The American Journal of Human Genetics, Volume *100*

Supplemental Data

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

Saskia C. Sanderson, Kyle B. Brothers, Nathaniel D. Mercaldo, Ellen Wright Clayton, Armand H. Matheny Antommaria, Sharon A. Aufox, Murray H. Brilliant, Diego Campos, David S. Carrell, John Connolly, Pat Conway, Stephanie M. Fullerton, Nanibaa' A. Garrison, Carol R. Horowitz, Gail P. Jarvik, David Kaufman, Terrie E. Kitchner, Rongling Li, Evette J. Ludman, Catherine A. McCarty, Jennifer B. McCormick, Valerie D. McManus, Melanie F. Myers, Aaron Scrol, Janet L. Williams, Martha J. Shrubsole, Jonathan S. Schildcrout, Maureen E. Smith, and Ingrid A. Holm

Table S1: Population, sample and response frequencies across the eMERGE Network and by participating institution. Response rates were calculated as the proportion responding among sampled population meeting consistency checks.

	N[1]	Sampled	Sampled + Checks[2]	Responded	Response Rate[3]
All					
	2,389,162	90,000	82,239	13,000	15.8
Population Type					
Adult	1,787,295	58,500	54,850	9,185	16.7
Pediatric	601,867	31,500	27,389	3,815	13.9
Centers					
Boston Children's Hospital	140,304	9,000	6,935	962	13.9
Cincinnati Children's Hospital	143,994	9,000	7,941	1,236	15.6
Children's Hospital of Philadelphia	209,755	9,000	8,439	1,039	12.3
Essentia Institute of Rural Health	243,092	4,500	4,129	680	16.5
Group Health Cooperative	217,959	9,000	8,973	1,429	15.9
Geisinger Health System	356,488	9,000	8,142	1,136	14.0
Mayo Clinic	136,391	9,000	8,443	1,983	23.5
Marshfield Clinic	134,212	4,500	4,280	811	18.9
Icahn School of Medicine at Mount Sinai	162,927	9,000	8,280	1,079	13.0
Northwestern University	206,554	9,000	8,521	1,380	16.2
Vanderbilt University	329,672	4,500	4,082	687	16.8
Vanderbilt Children's Hospital	107,814	4,500	4,074	578	14.2

[1] N indicates the number of eligible subjects that met the inclusion criteria. Patients with at least one inpatient or outpatient clinic visit at one of the participating sites confirmed to be between October 1, 2013 and September 1, 2014 were originally eligible. However, as is well known, electronic health record (EHR) data have a number of challenges and so data cleaning measures were taken that included: removing patients missing or with an invalid medical record number, removing duplicate patient identifiers, removing those with missing a household identifier (a center-specific anonymized address), removing those with a household identifier in common with at least 20 other patients (likely a clinic address and possibly a homeless shelter), and removing those with an address that could not be geocoded. Further, patients known to have died or were on a 'do not contact' list were excluded. Patients with missing age or gender and those with an age conflict (e.g., <18 years at an adult center, or ≥18 years at a pediatric center) were removed. Finally, we randomly selected one patient per household to be eligible for inclusion.

[2] Multiple surveys/reminders were sent to initially sampled subjects (if needed) and subjects were excluded if they did not have a valid address, were not currently living or previously participated in the pilot phase of the survey.

[3] Response rate = Responded/Sampled+Checks.

Table S2: Marginal distributions of stratification variables under random (RS) and maximum entropy sampling (MES) of 90,000 participants.

	RS, %	MES[1], %
Age Group		
<12, <35	34	48
>=12, >=35	66	52
Gender		
Female	56	52
Male	44	48
Race		
White	83	34
Black	9	20
Asian	2	16
American Indian/Alaska Native	1	5
Native Hawaiian/Pacific Islander	<1	4
Other	5	21
Ethnicity		
Not Hispanic/Latino	95	69
Hispanic/Latino	5	31
Education		
<HS	1	13
HS + Some college	76	53
>= Bachelors	23	35
Rurality		
Rural	43	37
Suburban/Urban	57	63

[1] Maximum entropy sampling was performed separately within each center to identify the number of individuals to sample from each stratum such that the Shannon entropy in demographic characteristics in the sample was maximized.[1; 2] This corresponds to, to the extent possible, minimizing the stratum to stratum variation in the number of subjects sampled across the possible 288 strata. It is worth noting that there were strata that were empty. Within each stratum, we then randomly sampled patients for inclusion into the study with complete (or non-imputed) stratification information and augmented with those subjects with imputed data when necessary. We recognize that the precise definition of the population and the sampling frame are imperfect. There are a number of reasons that include but are not limited to: 1) incomplete demographic data EHR data that varied from site to site, 2) misclassified EHR data: the extent of the misclassification is observable for those who responded to the survey but are not observable for those who did not; and 3) at all pediatric sites, stratum defining demographic data correspond to children even though it was the parents who answered the survey. We must therefore describe respondents as the 'population' of parents of children at a participating site. We further recognize that mothers or female guardians tended to respond to the surveys far more often than fathers or male guardians. We are currently investigating the impact of misclassification of survey stratification variables on regression estimates and associated inferences.

Table S3: Weight Truncation Sensitivity Analysis. Five trimming and redistribution rules were applied to post-stratified weights and used in separate survey-weighted regression models between biobank willingness and data-sharing and consent model. Regression coefficients and 95% linearized confidence intervals were transformed to the probability (or percentage) scale. Wald tests were used to evaluate data-sharing (broad-controlled, BC vs broad-open, BO) and consent (BC vs tiered-controlled, TC) comparisons.[1]

	Broad-controlled (BC)	**Broad-open (BO)**	**Tiered-controlled (TC)**	**All**	**BC vs BO X^2_1, p**	**BC vs TC X^2_1, p**
No trimming						
	12 (9,15)	14 (11,18)	11 (8,16)	13 (11,15)	2.08, 0.150	0.83, 0.362
	19 (16,24)	19 (14,25)	22 (18,25)	20 (17,23)		
	69 (63,74)	67 (61,73)	67 (62,71)	68 (64,71)		
95th Quantile						
	12 (10,14)	14 (12,17)	12 (9,15)	13 (11,15)	4.40, 0.036	1.63, 0.202
	19 (16,22)	20 (17,23)	22 (19,25)	20 (18,22)		
	69 (65,73)	66 (62,70)	66 (61,71)	67 (64,70)		
Maximum of median + 6*interquartile range and the 90th quantile						
	12 (10,14)	15 (13,18)	12 (9,15)	13 (12,15)	4.48, 0.034	1.07, 0.202
	20 (17,22)	21 (17,23)	22 (19,25)	20 (19,22)		
	68 (65,72)	65 (61,69)	66 (62,71)	66 (63,69)		
Median + 6*interquartile range						
	13 (11,15)	15 (13,18)	12 (10,15)	13 (12,15)	7.07, 0.008	0.67, 0.413
	20 (18,22)	21 (19,24)	22 (19,24)	21 (19,23)		
	68 (64,71)	64 (59,68)	66 (61,70)	66 (62,69)		
Median + 4*interquartile range						
	13 (11,15)	15 (13,18)	12 (10,15)	14 (12,15)	7.93, 0.005	0.37,0.542
	20 (18,23)	21 (19,24)	22 (19,24)	21 (19,23)		
	67 (64,71)	63 (59,67)	66 (62,70)	65 (62,69)		

[1] To account for the complex sampling design and the effects of survey non-response, post-stratified survey weights were computed. Because the design sought to enrich the sample with minority populations, sampling probabilities differed dramatically, and therefore there was substantial sampling weight variability within and among sites. We therefore conducted a weight trimming and redistribution scheme that was partly described in the Methods section. Even though weight trimming improves (lowers) variances, they are admittedly ad hoc. To address the tradeoffs associated with weight trimming, we conducted a sensitivity analysis to examine the impact of the trimming approach on the results from the primary analysis. Five trimming rules were considered: 1) no trimming, 2) trimming at the 95th quantile, 3) trimming at the maximum of a) the median + 6*interquartile range (IQR) and b) the 90th quantile, 4) trimming at the median + 6*IQR, and 5) trimming at the median + 4*IQR.[3; 4] These rules are ordered by the extent of total trimming in our data. For example, the range of trimmed weights using rule 2 was 1 to 3100, while the range of weights using rule 5 was 18 to 870. This Table summarizes the results of the marginal analysis of willingness (3-level; not willing, not sure, willing) and data-sharing and consent model each trimming rule. As can be seen, the trimming rule can indeed impact the statistical testing results if the 0.05 significance level is used. However, this is primarily due to sample size. Overall, the clinical conclusions remain largely unchanged, i.e., neither consent model nor data sharing model had a large impact on inclination to participate in a biobank.

Supplemental References:

1. 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.
2. Shannon, C.E. (2001). A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review 5, 3-55.
3. Chowdhury, S., Khare, M., and Wolter, K. (2007). Weight trimming in the national immunization survey. In Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, American Statistical Association. pp 2651-2658.
4. Potter, F. (1988). Survey of procedures to control extreme sampling weights. In Proceedings of the section on survey research methods, American Statistical Association. pp 453-458.