of biorepositories determined primarily by their size and composition.11 A major factor influencing investigators’ ability to create large-scale biorepositories is the need to ensure that subjects consent to participation. In the US regulatory setting, written informed consent is considered the gold standard for ensuring that research participation is fully intentional and voluntary.12 13 Despite this, some have argued that truly informed consent is impractical for large-scale biorepositories,12 14–17 and that obtaining informed consent from patients can be costly and time-consuming, and can introduce bias into genomic studies.18–23 To enable rapid, cost-efficient collection of biosamples from a diverse set of patients, investigators have adopted an opt-out model of patient inclusion as an alternative to full informed consent for biorepository inclusion.24

THE OPT-OUT MODEL FOR RESEARCH PARTICIPATION

With the opt-out approach, biologic samples obtained during routine clinical investigations are candidates for biorepository inclusion unless the individual patient proactively opts out.25–28 This approach overcomes the high cost and effort that would be required if investigators obtained informed consent for every potential subject when creating a large-scale biorepository. For example, according to published studies, the estimated cost to obtain informed consent was $248 per patient, equivalent to the cost of 2–7.5 full-time employees per study.22 23 29 Notifying patients and providing the opportunity to opt out of the biorepository is accepted as an efficient alternative.6 25–28 30–33 Furthermore, the opt-out approach may eliminate the potential bias that can result from large consented genetic studies, such as imbalances based on age, sex, race, income, education, and health status.18 20 21 34–37

The Vanderbilt University Medical Center (VUMC) biorepository, called BioVU, was developed to support genotype-phenotype research. BioVU combines an opt-out approach with comprehensive medical record de-identification and intensive oversight from the VUMC Institutional Review Board (IRB), internal and external ethics boards, and a community advisory board. The Federal Office of Human Research Protections (OHRP) previously affirmed that research using biosamples and de-identified health records qualifies as non-human subjects research under the Common Rule.38 Although there is no regulatory requirement for notification or consent under this designation, patients at VUMC can opt out from having their blood sample included in BioVU24 at the time they provide consent for medical treatment during standard patient intake processes. In BioVU, once a patient opts out, his or her sample is permanently excluded. The decision to allow potential subjects to opt out was based on ethical considerations and research demonstrating that patients were more likely to support such a biorepository when an opt-out mechanism was available.26 27 32 39 Additional public notification mechanisms increase awareness of the project, and ensure that patients are aware of the opportunity to opt out of participation.26 30
BioVU’s cumulative opt-out rate of 15% over the past 5 years indicates that large numbers of patients have recognized and exercised their right to opt out.26 40 As of April 15, 2013, BioVU contained DNA samples from over 163 000 individuals, and to our knowledge is the largest operating opt-out biorepository in the USA.24 The size and diversity of BioVU has supported a multitude of genomic studies, including studies conducted within the National Human Genome Research Institute (NHGRI) eMERGE consortium.24 41–45

**THE NEED TO STUDY THOSE WHO OPT OUT**

Differences between patients who do and do not opt out of biorepository-based research may influence the validity of any results from research leveraging the biorepository. Research evaluating whether the opt-out approach leads to subject skew relative to the non-enrolled population is lacking. Three general approaches can be considered when differences between those patients who opt out of biorepository-based research and those who do not are being measured. These include: (1) directly comparing differences between the two groups using existing de-identified medical record data; (2) assessing the potential sociocultural determinants of opting out by enriching EHR data with other potential correlates of culture and belief systems; and (3) using surveys or qualitative methods to study patients who opt out of biorepository-based research. These methods complement each other, yet can raise the specific ethical questions addressed below. In particular, a large-scale, population-based study can complement (and be complemented by) approaches that directly elicit patient perspectives when reasons for opting out of research are being studied. Both approaches raise ethical considerations in this setting.

**Comparisons using existing de-identified medical record data**

In the first approach we describe, investigators could compare the attributes of populations who do and do not opt out from a de-identified EHR to identify factors associated with whether or not patients opt out of biorepository-based research. It is important to note that while patients may opt out of having their biospecimens stored in the biorepository, their de-identified health records may still be available for non-biospecimen research as non-human subjects research and without any regulatory requirement for obtaining informed consent. This distinction would allow investigators to study differences between these two cohorts using non-biospecimen EHR data alone. Standard regression methods can be applied to compare differences between these two groups on the basis of basic patient demographics, patient health indicators, and clinical results. Demographic variables (eg, age, gender, and race) and health related variables (eg, annual rate of inpatient visits, outpatient visits, procedure and diagnostic codes, opportunities for opting out, and length in time of the patient record) can be obtained directly from most EHR systems or from a biorepository.

**Ethical considerations**

While there is no specific regulatory barrier, research to study individuals’ decisions around research participation, such as the choice to opt out, poses special ethical challenges because it requires investigators to study individuals who have chosen not to be included in at least some forms of research. The responsible conduct of research in innovative areas and developing policies to guide such conduct require empirical data on participants’ preferences, perceptions, and behaviors, including those who have declined participation. Some experts have questioned whether it is ethical to study patients who have opted out of being included in one type of research. However, we propose that comparing patients who have opted out of a biorepository with those who have not using de-identified, aggregate data is consistent with ethical and regulatory norms in this field, both in the USA and worldwide. For example, published studies exist that report on subjects’ willingness to participate in a second research study,46 to have biosamples collected for secondary genetic research,47 48 and to allow their research data to be linked with medical record or health insurance data.46 49–54 In addition, medical record and national health registry data have been used to report on the characteristics of patients who decline participation in research or fail to respond to an invitation to participate in research, and have highlighted the potential effects of these differences on scientific investigation.55–60

**Enriching medical record data with socio-cultural variables**

In the second approach we describe, investigators could compare differences among populations by incorporating additional relevant attributes not typically contained in medical records, including socio-economic indicators. This may be especially useful for research evaluating subjects’ beliefs about research participation. The lack of rich indicators of socio-economic status recorded within EHRs can limit epidemiological research.61 Investigators have addressed this problem by characterizing the mean socio-economic variables based on geospatial information about the community in which subjects reside.62 63 Correlates of socio-economic status could be estimated, for example, from the average income in the census tract where a subject’s residence is located.61 This method has been repeatedly validated as a reasonable estimate of socio-economic status,64 65 and has been shown to be an effective method for estimating the impact of socio-economic status on health behaviors and outcomes.66–69 A wide range of health discoveries across disciplines has applied this method, including a number published in high impact journals.63 70–72 Using geospatial variables, investigators could enrich de-identified biorepository data with correlates of patient geospatial information, including average race, ethnicity, education, and religious practices.

**Ethical considerations**

Geospatial information added to a de-identified resource could potentially increase the re-identification risk for the corresponding patients. This concern is strengthened by various investigations that have demonstrated how de-identified health information can be re-identified.73–80 However, there is a significant difference between the description of a path by which health information could be re-identified and the likelihood that such a path would be leveraged by an adversary in the real world.81 While geocodes may facilitate patient re-identification, they are not necessarily explicit identifiers in their own right. As such, geocodes may be retained in de-identified data according to the HIPAA Privacy Rule’s Expert Determination standard provided the risk to identification is deemed to be small. Various statistical and computational methodologies have been developed to disclose geocoded health information in a manner that respects privacy. Our prior work has demonstrated that algorithmic strategies can be constructed and executed to rescind pseudonymized ZIP codes82 if the sample has a relatively similar distribution to the public data and we suspect that similar attacks would permit identification of census tracts.
Directly studying patients who have opted out

In the third approach we describe, investigators could directly study those who have opted out using surveys or other qualitative methods to characterize preferences and perspectives surrounding their choice to opt out of research participation. Numerous established methods—including surveys, focus groups, and interviews—can elucidate these perspectives and preferences. \(^{31–33, 48, 83–90}\) Surveys of patients by mail, email, or phone are certainly reasonable ways to gather data on patients’ perspectives. In-person exit interviews would also be suitable, and could have the added benefit of taking place during clinical encounters, which is the pertinent time frame for ‘top-of-mind’ responses. These approaches require that investigators directly contact (or re-contact) potential subjects.

Ethical considerations

Using surveys or qualitative methods to evaluate the perspectives of patients who have opted out of biorepository research participation may be constrained in two ways, one ethical and one statistical. First, those who opt out may not agree to be contacted by researchers in any capacity. To identify such persons and invite them to participate in research on this basis is ethically problematic, and unlikely to be received positively by patients. Second, those who opt out of a biorepository may also be more likely to decline participation in other forms of research, thus raising a concern about participation bias. For these two reasons, direct research with patients about their perspectives on opting out might utilize recruitment methods that (1) give no consideration to documented opt-out decisions and (2) are designed to minimize participation bias.

DISCUSSION

While obtaining written informed consent may be ideal for ensuring that research participation is fully intentional and voluntary,\(^ {12, 13}\) this requirement may hinder the development and growth of large-scale biorepositories. Rapid and unbiased accrual of patient samples is critical for supporting large-scale biorepository-based research, and can be achieved with an opt-out model. However, a major barrier to adopting the opt-out model is uncertainty surrounding the ethical and policy implications of implementing such an approach. At the crux of this uncertainty is whether the opt-out approach is effective for enabling patients who do not want their samples to be included in research to exercise this preference. Research evaluating those who opt out will be critical to determine the effectiveness of an opt-out approach, and is timely given current national attention to methods of consent for biosample research. However, researchers studying the differences between those who do and do not opt out of biosample-based research have access to several complementary methods, each with their own ethical and pragmatic challenges. Addressing these challenges is critical; recently, the OHRP and Department of Health and Human Services (HHS) have distributed a document seeking input on a far-reaching revision of regulations related to human research protections. In this Advance Notice of Proposed Rule-Making (ANPRM) document, HHS has explicitly asked for input on the suitability of opt-out approaches for biosample research.\(^ {91, 92}\) Specifically, research measuring the characteristics of those who choose to opt out of biorepositories has the potential to provide valuable information about the generalizability of research using opt-out models, and may help inform the OHRP decision on the suitability of the notification and opt-out model. Additionally, this information will allow existing biorepositories with similar models to improve patient education and ultimately increase awareness of such programs.

Contributors All authors provided conceptual input; STR, JLM, KBB, and BAM drafted segments of the article; and all authors provided critical revision of the manuscript for important intellectual content.

Funding The project was supported, in part, by grants from the US National Library of Medicine (R01 LM009989), the National Center for Advancing Translational Sciences (UL1 TR000445), and the National Human Genome Research Institute (R01 HG006844, U01 HG006378).

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

51 Tate AR, Calderwood L, Dezateux C, Harris T, Cook DG, Victor C, Young AF, Dobson AJ, Byles JE. Health services research using linked records: who...


38 OHRP. Guidelines on research involving coded private information or biological specimens. Rockville, MD: Office of Human Research Protections, 2008.


59 Krieger N. Women and social class: a methodological study comparing individual, household, and census measures as predictors of black/white differences in reproductive history. J Epidemiol Community Health 1991;45:35–42.


