# Public Attitudes toward Consent and Data Sharing in Biobank Research: A Large Multi-site Experimental Survey in the US

Saskia C. Sanderson,[1,2,3,27,*] Kyle B. Brothers,[4,27,*] Nathaniel D. Mercaldo,[5] Ellen Wright Clayton,[6] Armand H. Matheny Antommaria,[7] Sharon A. Aufox,[8] Murray H. Brilliant,[9] Diego Campos,[10] David S. Carrell,[11] John Connolly,[12] Pat Conway,[13] Stephanie M. Fullerton,[14] Nanibaa' A. Garrison,[15,26] Carol R. Horowitz,[16] Gail P. Jarvik,[17] David Kaufman,[18] Terrie E. Kitchner,[9] Rongling Li,[19] Evette J. Ludman,[11] Catherine A. McCarty,[13] Jennifer B. McCormick,[20] Valerie D. McManus,[21] Melanie F. Myers,[22] Aaron Scrol,[11] Janet L. Williams,[23] Martha J. Shrubsole,[24] Jonathan S. Schildcrout,[5] Maureen E. Smith,[8] and Ingrid A. Holm[25]

Individuals participating in biobanks and other large research projects are increasingly asked to provide broad consent for open-ended research use and widespread sharing of their biosamples and data. We assessed willingness to participate in a biobank using different consent and data sharing models, hypothesizing that willingness would be higher under more restrictive scenarios. Perceived benefits, concerns, and information needs were also assessed. In this experimental survey, individuals from 11 US healthcare systems in the Electronic Medical Records and Genomics (eMERGE) Network were randomly allocated to one of three hypothetical scenarios: tiered consent and controlled data sharing; broad consent and controlled data sharing; or broad consent and open data sharing. Of 82,328 eligible individuals, exactly 13,000 (15.8%) completed the survey. Overall, 66% (95% CI: 63%–69%) of population-weighted respondents stated they would be willing to participate in a biobank; willingness and attitudes did not differ between respondents in the three scenarios. Willingness to participate was associated with self-identified white race, higher educational attainment, lower religiosity, perceiving more research benefits, fewer concerns, and fewer information needs. Most (86%, CI: 84%–87%) participants would want to know what would happen if a researcher misused their health information; fewer (51%, CI: 47%–55%) would worry about their privacy. The concern that the use of broad consent and open data sharing could adversely affect participant recruitment is not supported by these findings. Addressing potential participants' concerns and information needs and building trust and relationships with communities may increase acceptance of broad consent and wide data sharing in biobank research.

## Introduction

Health research increasingly requires very large numbers of participants to be willing to share their biological samples, genomic data, and clinical information with researchers.[1] The proposed Precision Medicine Initiative (PMI) is one prominent example of a very large biobank created to improve understanding of human health and disease.[2,3] In this type of biobank, participants are asked to share their data not only with the institutions making the request but also with other investigators at diverse sites, often for projects not yet conceived.[4] The National Institutes of Health (NIH) Genomic Data Sharing Policy already requires that NIH-funded researchers proposing to generate genomic data obtain broad consent from participants, that is, consent that permits wide data sharing.[5]

Obtaining broad consent has also been proposed by the Department of Health and Human Services and the White House Office of Science and Technology Policy as the preferred approach. Their 2015 *Notice of Proposed Rulemaking (NPRM)* required that informed consent be obtained for all research using tissue samples and most clinical data and

[1]Department of Behavioural Science and Health, University College London, London WC1E 6BT, UK; [2]Great Ormond Street Hospital, London WC1N 3JH, UK; [3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [4]Department of Pediatrics, University of Louisville, Louisville, KY 40202, USA; [5]Department of Biostatistics, Vanderbilt University, Nashville, TN 37203, USA; [6]Center for Biomedical Ethics and Society, Vanderbilt University, Nashville, TN 37203, USA; [7]Ethics Center, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA; [8]Center for Genetic Medicine, Northwestern University, Chicago, IL 60611, USA; [9]Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA; [10]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [11]Group Health Research Institute, Seattle, WA 98101, USA; [12]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [13]Essentia Institute of Rural Health, Duluth, MN 55805, USA; [14]Department of Bioethics and Humanities, University of Washington, Seattle, WA 98195, USA; [15]Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA 98101, USA; [16]Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [17]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [18]Division of Genomics and Society, National Human Genome Research Institute, Bethesda, MD 20892, USA; [19]Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, MD 20892, USA; [20]Biomedical Ethics Program, Mayo Clinic, Rochester, MN 55905, USA; [21]Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI 54449, USA; [22]Genetic Counseling Graduate Program, Cincinnati Children's Hospital Medical Center and University of Cincinnati, Cincinnati, OH 45229, USA; [23]Genomic Medicine Institute, Geisinger Health System, Danville, PA 17822, USA; [24]Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN 37203, USA; [25]Division of Genetics and Genomics and the Manton Center for Orphan Diseases Research, Boston Children's Hospital, Boston, MA 02115, USA; [26]Department of Pediatrics, Division of Bioethics, University of Washington, Seattle, WA 98101, USA
[27]These authors contributed equally to this work
*Correspondence: saskia.sanderson@ucl.ac.uk (S.C.S.), kyle.brothers@louisville.edu (K.B.B.)
http://dx.doi.org/10.1016/j.ajhg.2017.01.021.

stated that such consent can be obtained in a one-time, open-ended or "broad" fashion.[6]

Obtaining broad consent from participants in biobanks reduces administrative burdens and may accelerate discovery.[5] However, broad consent and wide data sharing reduces participants' control over how their data is used.[7,8] Individuals who object to decreased control may be less willing to take part in the research.[9–13]

Decreased willingness to participate in research is of particular concern for populations who are already underrepresented in medical research. Most genomics research to date has used data predominantly from individuals of Northern European ancestry, limiting the insights gained for individuals of other ancestries.[14–18] Ancestry is only imperfectly correlated with the social constructs of race and ethnicity, but this association does mean that perspectives on research participation within underserved groups can influence inclusion from underrepresented ancestry groups. Some racial and ethnic groups, including African Americans, have less trust in medical researchers than others,[19,20] so the move toward broad consent could undermine much-needed efforts to increase participation in medical research among underrepresented groups.

Members of the public are key stakeholders in this discussion.[21] A recent systematic literature review[22] reported that, when presented with different consent models and asked to choose among them, individuals often favor greater levels of control and select more specific types of consent over broad consent.[23–25] These findings could have serious implications for large-scale research efforts. Asking respondents to choose from among several consent or data sharing options is quite different, however, from asking them to enroll in a particular biobank with a defined data sharing policy. In addition, it is unclear whether respondents to these previous surveys truly understood the trade-offs involved, or whether they were simply endorsing the idea of greater choice more generally. As noted in the same review, previous studies have also tended to be small and local and to underrepresent key demographic groups.[22]

We therefore conducted a large survey of attitudes toward consent and data sharing in biobank research among diverse participants recruited at multiple healthcare systems participating in the Electronic Medical Records and Genomics (eMERGE) Network.[26] The eMERGE Network is ideal for this type of research, as it has the necessary infrastructure and access to a large diverse population of individuals who are among those most likely to be invited to participate in a biobank. The overarching aim of the present study was to examine patients' attitudes toward participating in biobank research using an experimental study design that randomly assigned participants to different consent and data sharing conditions.

We hypothesized that participants randomly assigned to a hypothetical biobank with broad consent and open data sharing would express less willingness to participate in, and have more negative attitudes toward, the biobank than those assigned to a biobank with tiered consent and controlled data sharing. We also hypothesized that willingness would be lower among participants of lower socioeconomic status and from underserved racial and ethnic groups. In addition, we examined participants' perceived benefits, concerns, and information needs regarding participating in biobank research.

## Subjects and Methods

### Study Design and Procedures

This was an experimental survey study. Individuals were randomly assigned to one of three conditions and completed self-report questionnaires. Participants were patients who had: (1) sought care for themselves or their minor child at one of 11 eMERGE Network sites[26] between October 1, 2013 and September 1, 2014; (2) a valid address that could be geocoded; and (3) age and sex available in the electronic health record. To maximize diversity of the sample, and specifically to enrich the observed sample with demographic groups that have been underrepresented in previous studies, we utilized a disproportionate stratified sampling scheme. Strata were defined by the cross-classification of patient age group (at adult centers: $\leq$ 35 years and $>$ 35 years; pediatric centers: $\leq$ 12 years and $>$ 12 years), sex, race (white, black, Asian, Native American/Alaska Native, Hawaiian/Pacific Islander, other), ethnicity (Hispanic/Latino or not), education group ($<$ 12 years, 12–15 years, 16+ years), and residence in a rural versus urban/suburban census block group. To be able to execute this complex study design, values for each stratification variable needed to be known (or estimated) prior to randomization. Many of these items, as well as home address, were readily available in the EHR. Others were incomplete or were inaccurate. Home addresses were geocoded and linked to unique census block groups. Information from the 2010 census and 2008–2012 American Community Survey were used to impute missing stratification information, using the mode of the participant's census block group value, in order to define the "approximate" sampling frame. Further details on the sampling frame and the stratified sampling strategy are provided in Tables S1 and S2. A thorough exposition of the sampling scheme, including how census data were utilized for the purpose of sampling, will be provided in a separate article. The sampling frame was approximately 2.4 million patients, of whom 90,000 were selected.

Participants within each sampling stratum were randomized to receive a survey which included one of three hypothetical biobank scenarios. The scenarios were identical except for the details regarding consent type and data sharing approach.[27] In the first scenario, donated samples and data could be used for all kinds of medical research and data could be shared with approved investigators only ("broad-controlled"). The second and third scenarios contained an alternative consent approach or data sharing policy: in the "tiered-controlled" scenario, the consent process allowed participants to select the types of research for which their samples and data could be used, and in the "broad-open" scenario, the data sharing policy allowed de-identified data to be shared through an online database open to the public.

Pre-notification postcards were mailed, after which optical scan surveys were mailed along with a non-contingent pre-incentive $2 bill, to potential participants in April 2015. Non-respondents received a reminder letter in May 2015 and a second survey in

July 2015. Participants could complete the survey on paper and return it in a self-addressed stamped envelope or complete an identical survey through a secure, online survey interface on the REDCap database platform.[28] This project received IRB approval at all 11 participating sites.

## Development of the Survey Instrument

The survey instrument, including the three hypothetical scenarios and the survey questions, was developed by a multidisciplinary expert working group, informed by the findings from a systematic review of the literature.[22] A complete draft of the survey instrument content was tested and refined using cognitive interviews with 40 patients across 6 sites. A pilot study to test the feasibility of the survey instrument and study procedures was then conducted across all 11 sites. In the pilot study, 166 respondents returned the survey, out of 1,500 patients who were sent the survey (response rate 11%). Analysis of pilot data suggested that planned study procedures were robust, and quality indicators including incomplete surveys and straight-lined responses necessitated only minor revisions to the survey. Respondents to the pilot survey were excluded from the main study, and responses to the pilot were not included in the main analysis.

## Survey Measures

### *Demographics*

Standard measures of demographic characteristics were used (see CDC Questionnaire in Web Resources).[29] Poverty was calculated from income and number of people in household. Rurality was assessed using census-level data. Religiosity was assessed using an item adapted from previous research.[30] Self-rated health, an indicator of quality of life, was assessed using the widely-used single item from the SF-12.[31]

### *Trust and Privacy*

Concern about privacy was assessed using two items.[32] Trust was assessed with two items.[33,34]

### *Willingness to Participate*

Willingness to participate in the hypothetical biobank described was assessed using a single item adapted from previous research.[25,35]

### *Attitudes toward Participating*

Attitude items were either generated specifically for this study or adapted from previous research.[24,25,36–39] In order to generate these items, the multidisciplinary expert working group defined three relevant sub-domains to be assessed within the overarching domain of "attitudes towards participating in a biobank:" perceived benefits of participating in the described biobank, concerns about participating in the described biobank, and information needs about the governance of the described biobank (e.g., how decisions are made regarding the use of the samples and data). Initial lists of items to assess each of these sub-domains were compiled based on a review of the literature on these topics[22] and on expert input. The lists were culled in an iterative manner, in order to produce a manageable number of prioritized items that would not over-burden participants. The final list comprised five items assessing perceived benefits, six items assessing concerns, and eight items assessing information needs. Likert-style scales were used with five responses ranging from "strongly disagree" to "strongly agree" for each item. Factor analysis confirmed that benefits, concerns, and information needs were distinct factors, with all items from each set loading on that factor with eigenvalues greater than 0.4. In order to describe responses to these items, responses were dichotomized and proportions responding "agree" or "strongly agree" with each statement were reported. In addition, composite scale scores were created by calculating the mean of each set of items (possible range from 1 to 5). Mean scores were described for each attitudinal scale. In addition, for the purposes of the regression analysis (see below), participant's scale scores were categorized: scale scores ranging from 1.0 to 2.50 were categorized as "low," scale scores ranging from 2.51 to 3.50 were categorized as "intermediate," and scale scores ranging from 3.51 to 5.0 were categorized as "high." The three survey instruments are available online (see Web Resources).

## Data Analysis

Response rates were calculated according to American Association for Public Opinion Research (AAPOR) criteria.[40] To determine whether randomization was maintained within the subset of respondents, sample counts and percentages were calculated for socio-demographic variables. These summaries were computed within consent and data sharing models across all sites and compared using the Pearson's chi-square test. For all other analyses, each participant was assigned a post-stratified, sampling weight (i.e., the inverse of the probability of being sampled and answering the survey) to account for the stratified sampling design. Because understudied populations were intentionally oversampled, sampling weights varied dramatically within and across sites, and so we conducted site-specific weight trimming and redistribution combining two commonly used approaches: weights were trimmed at (1) the 90$^{th}$ percentile of weights, or (2) the median+6*IQR, whichever was higher.[41,42] Recognizing that all trimming approaches are ad hoc, we conducted sensitivity analyses for the primary analysis using a number of approaches to trimming. Results from sensitivity analyses can be found in Table S3.

The impact of consent and data sharing models on willingness to participate in biobank research was estimated with a (survey weighted) three-level multinomial logistic regression (probably not/definitely not; not sure; yes probably/yes definitely) with linearized covariance estimates of uncertainty. For ease of exposition, all combined estimates (across multiply imputated datasets within a site and then across sites) were transformed and are reported as probabilities (or percentages). Comparisons among data sharing and consent models were performed via a Wald test from an ordinal logistic regression, proportional odds model. Similar methods were applied to estimate the impact of consent and data sharing models on each of the three attitudinal constructs (perceived benefits, concerns, and information needs).

To identify demographic and other characteristics associated with willingness to participate in biobank research, willingness was dichotomized (yes = agree/strongly agree; no = not sure/disagree/strongly disagree) and was regressed on covariates using unadjusted (marginal) and adjusted logistic regression analyses. Adjusted models were fitted hierarchically using socio-demographic variables first; adding trust and privacy items second; and adding the attitudinal constructs third. Unadjusted estimates were summarized with percentages, and adjusted estimates of covariates associations with willingness to participate were summarized with odds ratios and associated 95% confidence intervals. Sub-domain analyses were performed to quantify the extent to which the relationship between survey type and willingness differed across socio-demographic variables. There was little to

**Table 1. Sociodemographic and Health Characteristics of Survey Respondents by Consent and Data Sharing Model and across Models**

| | Broad-Controlled | | Broad-Open | | Tiered-Controlled | | Overall | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| Total | 4,405 | 34% | 4,371 | 34% | 4,224 | 32% | 13,000 | 100% |
| **Sex** | | | | | | | | |
| Female | 2,734 | 63% | 2,739 | 64% | 2,601 | 63% | 8,074 | 63% |
| Male | 1,583 | 37% | 1,561 | 36% | 1,557 | 37% | 4,701 | 37% |
| **Age** | | | | | | | | |
| 18–35 years | 1,058 | 25% | 1,023 | 24% | 1,030 | 25% | 3,111 | 25% |
| 36–50 years | 1,364 | 32% | 1,339 | 32% | 1,317 | 32% | 4,020 | 32% |
| 51–64 years | 1,019 | 24% | 1,042 | 25% | 942 | 23% | 3,003 | 24% |
| 65+ years | 823 | 19% | 822 | 19% | 790 | 19% | 2,435 | 19% |
| **Race** | | | | | | | | |
| White | 2,197 | 51% | 2,202 | 52% | 2,122 | 52% | 6,521 | 51% |
| Asian | 746 | 17% | 718 | 17% | 742 | 18% | 2,206 | 17% |
| Black or African American | 506 | 12% | 501 | 12% | 476 | 12% | 1,483 | 12% |
| Other | 438 | 10% | 405 | 9% | 385 | 9% | 1,228 | 10% |
| American Indian or Alaska Native | 223 | 5% | 245 | 6% | 221 | 5% | 689 | 5% |
| More than one race | 148 | 3% | 155 | 4% | 119 | 3% | 422 | 3% |
| Native Hawaiian or Pacific Islander | 45 | 1% | 49 | 1% | 47 | 1% | 141 | 1% |
| **Ethnicity** | | | | | | | | |
| Latino | 785 | 18% | 758 | 18% | 725 | 18% | 2,268 | 18% |
| Not Hispanic/Latino | 3,507 | 82% | 3,476 | 82% | 3,365 | 82% | 10,348 | 82% |
| **Educational Attainment** | | | | | | | | |
| Up to some high school (grades 9–12) | 328 | 8% | 279 | 7% | 282 | 7% | 889 | 7% |
| High school graduate or GED | 459 | 11% | 465 | 11% | 471 | 12% | 1,395 | 11% |
| Some college | 1,026 | 24% | 1,060 | 25% | 965 | 24% | 3,051 | 24% |
| Bachelors degree or equivalent | 1,180 | 28% | 1,166 | 28% | 1,132 | 28% | 3,478 | 28% |
| Masters degree or equivalent | 767 | 18% | 778 | 19% | 772 | 19% | 2,317 | 19% |
| PhD, MD, JD, or equivalent | 452 | 11% | 438 | 10% | 443 | 11% | 1,333 | 11% |
| **Annual Household Income** | | | | | | | | |
| Less than $30,000 | 902 | 22% | 915 | 23% | 889 | 23% | 2,706 | 23% |
| $30,000–$60,000 | 838 | 21% | 847 | 21% | 778 | 20% | 2,463 | 21% |
| $60,000–$90,000 | 599 | 15% | 644 | 16% | 650 | 17% | 1,893 | 16% |
| $90,000–$150,000 | 852 | 21% | 781 | 19% | 785 | 20% | 2,418 | 20% |
| More than $150,000 | 864 | 21% | 866 | 21% | 804 | 21% | 2,534 | 21% |
| **Total Number of People in Household** | | | | | | | | |
| 1 | 551 | 13% | 584 | 14% | 498 | 12% | 1,633 | 13% |
| 2 | 1,281 | 30% | 1,316 | 31% | 1,260 | 31% | 3,857 | 30% |

(*Continued on next page*)

**Table 1.** ***Continued***

| | Broad-Controlled | | Broad-Open | | Tiered-Controlled | | Overall | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| 3 | 826 | 19% | 805 | 19% | 758 | 18% | 2,389 | 19% |
| 4 or more | 1,623 | 38% | 1,550 | 36% | 1,604 | 39% | 4,777 | 38% |
| **Poverty**[a] | | | | | | | | |
| Below the poverty line | 586 | 15% | 615 | 15% | 590 | 15% | 1,791 | 15% |
| Not below the poverty line | 3,422 | 85% | 3,391 | 85% | 3,273 | 85% | 10,086 | 85% |
| **Work Situation** | | | | | | | | |
| Working | 2,313 | 53% | 2,288 | 53% | 2,233 | 54% | 6,834 | 53% |
| Retired | 796 | 18% | 805 | 19% | 785 | 19% | 2,386 | 19% |
| Disabled or unemployed | 496 | 11% | 493 | 11% | 471 | 11% | 1,460 | 11% |
| Other | 741 | 17% | 725 | 17% | 679 | 16% | 2,145 | 17% |
| **Healthcare Insurance** | | | | | | | | |
| Private insurance | 3,061 | 71% | 3,011 | 71% | 2,948 | 71% | 9,020 | 71% |
| Public insurance | 1,051 | 24% | 1,065 | 25% | 1,003 | 24% | 3,119 | 25% |
| Other type of insurance | 102 | 2% | 98 | 2% | 91 | 2% | 291 | 2% |
| No insurance | 87 | 2% | 82 | 2% | 94 | 2% | 263 | 2% |
| **Rurality (from Census-level Data)** | | | | | | | | |
| Suburban or urban | 2,549 | 58% | 2,530 | 58% | 2,427 | 57% | 7,506 | 58% |
| Rural | 1,856 | 42% | 1,841 | 42% | 1,797 | 43% | 5,494 | 42% |
| **Marital Status** | | | | | | | | |
| Married, living with someone | 2,696 | 63% | 2,701 | 64% | 2,630 | 64% | 8,027 | 64% |
| Not married, living with someone | 352 | 8% | 333 | 8% | 350 | 9% | 1,035 | 8% |
| Not married, not living with someone | 1,201 | 28% | 1,164 | 28% | 1,112 | 27% | 3,477 | 28% |
| **Number of Children** | | | | | | | | |
| No children | 862 | 20% | 850 | 20% | 838 | 21% | 2,550 | 20% |
| One or more child | 3,398 | 80% | 3,384 | 80% | 3,246 | 79% | 10,028 | 80% |
| **Parent of Child Less than 18 yrs of Age** | | | | | | | | |
| No | 2,508 | 57% | 2,480 | 57% | 2,344 | 55% | 7,332 | 56% |
| Yes | 1,897 | 43% | 1,891 | 43% | 1,880 | 45% | 5,668 | 44% |
| **Religiosity** | | | | | | | | |
| Not at all religious | 615 | 14% | 603 | 14% | 591 | 14% | 1,809 | 14% |
| Not very religious | 684 | 16% | 660 | 15% | 660 | 16% | 2,004 | 16% |
| Somewhat religious | 1,881 | 44% | 1,836 | 43% | 1,827 | 44% | 5,544 | 44% |
| Very religious | 1,131 | 26% | 1,172 | 27% | 1,057 | 26% | 3,360 | 26% |
| **Self-Rated Health** | | | | | | | | |
| Excellent | 544 | 13% | 558 | 13% | 501 | 12% | 1,603 | 13% |
| Very good | 1,364 | 31% | 1,345 | 31% | 1,287 | 31% | 3,996 | 31% |
| Good | 1,606 | 37% | 1,569 | 37% | 1,589 | 38% | 4,764 | 37% |

(*Continued on next page*)

**Table 1.** *Continued*

| | Broad-Controlled | | Broad-Open | | Tiered-Controlled | | Overall | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| Fair | 659 | 15% | 660 | 15% | 660 | 16% | 1,979 | 15% |
| Poor | 168 | 4% | 151 | 4% | 133 | 3% | 452 | 4% |
| **Diagnosis of a Genetic Disorder** | | | | | | | | |
| No | 3,795 | 93% | 3,783 | 95% | 3,664 | 94% | 11,242 | 94% |
| Yes | 264 | 7% | 218 | 5% | 248 | 6% | 730 | 6% |

Note: Observed frequencies and percentages are reported ignoring sampling design. Pearson's chi-square tests were performed to assess differences between consent and data sharing models and each characteristic; no differences were detected and thus for brevity test summaries were omitted, but are available from the authors.
[a]Federal poverty level guidelines were used to assign poverty status and are a function of income and number of individuals within a household (below poverty line = 1 if income < number in household*4,160+11,770) (Department of Health and Human Services).[53] Categorized income levels were collected in the survey and thus interval midpoints were used as a proxy for income in the threshold formula.

no evidence to suggest that such interactions existed (not shown). Responses to the individual attitudinal items were described, and 95% confidence intervals were computed.

Multiple imputation was conducted within each site to account for item non-response, which ranged from less than 1% to 5% for all key variables except income (8%). We used socio-demographic variables, biobank participation willingness, attitudinal constructs, and deciles of post-stratified weights to impute all missing data.[43] Ten complete imputation datasets were created for each site. Survey weighted regression analyses were performed on each complete dataset and combined using the standard "Rubin's rules."[44] For every analysis, site-specific estimates were then combined to summarize characteristics of the entire eMERGE Network using multivariate random-effect, meta-analytic methods.[45] All analyses were performed using R version 3.2.2 (R Project for Statistical Computing) and Stata Version 14 (StataCorp).

## Results

### Participant Characteristics

#### *Socio-demographics*

Of 90,000 surveys mailed, 7,672 individuals were ineligible due to invalid address, death, or incapacity, and 681 refused to participate. Of the 82,328 eligible individuals, exactly 13,000 responded (AAPOR reponse rate 15.8%). Among responders, 11,712 completed the paper (91.9%) and 1,288 the online (9.9%) survey. Sixty-three percent of participants were female; 51% self-identified as white; 18% as Hispanic; 42% had less than a Bachelors degree; and 44% had an annual household income of $60,000 or less (Table 1). There were no sociodemographic differences among participants receiving each of the three scenarios, indicating randomization was successful.

#### *Trust and Privacy*

Ninety percent of participants agreed health information privacy was important to them; 64% agreed that they worried about the privacy of their health information. Two thirds agreed that they trusted their healthcare system (64%) and medical researchers (61%). There were no differences in participants' trust and privacy among the three biobank scenario groups (Table 2).

### Willingness Compared between Scenario Groups

Overall, 66% (95% CI: 63%–69%) of participants stated that they would be willing to participate in the biobank described to them (Table 3). Willingness did not differ between broad and tiered consent models (68% versus 66%, $\chi^2 = 1.07$, $p = 0.30$). Willingness was slightly higher among participants presented with a controlled rather than an open data sharing model, although the difference was not large in absolute terms (68% versus 65%, respectively, $\chi^2 = 4.48$, $p = 0.03$).

### Attitudes Compared between Scenario Groups

Mean attitude scores (where 1 indicates low, 5 indicates high) were 3.85 for perceived benefits, 3.11 for concerns, and 3.95 for information needs overall. Mean scores did not differ between experimental conditions (see Table 3).

### Associations between Willingness and Participant Characteristics

Because patterns of associations between socio-demographic variables and willingness were the same within each of the consent and data sharing conditions separately, and because willingness did not differ at the 0.001 level between groups, all subsequent analyses were conducted on the sample as a whole.

The following participant characteristics were independently associated with willingness to participate before attitudes were entered into the model: race (as self-reported by the respondent in the survey), education, religiosity, and trust and privacy concerns (Table 4). Black or African American participants expressed the lowest levels (56%) and white participants the highest levels (70%) of willingness to participate (OR 0.59, 95% CI: 0.47 to 0.76). Participants who reported education "up to some high school" were less willing to participate (51%) than participants with doctoral training (76%) (OR 0.47, 95% CI: 0.33 to 0.67). "Very religious" participants were less willing to participate (63%) than "not at all religious" participants (73%) (OR 0.68, 95% CI: 0.54 to 0.85). Participants with lower levels of trust in medical researchers

**Table 2. Trust in the Healthcare System, Trust in Medical Researchers, and Concerns about Privacy of Survey Respondents by Consent and Data Sharing Model and across Models**

| | Broad-Controlled | | Broad-Open | | Tiered-Controlled | | Overall | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| **Privacy: Health Information Privacy Is Important to Me** | | | | | | | | |
| Disagree or strongly disagree | 117 | 3% | 96 | 2% | 100 | 2% | 313 | 2% |
| Neither agree nor disagree | 332 | 8% | 327 | 8% | 344 | 8% | 1,003 | 8% |
| Agree or strongly agree | 3,861 | 90% | 3,850 | 90% | 3,686 | 89% | 11,397 | 90% |
| **Privacy: I Worry about the Privacy of My Health Information** | | | | | | | | |
| Disagree or strongly disagree | 705 | 16% | 671 | 16% | 678 | 16% | 2,054 | 16% |
| Neither agree nor disagree | 856 | 20% | 822 | 19% | 838 | 20% | 2,516 | 20% |
| Agree or strongly agree | 2,745 | 64% | 2,783 | 65% | 2,607 | 63% | 8,135 | 64% |
| **Trust: I Trust my Healthcare System** | | | | | | | | |
| Disagree or strongly disagree | 522 | 12% | 499 | 12% | 471 | 11% | 1,492 | 12% |
| Neither agree nor disagree | 1,018 | 24% | 1,049 | 25% | 1,021 | 25% | 3,088 | 24% |
| Agree or strongly agree | 2,778 | 64% | 2,726 | 64% | 2,637 | 64% | 8,141 | 64% |
| **Trust: I Trust Medical Researchers** | | | | | | | | |
| Disagree or strongly disagree | 301 | 7% | 325 | 8% | 279 | 7% | 905 | 7% |
| Neither agree nor disagree | 1,336 | 31% | 1,362 | 32% | 1,306 | 32% | 4,004 | 32% |
| Agree or strongly agree | 2,648 | 62% | 2,565 | 60% | 2,535 | 62% | 7,748 | 61% |

Observed frequencies and percentages are reported ignoring sampling design. Pearson's chi-square tests were performed to assess differences between consent and data sharing models and each characteristic; no differences were detected and thus for brevity test summaries were omitted, but are available from the authors.

and/or their healthcare system, those who felt more strongly that the privacy of their health information was important, and those more worried about the privacy of their health information were less willing to participate.

When attitudes toward the biobank were entered into the model, each of the three composite scale variables was independently associated with willingness: participants were more willing to participate if they perceived more benefits, had fewer concerns, and had fewer information needs (Table 4). In this model, education and religiosity remained associated with willingness, but race, trust, and privacy concerns did not.

### Benefits, Concerns, and Information Needs

The most endorsed benefit in the sample overall was "I would feel that I was helping future generations" (84%, 81%–87%) and the least endorsed benefit was "I would feel that taking part could help me personally" (44%, 40%–47%). The most endorsed concern was "I would worry about my privacy" (51%, 47%–55%) and the least endorsed concern was "I would worry that someone might make money using my health information" (36%, 33%–39%). The most endorsed information need about biobank governance was "I would want to know what would happen if a researcher misused the health information in the biobank" (86%, 84%–87%) and the least endorsed information need about biobank governance was "I would want to know if my health information might be used by drug companies that make money" (59%, 56%–61%) (Table 5).

## Discussion

This is the largest survey of patients' attitudes toward participating in biobank research to date. In this study, we found no evidence to support the hypothesis that asking potential biobank participants to provide broad consent or permit open data sharing would lead to less willingness to participate than asking them to provide tiered consent or permit controlled data sharing. Models of consent and data sharing had limited relevance to participants' decision making when they were asked to make a decision about whether to participate in the single biobank scenario presented to them. These findings are consistent with previous research reporting that overall acceptance of broad consent is similar to that of specific or tiered consent, although a number of factors may influence this preference.[23–25] Our use of an experimental design and randomization means participants were presented with a decision that was closer to the real world than previous studies that have given participants choices and asked which consent model they preferred. Individuals may have pre-existing global views regarding

**Table 3. Willingness and Attitudes toward Participating in a Biobank by Consent and Data Sharing Model and across Models**

| | Biobank Consent and Data Sharing Model | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Broad-Controlled (BC) | | Broad-Open (BO) | | Tiered-Controlled (TC) | | All | | | |
| | N | % | N | % | N | % | N | % | BC versus BO | BC versus TC |
| **Primary Outcome** | | | | | | | | | | |
| **Willingness to Participate in a Biobank** | | | | | | | | | | |
| No definitely not / Probably not | 513 | 12 (10,14) | 611 | 15 (13,18) | 487 | 12 (9,15) | 1,611 | 13 (12,15) | $X_1^2 = 4.48$, $p = 0.03$ | $X_1^2 = 1.07$, $p = 0.30$ |
| Not sure | 853 | 20 (17,22) | 913 | 20 (17,23) | 853 | 22 (19,25) | 2,619 | 20 (19,22) | | |
| Yes probably / Yes definitely | 2,880 | 68 (65,72) | 2,702 | 65 (61,69) | 2,758 | 66 (62,71) | 8,340 | 66 (63,69) | | |
| **Secondary Outcomes** | | | | | | | | | | |
| **Perceived Benefits** | | | | | | | | | | |
| Low (1.0–2.5) | 143 | 4 (3,5) | 147 | 4 (3,5) | 146 | 3 (2,5) | 436 | 4 (3,5) | $X_1^2 = 0.70$, $p = 0.79$ | $X_1^2 = 1.42$, $p = 0.23$ |
| Intermediate (2.5–3.5) | 887 | 21 (18,23) | 926 | 22 (19,25) | 934 | 23 (20,27) | 2,747 | 22 (20,24) | | |
| High (3.5–5.0) | 3,164 | 75 (72,78) | 3,063 | 75 (71,78) | 2,952 | 73 (69,77) | 9,179 | 74 (72,77) | | |
| **Concerns** | | | | | | | | | | |
| Low (1.0–2.5) | 1,321 | 34 (30,38) | 1,231 | 32 (28,37) | 1,305 | 34 (29,39) | 3,857 | 33 (30,37) | $X_1^2 = 1.23$, $p = 0.27$ | $X_1^2 = 0.06$, $p = 0.80$ |
| Intermediate (2.5–3.5) | 1,263 | 30 (28,33) | 1,243 | 30 (27,33) | 1,175 | 30 (28,32) | 3,681 | 30 (29,32) | | |
| High (3.5–5.0) | 1,520 | 36 (32,40) | 1,636 | 38 (34,42) | 1,479 | 36 (32,40) | 4,635 | 36 (33,40) | | |
| **Information Needs** | | | | | | | | | | |
| Low (1.0–2.5) | 176 | 5 (3,6) | 179 | 6 (5,8) | 190 | 5 (3,6) | 545 | 5 (4,6) | $X_1^2 = 0.08$, $p = 0.77$ | $X_1^2 = 1.30$, $p = 0.25$ |
| Intermediate (2.5–3.5) | 831 | 22 (19,24) | 695 | 19 (17,21) | 767 | 19 (17,22) | 2,293 | 20 (18,22) | | |
| High (3.5–5.0) | 3,161 | 74 (71,77) | 3,267 | 75 (72,78) | 3,087 | 76 (72,80) | 9,506 | 75 (72,77) | | |

Secondary outcomes were defined as the average of the recoded survey items (1 = No, definitely not, …, 5 = Yes, definitely) that comprised the outcomes. Observed frequencies and survey-adjusted percentages (95% CI) are reported for all outcomes. Wald tests were performed to assess differences between data sharing models (broad-controlled [BC] versus broad-open [BO]) and between consent types (broad-controlled versus tiered-controlled).

participating in biobank research (i.e., they are generally open to, or generally against, participating), which are not swayed by the consent and data sharing models presented to them. It is also possible that many participants did not read the scenario carefully and were therefore responding to the general idea of participating in a biobank rather than the specific scenario presented to them. While this should certainly be considered a possible limitation of our findings, it is also possible that this lack of attention to detail simulates, in some ways, the attention that potential biobank participants give to analogous details in a real biobank consent document.

Our findings support the hypothesis that socio-demographic characteristics are associated with willingness to participate in biobank research. Consistent with previous research,[46,47] willingness to participate in a biobank was significantly lower among participants who self-identified as black or African American and those with lower levels of educational attainment. Although religiosity is difficult to assess, participants who self-identified as more religious on this simple measure were less willing to participate. Little research has previously explored associations between religiosity and attitudes regarding biobank participation,[48] but other investigations have shown religiosity is associated with negative perceptions of the value of science and technology.[49–52] Our findings support previous indications that certain socio-demographic groups will require greater efforts to ensure participation in large research initiatives going forward. Our results have implications particularly for biobanks such as the PMI that plan to rely on recruitment of volunteers rather than selection to represent the population. Unless there is an attempt to over-enroll underrepresented groups, as we did in our survey study, large studies like the PMI may end up with cohort samples that do not adequately represent the US population.

We also found that, consistent with previous research,[23,36] willingness to participate in biobank research was lower among respondents with more concerns about privacy and lower levels of trust. When designing recruitment strategies, researchers and institutions may be able to address these trust and privacy concerns at a local level by building relationships with their communities, and at a national level by implementing new policies and public education and

**Table 4. Univariate and Multivariate Associations between Consent and Data Sharing Models, Socio-demographics, Trust and Privacy Items, and Attitudinal Constructs and Willingness to Participate in a Biobank**

| Independent Variable | Percent (95% CI) | Socio-demographics | Multivariate Models, OR (95% CI) | |
|---|---|---|---|---|
| | | | Socio-demographics, Trust, and Privacy | Socio-demographics, Trust, Privacy, and Attitudes |
| **Consent and Data Sharing Model** | | | | |
| Broad-controlled | 68 (65, 71) | 1 | 1 | 1 |
| Broad-open | 65 (62, 68) | 0.84 (0.68, 1.03) | 0.81 (0.66, 1.00)[a] | 0.80 (0.63, 1.02) |
| Tiered-controlled | 66 (61, 70) | 0.94 (0.76, 1.17) | 0.93 (0.69, 1.24) | 0.90 (0.66, 1.22) |
| **Sex** | | | | |
| Female | 66 (63, 69) | 1 | 1 | 1 |
| Male | 67 (63, 71) | 0.92 (0.76, 1.11) | 0.87 (0.72, 1.07) | 0.91 (0.73, 1.13) |
| **Age** | | | | |
| 18–35 years | 65 (61,70) | 1.05 (0.81, 1.35) | 0.92 (0.65, 1.30) | 0.97 (0.70, 1.34) |
| 36–50 years | 65 (61,68) | 1 | 1 | 1 |
| 51–64 years | 66 (62,70) | 1.06 (0.76, 1.46) | 0.99 (0.64, 1.53) | 0.91 (0.59, 1.40) |
| 65+ years | 70 (65,75) | 1.01 (0.63, 1.62) | 0.80 (0.43, 1.50) | 0.79 (0.39, 1.58) |
| **Race** | | | | |
| White | 70 (67,74) | 1 | 1 | 1 |
| Black or African American | 56 (51,60) | 0.59 (0.47, 0.76)[a] | 0.68 (0.52, 0.88)[a] | 0.74 (0.53, 1.04) |
| Asian | 60 (54,66) | 0.62 (0.49, 0.76)[a] | 0.67 (0.53, 0.84)[a] | 0.79 (0.59, 1.04) |
| American Indian or Alaska Native | 57 (49,65) | 0.70 (0.51, 0.95)[a] | 0.78 (0.57, 1.05) | 0.81 (0.57, 1.15) |
| Other[b] | 56 (49,63) | 0.67 (0.49, 0.94)[a] | 0.82 (0.46, 1.46) | 0.89 (0.61, 1.31) |
| More than one race | 65 (56,73) | 0.83 (0.54, 1.27) | 0.90 (0.60, 1.35) | 1.10 (0.67, 1.81) |
| **Ethnicity** | | | | |
| Latino | 67 (64, 70) | 1 | 1 | 1 |
| Not Hispanic/Latino | 61 (55, 66) | 0.94 (0.71, 1.24) | 0.91 (0.59, 1.40) | 0.89 (0.66, 1.21) |
| **Educational Attainment** | | | | |
| Up to some high school (grades 9–12) | 51 (44, 57) | 0.47 (0.33, 0.67)[a] | 0.40 (0.26, 0.59)[a] | 0.34 (0.21, 0.54)[a] |
| High school graduate or GED | 58 (53, 64) | 0.63 (0.43, 0.93)[a] | 0.60 (0.39, 0.90)[a] | 0.52 (0.33, 0.84)[a] |
| Some college | 64 (60, 68) | 0.78 (0.58, 1.04) | 0.79 (0.56, 1.12) | 0.62 (0.44, 0.88)[a] |
| Bachelors degree or equivalent | 70 (67, 73) | 0.83 (0.60, 1.14) | 0.83 (0.58, 1.18) | 0.75 (0.52, 1.08) |
| Masters degree or equivalent | 73 (68, 77) | 0.96 (0.73, 1.28) | 0.96 (0.73, 1.28) | 0.91 (0.64, 1.29) |
| PhD, MD, JD, or equivalent | 76 (70, 81) | 1 | 1 | 1 |
| **Annual Household Income** | | | | |
| Less than $30,000 | 56 (52,60) | 0.58 (0.42, 0.79)[a] | 0.67 (0.46, 0.97)[a] | 0.74 (0.48, 1.13) |
| $30,000–$60,000 | 63 (58,67) | 0.66 (0.43, 0.93)[a] | 0.71 (0.52, 0.96)[a] | 0.81 (0.57, 1.13) |
| $60,000–$90,000 | 68 (64,72) | 0.84 (0.63, 1.13) | 0.89 (0.63, 1.25) | 0.99 (0.68, 1.43) |
| $90,000–$150,000 | 74 (70,77) | 1.12 (0.83, 1.50) | 1.11 (0.85, 1.46) | 1.15 (0.83, 1.61) |
| More than $150,000 | 73 (68,77) | 1 | 1 | 1 |
| **Total Number of People in Household** | | | | |
| 1 | 84 (66, 94) | 2.23 (0.68, 7.32) | 2.77 (0.29, 26.80) | 2.08 (0.38, 11.48) |
| 2 | 68 (63, 72) | 1.10 (0.87, 1.40) | 1.10 (0.79, 1.53) | 1.12 (0.79, 1.58) |
| 3 | 66 (62,69) | 1.01 (0.81, 1.25) | 1.09 (0.83, 1.43) | 1.10 (0.85, 1.43) |

(*Continued on next page*)

**Table 4.** ***Continued***

| Independent Variable | Percent (95% CI) | Multivariate Models, OR (95% CI) | | |
|---|---|---|---|---|
| | | Socio-demographics | Socio-demographics, Trust, and Privacy | Socio-demographics, Trust, Privacy, and Attitudes |
| 4 or more | 65 (62,68) | 1 | 1 | 1 |
| **Work Situation** | | | | |
| Working | 67 (64,70) | 1 | 1 | 1 |
| Retired | 69 (65,73) | 1.24 (0.94, 1.64) | 1.22 (0.90, 1.66) | 1.10 (0.73, 1.66) |
| Disabled / Unemployed | 60 (55,65) | 0.90 (0.70, 1.16) | 0.99 (0.77, 1.28) | 0.98 (0.70, 1.37) |
| Other | 64 (59,68) | 0.97 (0.75, 1.26) | 1.00 (0.65, 1.53) | 0.95 (0.66, 1.37) |
| **Healthcare Insurance** | | | | |
| Private insurance | 67 (64,71) | 1 | 1 | 1 |
| Public insurance | 64 (60,69) | 1.45 (1.04, 2.03)[a] | 1.33 (0.96, 1.84) | 1.37 (0.89, 2.12) |
| Other type of insurance | 59 (49,68) | 1.05 (0.65, 1.68) | 0.95 (0.57, 1.58) | 1.09 (0.63, 1.87) |
| No insurance | 63 (54,71) | 1.79 (1.08, 2.96)[a] | 1.79 (1.00, 3.23)[a] | 1.65 (0.81, 3.35) |
| **Rurality (from Census-level Data)** | | | | |
| Suburban or urban | 66 (63,70) | 1 | 1 | 1 |
| Rural | 67 (64,71) | 0.94 (0.79, 1.11) | 0.95 (0.80, 1.13) | 0.99 (0.77, 1.27) |
| **Marital Status** | | | | |
| Married, living with someone | 69 (65,72) | 1 | 1 | 1 |
| Not married, living with someone | 68 (63,72) | 1.08 (0.79, 1.47) | 1.08 (0.76, 1.54) | 0.95 (0.67, 1.35) |
| Not married, not living with someone | 61 (56,66) | 0.91 (0.71, 1.17) | 0.88 (0.66, 1.19) | 0.86 (0.60, 1.25) |
| **Religiosity** | | | | |
| Not at all religious | 73 (69,77) | 1 | 1 | 1 |
| Not very religious | 71 (67,74) | 0.89 (0.68, 1.16) | 0.85 (0.64, 1.13) | 0.68 (0.48, 0.97)[a] |
| Somewhat religious | 65 (61,69) | 0.74 (0.59, 0.93)[a] | 0.73 (0.56, 0.96)[a] | 0.62 (0.46, 0.83)[a] |
| Very religious | 63 (59,66) | 0.68 (0.54, 0.85)[a] | 0.68 (0.53, 0.88)[a] | 0.59 (0.44, 0.80)[a] |
| **I Trust My Healthcare System** | | | | |
| Disagree | 48 (43,54) | – | 0.73(0.55, 0.98)[a] | 1.02 (0.70, 1.47) |
| Neither | 56 (53,60) | – | 0.81 (0.60, 1.08) | 0.95 (0.69, 1.29) |
| Agree | 74 (71,77) | – | 1 | 1 |
| **I Trust Medical Researchers** | | | | |
| Disagree | 34 (29,41) | – | 0.18 (0.13, 0.25)[a] | 0.52 (0.36, 0.75)[a] |
| Neither | 52 (49,55) | – | 0.38 (0.30, 0.49)[a] | 0.61 (0.47, 0.80)[a] |
| Agree | 78 (75,80) | – | 1 | 1 |
| **Health Information Privacy Is Important to Me** | | | | |
| Disagree | 76 (67,83) | – | 1.10 (0.63, 1.94) | 1.18 (0.80, 2.25) |
| Neither | 80 (75,84) | – | 1.47 (1.02, 2.13)[a] | 1.14 (0.80, 1.64) |
| Agree | 65 (61,68) | – | 1 | 1 |
| **I Worry about the Privacy of My Health Information** | | | | |
| Disagree | 87 (85,89) | – | 3.64 (2.63, 5.02)[a] | 1.44 (1.01, 2.07)[a] |
| Neither | 77 (74,80) | – | 2.15 (1.64, 2.83)[a] | 1.10 (0.81, 1.49) |
| Agree | 57 (53,61) | – | 1 | 1 |

(*Continued on next page*)

**Table 4.** *Continued*

| Independent Variable | Percent (95% CI) | Multivariate Models, OR (95% CI) | | |
|---|---|---|---|---|
| | | Socio-demographics | Socio-demographics, Trust, and Privacy | Socio-demographics, Trust, Privacy, and Attitudes |
| **Perceived Benefits of Participating in a Biobank** | | | | |
| Low (1.0–2.5) | 6 (3, 10) | – | – | 1 |
| Intermediate (2.5–3.5) | 32 (29,36) | – | – | 8.10 (3.47, 18.90)[a] |
| High (3.5–5.0) | 80 (77,82) | – | – | 62.21 (28.72, 134.75)[a] |
| **Concerns about Participating in a Biobank** | | | | |
| Low (1.0–2.5) | 91 (90,93) | – | – | 1 |
| Intermediate (2.5–3.5) | 76 (72,80) | – | – | 0.32 (0.24, 0.43)[a] |
| High (3.5–5.0) | 40 (37,43) | – | – | 0.07 (0.05, 0.10)[a] |
| **Information Needs about Participating in a Biobank** | | | | |
| Low (1.0–2.5) | 76 (68,82) | – | – | 1 |
| Intermediate (2.5–3.5) | 78 (73,82) | – | – | 1.36 (0.88, 2.09) |
| High (3.5–5.0) | 62 (59,65) | – | – | 1.62 (1.02, 2.58)[a] |

Survey-adjusted logistic regression estimates from the univariate models have been transformed to percentages (95% CI) while odds ratios (OR, 95% CI) summarize multivariate models.
[a]Significant differences at the 0.05 level.
[b]Due to small cell counts, Native Hawaiian/Pacific Islanders were grouped with the Other racial category.

engagement strategies. We also found, however, that these associations between privacy, trust, and willingness were no longer statistically significant after attitudes toward biobank research (benefits, concerns, information needs about governance) were added in to the model. This further suggests that outreach and engagement conveying the benefits of biobank research, and addressing concerns and information needs, may be important and achievable strategies to including more diverse populations in these research efforts.

We found that participants were more concerned about some risks than others and wanted to know more about some aspects of governance than others. These findings may be of particular value to researchers, institutions, and organizations involved in developing public education and information materials about biobanks, as they shed valuable light on what potential participants want to know about research.

Limitations include a low response rate. Although this was not unexpected given our recruitment method and efforts to oversample populations typically less willing to participate in research, we cannot know how much opinions of nonresponders would differ. However, it seems unlikely that nonresponders would be *more* willing to enroll in biobanks. Our study also relied on participants' self-reported intentions and hypothetical scenarios, rather than actual behavior. It is also possible that participants did not read all of the hypothetical scenario, and so may not have appreciated the detail regarding the consent and data sharing model. They did, however, have opinions about other aspects of the proposal that suggests substantial engagement with the survey. The limitations need to be weighed against the study's strengths, which include broad geographical and diverse institutional coverage, a rigorous sampling strategy, and experimental design.

In conclusion, the results from this study suggest that biobanks using broad consent may not be less successful in recruiting participants than if they use more specific consent approaches. Open data sharing may be almost as acceptable to participants as controlled data sharing. Some socio-demographic groups differ in their willingness to participate in biobank research. Individuals discriminate between different positive and negative aspects of biobank participation and feel more strongly about knowing about some aspects of biobank governance than others. Targeted interventions designed to recruit underrepresented groups, to make biobank information easier to understand, and to address individuals' specific attitudes about participating in a biobank may help increase acceptance of broad consent and open data sharing in biobank research. These findings may be of use to biobank investigators concerned with how biobanks are governed and to policy makers working to revise regulations on the protection of human research subjects.

## Supplemental Data

Supplemental Data include three tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2017.01.021.

## Acknowledgments

We would like to acknowledge the work of the following: Melissa Basford, Meckenzie Behr, Jess Behrens, Laura Beskow, Ariel Chandler, Beth Chau, Rosetta Chiavacci, Kurt Christensen, Beth Cobb, Natalia Connolly, Stephanie Devaney, Joe DeWalle, Steve

**Table 5. Attitudes toward Participating in a Biobank**

| Attitudes | Agree or Strongly Agree | |
|---|---|---|
| | N | Percent (95% CI) |
| **Perceived Benefits of Participating in a Biobank** | | |
| I would feel that I was helping future generations. | 10,773 | 84 (81, 87) |
| I would feel that taking part could lead to better medical treatments. | 10,564 | 83 (80, 85) |
| I would feel that taking part would help doctors where I get my medical care take better care of patients. | 9,957 | 78 (75, 80) |
| I would feel that taking part could help my family. | 8,139 | 65 (62, 67) |
| I would feel that taking part could help me personally. | 5,667 | 44 (40, 47) |
| **Concerns about Participating in a Biobank** | | |
| I would worry about my privacy. | 6,578 | 51 (47, 55) |
| I would worry about my medical record being shared. | 5,903 | 45 (42, 49) |
| I would worry about how researchers would use my health information. | 5,404 | 41 (38, 45) |
| I would worry about my genetic information being shared. | 4,866 | 38 (34, 41) |
| I would worry that some research would be done that I did not want to take part in. | 4,713 | 37 (34, 40) |
| I would worry that someone might make money using my health information. | 4,856 | 36 (33, 39) |
| **Information Needs Regarding Governance of a Biobank** | | |
| I would want to know what would happen if a researcher misused the health information in the biobank. | 11,055 | 86 (84, 87) |
| I would want to know what kind of knowledge would result from the use of my health information. | 10,827 | 84 (82, 86) |
| I would want to know who makes sure that my health information is used in the right way. | 10,752 | 84 (81, 86) |
| I would want to know if my health information might be used by insurance companies. | 10,160 | 79 (77, 81) |
| I would want to know the types of research my health information would be used for. | 9,748 | 74 (71, 77) |
| I would want to know who runs the biobank. | 9,500 | 73 (71, 75) |
| I would want to know how the biobank covers costs. | 7,716 | 60 (57, 62) |
| I would want to know if my health information might be used by drug companies that make money. | 7,625 | 59 (56, 61) |

Each item was dichotomized (1 = agree or strongly agree, 0 = otherwise) and the observed frequencies and ordered survey-adjusted percentages (95% CI) are reported. Differences were not detected between data sharing and consent models on the construct level (Table 3) and thus only overall summaries are provided.

Ellis, Alexander Fiksdal, John Harley, Diana Harris, Paul Hitz, Yolanda Keppel, Nicole Lockhart, Keith Marsolo, Catherine MarxKate Nowakowski, Jen Pacheco, Josh Pankratz, Lisa Price, Michelle Ramos, Sarah Stallings, Allise Taran, Matt Verkemp, Gordon Willis, and Sonja Ziniel.

The CERC Survey project within the eMERGE Network was initiated and funded by NHGRI with additional funding by the NIH Office of the Director through the following grants: U01HG006828 (Cincinnati Children's Hospital Medical Center/Boston Children's Hospital), U01HG006830 (Children's

Hospital of Philadelphia), UOIHG006389 (Essentia Institute of Rural Health, Marshfield Clinic Research Foundation, and Pennsylvania State University), U01HG006382 (Geisinger Clinic), U01HG006375 (Group Health Cooperative/University of Washington), 3U01HG006379 (Mayo Clinic), U01HG006380 (Icahn School of Medicine at Mount Sinai), 3U01-HG006388 (Northwestern University), U01HG006378 (Vanderbilt University Medical Center), and 3U01HG0006385 (Vanderbilt University Medical Center serving as the Coordinating Center).

Received: September 23, 2016
Accepted: January 11, 2017
Published: February 9, 2017

## Web Resources

CDC, BRFSS Questionnaires, https://www.cdc.gov/brfss/questionnaires/

eMerge, https://emerge.mc.vanderbilt.edu/projects/emerge-cerc-survey-2/

## References

1. Manolio, T.A., Rodriguez, L.L., Brooks, L., Abecasis, G., Ballinger, D., Daly, M., Donnelly, P., Faraone, S.V., Frazer, K., Gabriel, S., et al.; GAIN Collaborative Research Group; Collaborative Association Study of Psoriasis; International Multi-Center ADHD Genetics Project; Molecular Genetics of Schizophrenia Collaboration; Bipolar Genome Study; Major Depression Stage 1 Genomewide Association in Population-Based Samples Study; and Genetics of Kidneys in Diabetes (GoKinD) Study (2007). New models of collaboration in genome-wide association studies: the Genetic Association Information Network. Nat. Genet. *39*, 1045–1051.
2. Hudson, K., Lifton, R., Patrick-Lake, B., Gonzalez Burchard, E., Coles, T., Collins, R., Conrad, A., Denny, J., Desmond-Hellmann, S., Dishman, E., et al. (2015). The Precision Medicine Initiative Cohort Program: Building a Research Foundation for 21st Century Medicine (Bethesda, MD: National Institutes of Health).
3. Ashley, E.A. (2015). The precision medicine initiative: a new national effort. JAMA *313*, 2119–2120.
4. Collins, F.S., and Varmus, H. (2015). A new initiative on precision medicine. N. Engl. J. Med. *372*, 793–795.
5. Department of Health and Human Services (2014). Final NIH genomic data sharing policy. Fed. Regist. *79*, 51345–51354.
6. Department of Health and Human Services (2015). Federal policy for the protection of human subjects. Fed. Regist. *80*, 53933–54061.
7. Shabani, M., Bezuidenhout, L., and Borry, P. (2014). Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review. Expert Rev. Mol. Diagn. *14*, 1053–1065.
8. Kaye, J. (2012). The tension between data sharing and the protection of privacy in genomics research. Annu. Rev. Genomics Hum. Genet. *13*, 415–431.
9. Kaufman, D., Murphy, J., Erby, L., Hudson, K., and Scott, J. (2009). Veterans' attitudes regarding a database for genomic research. Genet. Med. *11*, 329–337.
10. McGuire, A.L., Hamilton, J.A., Lunstroth, R., McCullough, L.B., and Goldman, A. (2008). DNA data sharing: research participants' perspectives. Genet. Med. *10*, 46–53.
11. Trinidad, S.B., Fullerton, S.M., Ludman, E.J., Jarvik, G.P., Larson, E.B., and Burke, W. (2011). Research ethics. Research practice and participant preferences: the growing gulf. Science *331*, 287–288.
12. Murphy, J., Scott, J., Kaufman, D., Geller, G., LeRoy, L., and Hudson, K. (2009). Public perspectives on informed consent for biobanking. Am. J. Public Health *99*, 2128–2134.
13. Steinsbekk, K.S., Kåre Myskja, B., and Solberg, B. (2013). Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? Eur. J. Hum. Genet. *21*, 897–902.
14. Collins, F.S., and Manolio, T.A. (2007). Merging and emerging cohorts: necessary but not sufficient. Nature *445*, 259–259.
15. Fullerton, S.M. (2011). The input-output problem: whose DNA do we study, and why does it matter? In Achieving Justice in Genomic Translation Re-thinking the Pathway to Benefit, W. Burke, K.A. Edwards, and S. Goering, eds. (Oxford: Oxford University Press), pp. 40–55.
16. Need, A.C., and Goldstein, D.B. (2009). Next generation disparities in human genomics: concerns and remedies. Trends Genet. *25*, 489–494.
17. Rosenberg, N.A., Huang, L., Jewett, E.M., Szpiech, Z.A., Jankovic, I., and Boehnke, M. (2010). Genome-wide association studies in diverse populations. Nat. Rev. Genet. *11*, 356–366.
18. Bustamante, C.D., Burchard, E.G., and De la Vega, F.M. (2011). Genomics for the world. Nature *475*, 163–165.
19. Shavers, V.L., Lynch, C.F., and Burmeister, L.F. (2000). Knowledge of the Tuskegee study and its impact on the willingness to participate in medical research studies. J. Natl. Med. Assoc. *92*, 563–572.
20. Bates, B.R., and Harris, T.M. (2004). The Tuskegee Study of Untreated Syphilis and public perceptions of biomedical research: a focus group study. J. Natl. Med. Assoc. *96*, 1051–1064.
21. Hartzler, A., McCarty, C.A., Rasmussen, L.V., Williams, M.S., Brilliant, M., Bowton, E.A., Clayton, E.W., Faucett, W.A., Ferryman, K., Field, J.R., et al. (2013). Stakeholder engagement: a key component of integrating genomic information into electronic health records. Genet. Med. *15*, 792–801.
22. Garrison, N.A., Sathe, N.A., Antommaria, A.H., Holm, I.A., Sanderson, S.C., Smith, M.E., McPheeters, M.L., and Clayton, E.W. (2016). A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. Genet. Med. *18*, 663–671.
23. Kaufman, D.J., Murphy-Bollinger, J., Scott, J., and Hudson, K.L. (2009). Public opinion about the importance of privacy in biobank research. Am. J. Hum. Genet. *85*, 643–654.
24. Platt, J., Bollinger, J., Dvoskin, R., Kardia, S.L., and Kaufman, D. (2014). Public preferences regarding informed consent models for participation in population-based genomic research. Genet. Med. *16*, 11–18.
25. Simon, C.M., L'heureux, J., Murray, J.C., Winokur, P., Weiner, G., Newbury, E., Shinkunas, L., and Zimmerman, B. (2011). Active choice but not too active: public perspectives on biobank consent models. Genet. Med. *13*, 821–831.
26. Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W.A., Li, R., Manolio, T.A., Sanderson, S.C., Kannry, J., Zinberg, R., Basford, M.A., et al.; eMERGE Network (2013). The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet. Med. *15*, 761–771.
27. Smith, M.E., Sanderson, S.C., Brothers, K.B., Myers, M.F., McCormick, J., Aufox, S., Shrubsole, M.J., Garrison, N.A.,

Mercaldo, N.D., Schildcrout, J.S., et al. (2016). Conducting a large, multi-site survey about patients' views on broad consent: challenges and solutions. BMC Med. Res. Methodol. *16*, 162.
28. Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J.G. (2009). Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J. Biomed. Inform. *42*, 377–381.
29. Hamilton, C.M., Strader, L.C., Pratt, J.G., Maiese, D., Hendershot, T., Kwok, R.K., Hammond, J.A., Huggins, W., Jackman, D., Pan, H., et al. (2011). The PhenX Toolkit: get the most from your measures. Am. J. Epidemiol. *174*, 253–260.
30. Fetzer Institute (1999). Multidimensional Measurement of Religiousness/Spirituality for Use in Health Research: A Report of the Fetzer Institute/National Institute on Aging Working Group (Kalamazoo, MI: John E. Fetzer Institute).
31. Ware, J., Jr., Kosinski, M., and Keller, S.D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med. Care *34*, 220–233.
32. Terry, A.L., Chesworth, B.M., Stolee, P., Bourne, R.B., and Speechley, M. (2008). Joint replacement recipients' post-surgery views about health information privacy and registry participation. Health Policy *85*, 293–304.
33. Hall, M.A., Camacho, F., Lawlor, J.S., Depuy, V., Sugarman, J., and Weinfurt, K. (2006). Measuring trust in medical researchers. Med. Care *44*, 1048–1053.
34. Shea, J.A., Micco, E., Dean, L.T., McMurphy, S., Schwartz, J.S., and Armstrong, K. (2008). Development of a revised Health Care System Distrust scale. J. Gen. Intern. Med. *23*, 727–732.
35. Gaskell, G., Allansdottir, A., Allum, N., Castro, P., Esmer, Y., Fischler, C., Jackson, J., Kronberger, N., Hampel, J., Mejlgaard, N., et al. (2011). The 2010 Eurobarometer on the life sciences. Nat. Biotechnol. *29*, 113–114.
36. Brothers, K.B., Morrison, D.R., and Clayton, E.W. (2011). Two large-scale surveys on community attitudes toward an opt-out biobank. Am. J. Med. Genet. A. *155A*, 2982–2990.
37. Helft, P.R., Champion, V.L., Eckles, R., Johnson, C.S., and Meslin, E.M. (2007). Cancer patients' attitudes toward future research uses of stored human biological materials. J. Empir. Res. Hum. Res. Ethics *2*, 15–22.
38. Ludman, E.J., Fullerton, S.M., Spangler, L., Trinidad, S.B., Fujii, M.M., Jarvik, G.P., Larson, E.B., and Burke, W. (2010). Glad you asked: participants' opinions of re-consent for dbGap data submission. J. Empir. Res. Hum. Res. Ethics *5*, 9–16.
39. Lemke, A.A., Wolf, W.A., Hebert-Beirne, J., and Smith, M.E. (2010). Public and biobank participant attitudes toward genetic research participation and data sharing. Public Health Genomics *13*, 368–377.
40. American Association for Public Opinion Research (2011). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, Seventh Edition (AAPOR).
41. Chowdhury, S., Khare, M., and Wolter, K. (2007). Weight trimming in the National Immunization Survey. Proceedings of the Survey Research Methods Section of the American Statistical Association, 2651–2658.
42. Potter, F. (1988). Survey of procedures to control extreme sampling weights. Proceedings of the Survey Research Methods Section of the American Statistical Association, 2651–2658.
43. Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika *70*, 41–55.
44. Rubin, D.B. (2004). Multiple Imputation for Nonresponse in Surveys (John Wiley & Sons).
45. Chen, H., Manning, A.K., and Dupuis, J. (2012). A method of moments estimator for random effect multivariate meta-analysis. Biometrics *68*, 1278–1284.
46. McQuillan, G.M., Porter, K.S., Agelli, M., and Kington, R. (2003). Consent for genetic research in a general population: the NHANES experience. Genet. Med. *5*, 35–42.
47. Chen, D.T., Rosenstein, D.L., Muthappan, P., Hilsenbeck, S.G., Miller, F.G., Emanuel, E.J., and Wendler, D. (2005). Research with stored biological samples: what do research participants want? Arch. Intern. Med. *165*, 652–655.
48. De Vries, R.G., Tomlinson, T., Kim, H.M., Krenz, C.D., Ryan, K.A., Lehpamer, N., and Kim, S.Y. (2016). The moral concerns of biobank donors: the effect of non-welfare interests on willingness to donate. Life Sci. Soc. Policy *12*, 3.
49. Johnson, D.R., Scheitle, C.P., Ecklund, E.H., Sørensen, J., and Soule, S. (2015). Individual religiosity and orientation towards science: reformulating relationships. Sociol. Sci. *2*, 106–124.
50. Brossard, D., Scheufele, D.A., Kim, E., and Lewenstein, B.V. (2008). Religiosity as a perceptual filter: examining processes of opinion formation about nanotechnology. Public Underst. Sci. *18*, 546–558.
51. Ho, S.S., Scheufele, D.A., and Corley, E.A. (2013). Factors influencing public risk-benefit considerations of nanotechnology: Assessing the effects of mass media, interpersonal communication, and elaborative processing. Public Underst. Sci. *22*, 606–623.
52. Scheufele, D.A., Corley, E.A., Shih, T.J., Dalrymple, K.E., and Ho, S.S. (2009). Religious beliefs and public attitudes toward nanotechnology in Europe and the United States. Nat. Nanotechnol. *4*, 91–94.
53. Department of Health and Human Services (2016). Annual update of the HHS poverty guidelines. Fed. Regist. *81*, 4036–4037.