Title
Self-reported race and ethnicity of US biobank participants compared to the US Census

Journal
Journal of Community Genetics, 8(3)

ISSN
1868-310X

Authors
Cohn, Elizabeth Gross
Hamilton, Nalo
Larson, Elaine L
et al.

Publication Date
2017-07-01

DOI
10.1007/s12687-017-0308-6

Peer reviewed

J Community Genet (2017) 8:229–238
DOI 10.1007/s12687-017-0308-6

ORIGINAL ARTICLE

Self-reported race and ethnicity of US biobank participants compared to the US Census

Elizabeth Gross Cohn[1,2] · **Nalo Hamilton**[3] · **Elaine L. Larson**[1] · **Janet K. Williams**[4]

Received: 24 October 2016 / Accepted: 11 May 2017 / Published online: 16 June 2017

Abstract Precision medicine envisions a future of effective diagnosis, treatment, and prevention grounded in precise understandings of the genetic and environmental determinants of disease. Given that the original genome-wide association studies represented a predominately European White population, and that diversity in genomic studies must account for genetic variation both within and across racial categories, new research studies are at a heightened risk for inadequate representation. Currently biological samples are being made available for sequencing in biobanks across the USA, but the diversity of those samples is unknown. The aims of this study were to describe the types of recruitment and enrollment materials used by US biobanks and the diversity of the samples contained within their collection. Biobank websites and brochures were evaluated for reading level, health literacy, and factors known to encourage the recruitment of minorities, such as showing pictures of diverse populations. Biobank managers were surveyed by mail on the methods and materials used for enrollment, recruitment, consent, and the self-reported race/ethnicity of biobank participants. From 51 US biobanks (68% response rate), recruitment and enrollment materials were in English only, and most of the websites and brochures exceeded a fifth-grade reading level. When compared to the 2015 US Census, self-reported race/ethnicity of participants was not significantly different for Whites (61%) and blacks (13%). The percentages were significantly lower for Hispanics and Latinos (18 vs. 7%, $p = 0.00$) and Hawaiian/Pacific Islanders (0.2 vs. 0.01%; $p = 0.01$) and higher for Asians (13 vs. 5%, $p = 0.01$). Materials for recruitment predominantly in English may limit participation by underrepresented populations.

Keywords Biobanks · Minority · Health equity · Population health

✉ Elizabeth Gross Cohn
cohn@adelphi.edu

[1] School of Nursing, Columbia University, New York, NY, USA

[2] Adelphi University, Garden City, NY, USA

[3] School of Nursing, University of California, Los Angeles, Los Angeles, CA, USA

[4] University of Iowa, Iowa City, IA, USA

Introduction

Medical, ethical, legal, and social considerations require that investigators set targets for representation in clinical trials and registries, but enrollments often fail to meet these projections (Frieden and Centers for Disease C, Prevention 2011a, b). In precision medicine, this lack of representation can impede the evaluation of tools developed for clinical diagnosis (Burkardt et al. 2014) and limit the ability to gauge the efficacy of treatments (Lynch et al. 2014). In the past, underrepresentation has been attributed to historical transgressions (Corbie-Smith 1999), mistrust (Suther and Kiros 2009), personal preference (Sullivan et al. 2007), and lack of access (e.g., when patients have no access to the health care system, or patients and providers are either unaware of studies or unsure of how to enroll in them) (Gill et al. 2013). For over 20 years, the Food and Drug Administration (FDA) and the National Institutes of Health have advocated for greater inclusion in clinical trials of women (Foulkes 2011); racial, ethnic, and ancestral minority groups (Green et al. 2013); people with co-morbid or multimorbid conditions (Ritchie and Zulman 2013); and the elderly (Yoon et al. 2014). Given that the original genome-

wide association studies represented a predominately European white population, and that diversity in genomic studies must account for genetic variation both within and across racial categories, new studies are at a heightened risk for inadequate representation (Bustamante et al. 2011). In addition, the history of egregious incidents such as eugenics and Tuskegee in the name of "genetic research" in minority populations (Hoge and Appelbaum 2012; Roberts 2015) has laid down a foundation of mistrust and skepticism for minorities when considering enrollment in genetic and genomic studies.

Internationally, biobanks have taken various approaches to diversity and inclusion. In the UK, UK Biobank (http://www.ukbiobank.ac.uk/) is model for large-scale, population-based genomic research. Funded through a public-private partnership, with a goal of improving population health, the UK Biobank established 22 sites country-wide and enrolled community residents who voluntarily donated specimens for biomarkers and DNA research. As of June 2016, the UK Biobank reported the registration of participants who self-identified as white at just over 500,000 (reported by separate categories for British, ~470,000; Irish, ~14,000; and all other white backgrounds, ~17,000). For minority participants, ~10,000 self-identified as Asian or Asian British including Indian, Pakistani, and Bangladeshi and ~1700 self-reported as Chinese. Taken together, this represents a model of community engagement, but the participation of minorities remains low.

In an example of a merged model, the European Research Infrastructure Consortium (ERIC) in Graz, Austria, formed a consortium of 225 organizations spanning 30 countries, designed to provide a broad array of specimens. Unlike the UK Biobank, which used community-based samples, ERIC merges research and clinical biobanks with the pan-European Biobanking and Biomolecular Resources Research Infrastructure (http://bbmri-eric.eu/). ERIC has an overarching goal of interoperability of existing biobanks—some population-based and others clinically oriented—from different subpopulations. These collections include associated data on factors such as health status, nutrition, lifestyle, and environmental exposure. Data on biobank diversity are not available in the aggregate for the ERIC Consortium, making it difficult to determine the number or percent of minorities reflected in the biobank collection.

The USA is gearing up for increasing its efforts in large-scale biobanking. The publicly funded electronic *MEdical Records and GEnomics* (eMERGE) network, established in 2007 and expanded in 2011, is a model collaboration between the National Human Genome Research Institute (NHGRI) and nine academic medical centers (McCarty et al. 2011). Exploring both the utility and limitations of this network provides information about how the remainder of the nation's biobanks—not linked by this network—might benefit from its discoveries and be positioned to extend this model.

A recent review article described both the positive influence of eMERGE and how it might be improved (Crawford et al. 2014; Crosslin et al. 2014). According to that analysis, a critical limitation is the lack of representation of diverse racial/ethnic and ancestral groups (Rosenberg et al. 2010). The eMERGE investigators recommend strongly that future programs focus on intentional diversity in sampling (Crawford et al. 2014). Addressing the need to increase diversity within eMERGE has begun at three of the sites: Northwestern University, Vanderbilt University, and the Icahn School of Medicine at Mount Sinai (Siani 2013), but how this intentional effort is constructed and who the outreach and recruitment efforts should target remain open questions. As the USA prepares to launch a cohort study of a million or more participants, lessons learned from what we currently have may help to inform future efforts.

A majority of our nation's biobanks are not linked to a central system. Over the last decade, many hospital-based and private institutions have established biobanks in which to collect and store specimens that researchers can access with appropriate clearance and permissions. These samples provide the basis for the research being conducted in precision medicine in the USA to date (Bonham et al. 2016; Koretzky et al. 2016; Lu et al. 2014).

We sought to more fully describe the self-reported race and ethnicity of the participants providing the samples available in research biobanks across the nation. The aims of the study were to describe the content and presentation of recruitment and enrollment materials used by biobanks and the self-identified race/ethnicity of the participants.

Methods

Biobank definition

We used the National Heart, Lung, and Blood Institute definition of a biobank: "a collection of human specimens and associated data for research purposes, the physical structure where the collection is stored and all relevant process and policies." (National Heart, Lung, and Blood Institute. n.d.). "Open research biobanks" were defined as those in which investigators—independently or in collaboration with researchers within established systems and networks—could apply for access and conduct additional research on previously collected samples.

Procedure

The study, including the survey development and the national data collection, was approved by the IRB at Columbia University Medical Center.

Instrument development

Two instruments were developed for this study. First, to categorize the enrollment and recruitment materials, we developed a 15-item tool based on the extensive body of guidance literature for recruitment, enrollment, and consent. Our survey tool includes items identified by the Agency for Health Research and Quality (AHRQ) Health Literacy Universal Precautions Toolkit which describes best practices for health literacy, the Department of Health and Human Services 45 CFR 46.116 (Department of Health and Human Services 2009), which governs informed consent, and the principles described by Doak, Doak, and Root to enhance minority recruitment and develop materials which are accessible at all literacy levels (Doak et al. 1996). The development of this tool was guided by findings and principles emphasizing the importance of genomic literacy and the interest of the public towards genetic testing, biobanking for individuals, and facilitating care and treatment based on personalized medicine (Ashida et al. 2011; Hurle et al. 2013; Lea et al. 2011). These recommendations suggest the following:

- Pictures of diverse populations and clear messaging—what a biobank is, how it works, and the type of specimens that will be collected
- Built-in stimulation and motivation (i.e., anticipating and answering commonly asked questions such as why do people chose to participate)
- Attention to literacy level and complexity of information and the use of graphics including white space
- "Logic, language, and experience" considerations which include cultural appropriateness, tailored information, and authentic language translation

We piloted the tool by having two investigators independently rate a set of five websites and five brochures and then comparing results. Variation of more than one point was discussed among the raters, and when consensus was reached, a rule was created to guide future ratings.

Second, we developed a written survey for biobank managers to provide data which described the diversity of the biobank samples (based on self-reported race/ethnicity of the participants). To develop the survey, we met three times over 2 months with personnel of Feinstein Institute for Medical Research of the Northwell Health System, which was representative of the types of biobanks that we were soliciting. The group consisted of consent administrators, the clinical director, and several clinical staff members with whom we discussed and solidified the key topics and types of items that the survey should capture. Dillman's methodology (Dillman et al. 2014) was then used to develop a 15-question short-answer and multiple-choice survey with a narrative section to describe techniques for enrollment and recruitment.

Questions included the type of specimens collected (blood, saliva, tissue, urine, spinal fluid), total number of specimens and individuals enrolled in the biobank, and participants' self-reported race/ethnicity, including how the biobank captured that information. An open answer field was provided for respondents to describe their process of participant identification, recruitment, and enrollment. For example, biobanks may recruit from a geographic area, related to a hospital admission, or based on a specific disease.

A set of enrollment and recruitment materials were requested including a copy of the written informed consent form, a summary of recruitment strategies, and a blank copy of the participant enrollment or registration sheet. We asked for the enrollment or registration sheet to determine if the social determinants were addressed: for example, did the biobank collect information related to prolonged exposure to lead paint, if the participant grew up in a smoking household, or other exposures or factors known or thought to affect health.

To assure that the initial questions were clear and understandable, we tested the survey using a "think-aloud" technique (Dillman et al. 2014), in which a person with a role and profile similar to those who will be involved in the actual study completes the survey while talking aloud, commenting on each part, while the investigator observes but does not assist. This exercise assures that the questions are clear and concise and can stand on their own. For example, if the think-aloud participant reads the same question a few times, and then says, 'There are two answers here that would be correct for me but the instructions say I may choose only one, and I don't know what to pick" and then goes to the next question, the investigator would examine that item, clarify the selection or the instruction, and correct it for the final survey. Two think-alouds were conducted with biobank managers from Northwell who had not seen the survey, and items were amended based on their feedback. An example of an item that was amended was the addition of "spinal fluid" as an option in the type of materials collected. The results of the individuals who participated in the think-aloud were not included in the final data collection for the study.

Biobank identification

Since no central registry exists, it is difficult to ascertain the exact number of biobanks nationally. However, Henderson and colleagues (Henderson et al. 2013) examined 456 biobanks and categorized their organizational structure, affiliations, associations, and the type of biological specimen that the collections contained. Biobanks were identified through the search strategy based on methods previously employed by RAND and adapted for this study (Boyer et al. 2012). The search techniques reflected how a researcher interested in accessing samples from an existing research biobank might be expected to find such a resource. We searched Google and

Google Scholar using the search terms biobank, biorepository, genetic testing, whole genome, and whole-exome sequencing, the most common search terms for genomic work at the time of the study. We also reviewed NIH RePORTER for investigators funded for genomic sequencing with identified biobanks. Lastly, the agendas of major conferences were reviewed for speakers who were presenting results of studies that incorporated specimens from large biobanks. The search was conducted and the data collected between February 1, 2013, and January 10, 2015.

Biobank selection

Biobanks were included if they were located in the USA and contained samples that would be expected to reflect the diversity of the population and specimens were available for qualified researchers (qualified researchers were defined as someone who minimally has a terminal degree and the skills, knowledge, and ability to carry out appropriate independent research on samples). If biobanks were disease specific, to be included, they needed to contain specimens that would be expected to represent the population-at-large (for example, a biobank focused exclusively on Tay-Sachs would be excluded, whereas a biobank focused on heart failure would be included). Contact information for each biobank manager was recorded. Research assistants called each biobank to verify the name and contact information of the director, manager, or supervisor who could complete our survey.

When they agreed to participate, they received an information sheet, our 15-item survey tool, a tally sheet on which to report the self-reported race ethnicity and type of specimen, a self-addressed Federal Express envelope in which to return a copy of the recruitment materials, enrollment form, informed consent, and a $5.00 Starbucks gift card. As recommended by Dillman, reminder postcards were sent at 2 weeks, and phone calls were made at 4 weeks (Dillman et al. 2009). Those who returned the survey received a limited edition genomic art coffee mug.

Analysis

To assess the enrollment brochures and websites, we used the 13-item scale which incorporated items that were (1) shown to enhance minority recruitment (Green et al. 2013; Larkey et al. 2009; Lindenstruth et al. 2006) or (2) reflect factors critical to participants considering joining a biobank (Cadigan et al. 2014; Edwards et al. 2014; McDonald et al. 2014). Our tool recorded such items as the presence or absence of pictures of diverse participants; a simple language definition of a biobank; details of the type of samples collected; an explanation of DNA testing; the risks and benefits of participation; privacy protocols and safeguards; how to get additional information; details of any incentives; and the contact phone numbers and e-mails for follow-up. A sample of a website assessed is presented in Fig. 1 and our tool items as Table 1.

In addition to the content, we assessed reading levels of all materials. The premium version of Readability-Score.com (https://readability-score.com/) was used to measure the reading levels of websites and brochures. When websites had separate sections for patients and researchers, we evaluated the patient version. We calculated the Flesch-Kincaid Score, the Flesch-Kincaid Grade Level, the Gunning Fog Score, and the Simple Measure of Gobbledygook (SMOG) Index. Because the scores on all tools were closely correlated, we report the Flesh-Kincaid Score (where the lower the number, the more difficult the reading level; scale 0–100) and the Flesh-Kincaid Grade Level, which translates the score into the grade level equivalent for schools based in the USA.

Survey data on self-reported race ethnicity were analyzed using StatPac 4.s Statistics program (StatPac, Inc. Bloomington, MN). Descriptive data, frequencies, and cross-tabulations are presented with percentages where appropriate. Representativeness was determined as a percentage: the number from a specific particular ethnic group divided by overall number in that biobank. The general population percentage for ethnic groups was determined from US Census data of 2015. We compared the enrollment in biobanks for the US Census for city, state, and nation (some percentages do not add to 100 due to rounding). When biobanks collected more than one type of specimen from a single participant, the total reflects the number of participants, not the number of specimens. We used a two-sample t test to explore the difference between the representation in biobanks and in the US Census, overall and by city and state.

Results

Our national search identified 550 eligible biobanks. Biobanks associated with multicenter trials (280) were eliminated because their specimens were not available to outside researchers. We could verify contact information for 142 of the remaining 270 biobanks. Seventy-five biobanks agreed to receive the survey, and 51 returned it completed (68% response rate; Fig. 2).

We separated each biobank into a category type: (1) affiliated with an academic medical center, (2) affiliated with a hospital, (3) disease-specific, and (4) independent (unaffiliated) biospecimen repositories designed for research only. When the categories were not mutually exclusive, we assigned the biobank to the larger category. For example, if a disease-specific biobank was located within an academic medical center, we characterized it as an academic medical center biobank.

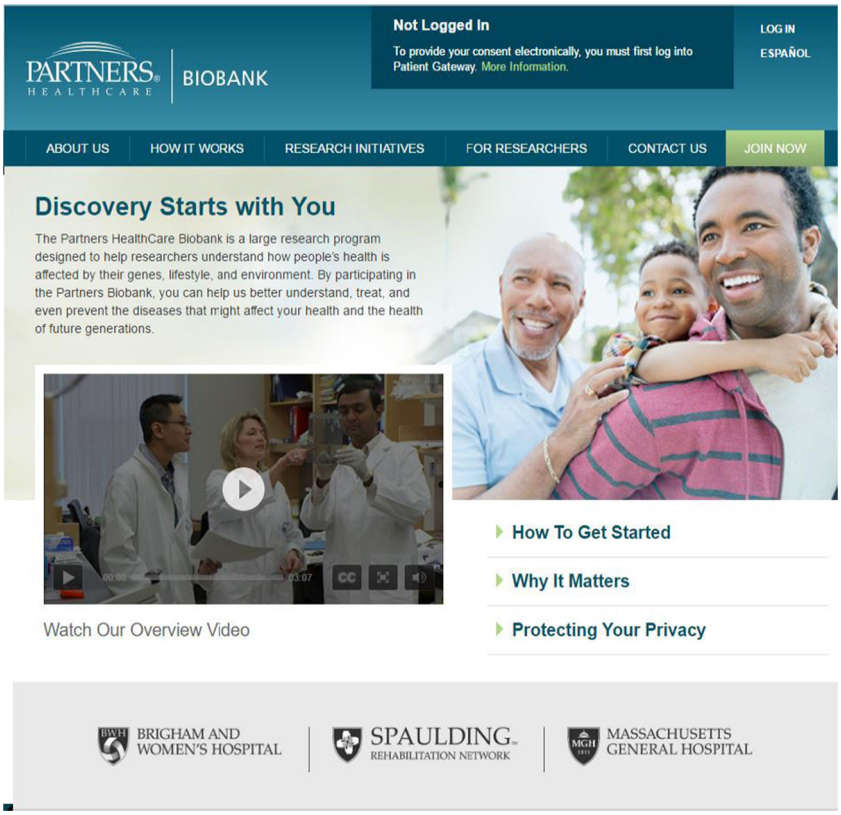

Fig. 1 Website for Partners HealthCare Biobank, a repository of consented patient samples and data at Partners HealthCare System (parent organization of Massachusetts General Hospital and Brigham and Women's Hospital). Reproduced with permission

Our study reports on 10,044,388 individual participants who provided specimens that were identified and catalogued. The sample reflected biobanks from across the USA; most were affiliated with academic medical centers and located on the East Coast and in the Midwest (Table 2). However, we report the results of enrollment in biobanks relative to the national data because most of our samples are from academic medical centers that draw a population greater than their immediate catchment area. The majority of the biobanks studied collected both blood and tissue samples. Cheek swab saliva, kidney biopsy, and umbilical cord blood were collected, depending on the disease and type of biobank (Table 2). Though

Table 1 Incidence of criteria of biobank websites and brochures from Federal Regulations 45 CFR 46.116 and AHRQ Health Literacy Universal Precautions Toolkit

Criterion	Websites ($N = 51$)	Websites (%)	Brochures ($N = 35$)	Brochures (%)	Criterion source
Listed contact phone numbers	47	92	32	91	45 CFR
Listed contact e-mail addresses	35	68	30	85	45 CFR
Describes the type of specimens collected	32	63	25	71	45 CFR
Outlines benefits to society	29	57	20	57	45 CFR
Outline privacy protocols/safeguards	23	45	16	45	45 CFR
Describes a biobank	20	39	15	43	45 CFR
Outlines benefits of participation	20	39	15	43	45 CFR
Outlines risks of participation	16	31	7	2	45 CFR
Pictures show a diverse group of participants	12	23	8	22	AHRQ
Explains DNA testing	10	19	2	5	45 CFR
Patient quotes	7	13	2	5	AHRQ
Offer an incentive to participation	6	11	4	11	45 CFR
Reading at or below a 5th-grade level	5	10	18	51	AHRQ

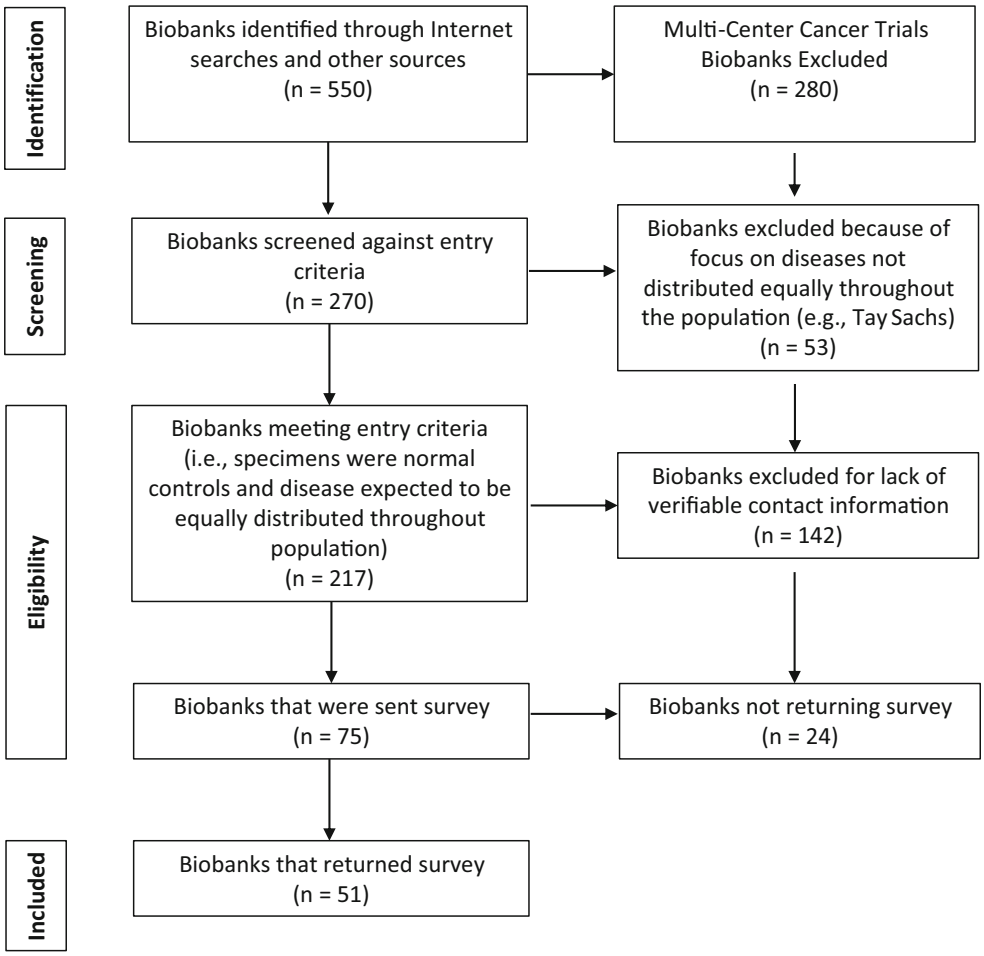

Fig. 2 Search strategy for biobanks

genetic material should be the same for an individual no matter the source, we report these findings for a more complete description of the biobanks. When there was a question about the number of specimens, as in cases where

Table 2 Characteristics of biobanks by location, affiliation, and specimen type

Number of participating biobanks	Total	*N* = 51	Percent
Location	Northeast	13	25
	Midwest	12	23
	West	12	23
	Southwest	7	13
	Southeast	7	13
Affiliation[a]	Academic medical center	32	62
	Independent research facility	17	33
	Hospital-based (non-academic)	7	13
	Disease-specific	5	10
Type of specimen	Blood	25	49
	Tissue	14	27
	Saliva	5	10
	Umbilical cord blood	3	5
	Postmortem fluids	4	7
	Kidney biopsy	3	5

[a] Not exclusive categories; total >100%

Table 3 US biobank sample compared to the US 2015 Census

Race/ethnicity	Individual biobank enrollment	Percent	US Census 2015	Percent	*p* value
White	5,681,516	59	196,065,480	61	.66
Asian	1,226,198	13	16,070,941	5	.04
Black/African American	1,185,105	12	41,784,446	13	1.0
Hispanic/Latino	708,550	7	57,855,387	18	.01
Native American	132,975	1	32,141,882	1	.94
Hawaiian/Pacific Islander	481,041	0.005	6,428,376	0.2	Unable
Multiple	214,517	2		2	1.0
Unknown	414,486	6	–	–	Unavailable
Totals	10,044,388	100	~321,418,820[a]	100	

[a] Numbers are correct but do not add due to rounding

multiple specimens were collected from the same participant, we called to verify the number of individuals and used that as our denominator (Table 2.)

Recruitment and enrollment materials were evaluated and categorized using a 13-item tool (Table 3). Pictures of a diverse group of participants were included on just under a quarter of the websites and the brochures. The function of a biobank was described in fewer than half of the websites and brochures. The type of specimens collected was described by more than half the websites and almost two thirds of the brochures reviewed. The benefits to society were described by over half of the websites and brochures. Most of the material was written above the recommended fifth-grade reading level.

Contact information was readily available on most websites and brochures. E-mail addresses were frequently found on biobank brochures, and links to more information were found on the majority of the websites. Answers to the most frequently asked questions were addressed on both website and brochures, but more frequently on brochures. An explanation of DNA testing and actual patient quotes were seen more frequently on websites and less frequently in the brochures. Individual risks were described on less than half of websites and brochures; however, the risks were more often described on websites than in brochures. Incentives for participation were few and varied from monetary incentives to cremation of the deceased person's remains (Table 1).

Compared with US Census data, whites, blacks, and Native Americans were represented in accordance with their representation in the US population. Asians were overrepresented when compared to their US Census percentage in the population. Hispanics and Latinos were represented in biobanks at significantly less than their representation in the population. Hawaiian/Pacific Islanders were represented in numbers too small to be properly analyzed relative to the population (Table 3).

Discussion

For biobanks to advance science equitably, their enrollment should reflect the rich diversity of the populations that they seek to represent. To do so, it is critical to understand who is currently represented and what groups might be at risk for underrepresentation. In this national survey, biobanks were primarily located in academic health centers in the northeast, east, and west regions of the USA. Hispanics/Latinos and Pacific Islanders were underrepresented as biobank participants, but whites and blacks were not. These findings counter the widely held belief that blacks are underrepresented (Bustamante et al. 2011; Knerr et al. 2011). Our previous work suggested that self-identified African Americans in Harlem, New York (and other minorities), may be open to participation in genetic and genomic research with proper safeguards in place and culturally and historically appropriate recruitment (Cohn et al. 2014; Green et al. 2013). The rate of enrollment in the biobanks that we studied suggested that the perceived barriers for recruitment of these populations can be overcome and higher enrollment can be achieved. We found a significantly lower participation rate of Hispanics/Latinos and Hawaiian/Pacific Islanders, as has also been described in clinical trials of stroke (Nakagawa et al. 2013), depression (Martin et al. 2013), cancer (Ashing-Giwa and Rosales 2012), and diabetes (Toobert et al. 2011). There are several junctures in the process that may account for this finding. Self-report or self-identification may not sufficiently capture those identifying as both white and Hispanic or Latino (Eisenhower et al. 2014), or there may be insufficient recruitment and enrollment of this population (Martin et al. 2013; Tenorio et al. 2014). A tailored focus on recruitment and enrollment, such as work being done at the University of Miami, El Centro, a National Institutes of Health—National Institute on Minority Health and Health Disparities Centers of Excellence—focused on improving health equity among groups of Hispanic and African descent and Caribbean and Latin American nations,

demonstrates the effectiveness of a personalized approach as one possible means to increase representation of this population (Liu et al. 2016; Madhivanan et al. 2016; Messiah et al. 2015).

Among biobanks assessed in this study, Asians contributed a significantly higher percentage of biobank specimens than the general population. Although Census categories are too large and diverse to be meaningful in general, this is especially the case for the diversity within the Asian category. Therefore, this outcome could be attributed to the inclusion of respondents from the Far East, Southeast Asia, or the Indian subcontinent, including at least ten countries varying in geography and culture. All Census categories may reflect a wider range of ancestral and ethnic variations than is practical for understanding and interpreting participant representation, this one included. We also noted the existence of a biobank specifically focused on Asian participants, which enrolled exclusively Asians with a group of race-concordant researchers and clinicians. This finding highlighted the potential effectiveness of specialized practices and race concordance in overcoming some of the barriers to enrollment (Evans 2004).

The educational materials available can adversely affect patient enrollment (or deposit of specimens into) in biobanks. The brochures that we received were only in English. Providing brochures in languages other than English, targeted and tailored, would be an important step in recruitment of Hispanics and Latinos. Additionally, despite foundational work on developing principles of genetic and functional health literacy (Bonham et al. 2009; Coleman et al. 2014; Hurle et al. 2013; Lachance et al. 2010; Modell et al. 2014, 2016) and the wide availability of guidance documentation (Louis et al. 2014; Marcantoni et al. 2014), enrollment and recruitment websites and brochures remain at a reading level too high for a large portion of the general public. Even basic strategies such as displaying pictures of diverse groups (Doak et al. 1996) appeared on only one fourth of websites and brochures. Currently, the use and effectiveness of translated and culturally relevant materials are being examined and tested by the National Cancer Institute as part of an integrated approach of tailored education and social media (http://www.cancer.gov/cancertopics/disparities/mmo). We note that some biobank websites offer Google Translator; for example, the Mayo Clinic offers automatic website translation to Spanish, Portuguese, Arabic, and Mandarin, whereas others did not. The use of automated translation programs for complex scientific concepts has been subject to criticism due to the loss of nuance and the likelihood that a literal translation may not represent the conceptual nature of the science and the complexity of the ideas being presented. However, it is an option that can be explored in terms of translation (Munday 2016) since such approaches may improve access to materials and information as a necessary first step in enrollment. The areas where brochures fell short of websites, such as risks of participation, patient quotes, and a definition of DNA testing, may put those with fewer computer skills or access, at increased risk for not receiving that information.

Limitations

Though we conducted a thorough search, we may have missed biobanks that would have met inclusion criteria but did not have a website causing them to be overlooked. We could not verify contact information for a large number of biobanks, and therefore non-response bias is likely. This was a descriptive study of a one-time national sample of biobanks. Reasons for the varying representation of ethnic groups require further investigation.

Conclusions

This study suggests that Hispanics and Latinos are the most at risk for underrepresentation in open research biobanks in the USA. Translation of enrollment and recruitment materials into languages other than English and reducing the language reading level on websites and in brochures are some immediate steps that may increase participation in biobanking.

Acknowledgements Funding was provided by the Robert Wood Johnson Foundation Nurse Faculty Scholars Program Grant #72375 and National Human Genome Research Institute grant NHGRI P50 HG007257 and the National Institutes of Clinical and Translational Science UL1RR024156. The authors would like to thank the faculty of the Feinstein Institute for Medical Research at Northwell Health System, Dr. Peter Gregersen, and Marlena Kern for guidance in the development of the study.

Compliance with ethical standards All procedures were followed in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants for being included in this study.

Conflict of interest The authors declare that they have no conflict of interest.

References

Ashida S, Goodman M, Pandya C, Koehly LM, Lachance C, Stafford J, Kaphingst KA (2011) Age differences in genetic knowledge, health literacy and causal beliefs for health conditions. Public Health Genomics 14:307–316. doi:10.1159/000316234

Ashing-Giwa K, Rosales M (2012) Recruitment and retention strategies of African American and Latina American breast cancer survivors in a longitudinal psycho-oncology study. Oncol Nurs Forum 39:E434–E442. doi:10.1188/12.ONF.E434-E442

Bonham VL, Citrin T, Modell SM, Franklin TH, Bleicher EW, Fleck LM (2009) Community-based dialogue: engaging communities of color

in the United States' genetics policy conversation. J Health Polit Policy Law 34:325–359. doi:10.1215/03616878-2009-009

Bonham VL, Callier SL, Royal CD (2016) Will precision medicine move us beyond race? N Engl J Med 374:2003–2005. doi:10.1056/NEJMp1511294

Boyer GJ, Whipple W, Cadigan RJ, Henderson GE (2012) Biobanks in the United States: how to identify an undefined and rapidly evolving population. Biopreservation Biobanking 10:511–517. doi:10.1089/bio.2012.0034

Burkardt DD, Graham JM Jr, Short SS, Frykman PK (2014) Advances in Hirschsprung disease genetics and treatment strategies: an update for the primary care pediatrician. Clin Pediatr 53:71–81. doi:10.1177/0009922813500846

Bustamante CD, Burchard EG, De la Vega FM (2011) Genomics for the world. Nature 475:163–165. doi:10.1038/475163a

Cadigan RJ, Juengst E, Davis A, Henderson G (2014) Underutilization of specimens in biobanks: an ethical as well as a practical concern? Genet Med : Off J Am Coll Med Genet 16:738–740. doi:10.1038/gim.2014.38

Cohn EG, Husamudeen M, Larson EL, Williams JK (2014) Increasing participation in genomic research and biobanking through community-based capacity building. J Genet Couns. doi:10.1007/s10897-014-9768-6

Coleman B et al (2014) Multi-ethnic minority nurses' knowledge and practice of genetics and genomics. J Nurs Scholarsh : Off Publ Sigma Theta Tau Int Honor Soc Nurs / Sigma Theta Tau 46:235–244. doi:10.1111/jnu.12083

Corbie-Smith G (1999) The continuing legacy of the Tuskegee Syphilis Study: considerations for clinical investigation. Am J Med Sci 317:5–8

Crawford DC et al (2014) eMERGEing progress in genomics-the first seven years. Front Genet 5:184. doi:10.3389/fgene.2014.00184

Crosslin DR et al (2014) Controlling for population structure and genotyping platform bias in the eMERGE multi-institutional biobank linked to electronic health records. Front Genet 5:352. doi:10.3389/fgene.2014.00352

Department of Health and Human Services (2009) CRF 46.116. Department of Health and Human Services. http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.116. Accessed October 5, 2014

Dillman DA, Smyth JD, Christian LM, Dillman DA (2009) Internet, mail, and mixed-mode surveys: the tailored design method, 3rd edn. Wiley & Sons, Hoboken

Dillman DA, Smyth JD, Christian LM (2014) Internet, phone, mail, and mixed-mode surveys : the tailored design method, 4th edn. Wiley, Hoboken

Doak C, Doak L, Root J (1996) Teaching patients with low literacy skills, 2nd edn. Lippincott, Philadelpia

Edwards T, Cadigan RJ, Evans JP, Henderson GE (2014) Biobanks containing clinical specimens: defining characteristics, policies, and practices. Clin Biochem 47:245–251. doi:10.1016/j.clinbiochem.2013.11.023

Eisenhower A, Suyemoto K, Lucchese F, Canenguez K (2014) "Which box should I check?": examining standard check box approaches to measuring race and ethnicity. Health Serv Res 49:1034–1055. doi:10.1111/1475-6773.12132

Evans M (2004) Healthcare's minority report. Sullivan Commission, IOM try to make patient, hospital staff makeup more reflective of the nation's ever-changing population. Mod Healthc 34(6–7):14–11

Foulkes MA (2011) After inclusion, information and inference: reporting on clinical trials results after 15 years of monitoring inclusion of women. J Women's Health 20:829–836. doi:10.1089/jwh.2010.2527

Frieden TR, Centers for Disease C, Prevention (2011a) Forward: CDC health disparities and inequalities report—United States, 2011. Morb Mortal Wkly Rep Surveill Summ 60(Suppl):1–2

Frieden TR, Centers for Disease C, Prevention (2011b) Public health then and now: celebrating 50 years of MMWR at CDC. Foreword Morb Mortal Wkly Rep Surveill Summ 60(Suppl 4):1

Gill PS, Plumridge G, Khunti K, Greenfield S (2013) Under-representation of minority ethnic groups in cardiovascular research: a semi-structured interview study. Fam Pract 30:233–241. doi:10.1093/fampra/cms054

Green MA, Kim MM, Barber S, Odulana AA, Godley PA, Howard DL, Corbie-Smith GM (2013) Connecting communities to health research: development of the Project CONNECT minority research registry. Contemp Clin Trials 35:1–7. doi:10.1016/j.cct.2013.01.001

Henderson GE et al (2013) Characterizing biobank organizations in the U.S.: results from a national survey. Genome Med 5:3. doi:10.1186/gm407

Hoge SK, Appelbaum PS (2012) Ethics and neuropsychiatric genetics: a review of major issues. Int J Neuropsychopharmacol / Off Sci J Collegium Internationale Neuropsychopharmacologicum 15:1547–1557. doi:10.1017/S1461145711001982

Hurle B, Citrin T, Jenkins JF, Kaphingst KA, Lamb N, Roseman JE, Bonham VL (2013) What does it mean to be genomically literate?: National Human Genome Research Institute Meeting Report. Genet Med : Off J Am Coll Med Genet 15:658–663. doi:10.1038/gim.2013.14

Knerr S, Wayman D, Bonham VL (2011) Inclusion of racial and ethnic minorities in genetic research: advance the spirit by changing the rules? J Law Med Ethics : J Am Soc Law Med Ethics 39:502–512. doi:10.1111/j.1748-720X.2011.00617.x

Koretzky M, Bonham VL, Berkman BE, Kruszka P, Adeyemo A, Muenke M, Hull SC (2016) Towards a more representative morphology: clinical and ethical considerations for including diverse populations in diagnostic genetic atlases. Genet Med : Off J Am Coll Med Genet. doi:10.1038/gim.2016.7

Lachance CR, Erby LA, Ford BM, Allen VC Jr, Kaphingst KA (2010) Informational content, literacy demands, and usability of websites offering health-related genetic tests directly to consumers. Genet Med : Off J Am Coll Med Genet 12:304–312. doi:10.1097/GIM.0b013e3181dbd8b2

Larkey LK, Gonzalez JA, Mar LE, Glantz N (2009) Latina recruitment for cancer prevention education via community based participatory research strategies. Contemp Clin Trials 30:47–54. doi:10.1016/j.cct.2008.08.003

Lea DH, Kaphingst KA, Bowen D, Lipkus I, Hadley DW (2011) Communicating genetic and genomic information: health literacy and numeracy considerations. Public Health Genomics 14:279–289. doi:10.1159/000294191

Lindenstruth KA, Curtis CB, Allen JK (2006) Recruitment of African American and white postmenopausal women into clinical trials: the beneficial effects of soy trial experience. Ethn Dis 16:938–942

Liu AY et al (2016) Preexposure prophylaxis for HIV infection integrated with municipal- and community-based sexual health Services. JAMA Intern Med 176:75–84. doi:10.1001/jamainternmed.2015.4683

Louis AJ, Arora VM, Press VG (2014) Evaluating the brief health literacy screen. J Gen Intern Med 29:21. doi:10.1007/s11606-013-2655-2

Lu YF, Goldstein DB, Angrist M, Cavalleri G (2014) Personalized medicine and human genetic diversity. Cold Spring Harbor Perspect Med 4:a008581. doi:10.1101/cshperspect.a008581

Lynch HT, Drescher K, Knezetic J, Lanspa S (2014) Genetics, biomarkers, hereditary cancer syndrome diagnosis, heterogeneity and treatment: a review. Curr Treat Options in Oncol 15:429–442. doi:10.1007/s11864-014-0293-5

Madhivanan P, Valderrama D, Krupp K, Ibanez G (2016) Family and cultural influences on cervical cancer screening among immigrant Latinas in Miami-Dade County, USA. Cult Health Sex 18:710–722. doi:10.1080/13691058.2015.1116125

Marcantoni JR, Finney K, Lane MA (2014) Using health literacy guidelines to improve discharge education and the post-hospital

transition: a quality improvement project. Am J Med Qual : Off J Am Coll Med Qual 29:86. doi:10.1177/1062860613508905

Martin A, Negron R, Balbierz A, Bickell N, Howell EA (2013) Recruitment of black and Latina women to a randomized controlled trial. J Health Care Poor Underserved 24:1102–1114. doi:10.1353/hpu.2013.0125

McCarty CA et al (2011) The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genet 4:13. doi:10.1186/1755-8794-4-13

McDonald JA et al (2014) Intentions to donate to a biobank in a national sample of African Americans. Public Health Genomics 17:173–182. doi:10.1159/000360472

Messiah A et al (2015) Combining community-based participatory research (CBPR) with a random-sample survey to assess smoking prevalence in an under-served community. J Natl Med Assoc 107:97–101. doi:10.1016/S0027-9684(15)30030-4

Modell SM, Citrin T, King SB, Kardia SL (2014) The role of religious values in decisions about genetics and the public's health. J Relig Health 53:702–714. doi:10.1007/s10943-013-9814-y

Modell SM, Greendale K, Citrin T, Kardia SL (2016) Expert and advocacy group consensus findings on the horizon of public health genetic testing. Healthcare:4. doi:10.3390/healthcare4010014

Munday J (2016) Introducing translation studies: theories and applications Routledge, New York

Nakagawa K, Koenig MA, Asai SM, Chang CW, Seto TB (2013) Disparities among Asians and native Hawaiians and Pacific Islanders with ischemic stroke. Neurology 80:839–843. doi:10.1212/WNL.0b013e3182840797

Ritchie CS, Zulman DM (2013) Research priorities in geriatric palliative care: multimorbidity. J Palliat Med 16:843–847. doi:10.1089/jpm.2013.9491

Roberts D (2015) Can research on the genetics of intelligence be "socially neutral"? Hast Cent Rep 45:S50–S53. doi:10.1002/hast.499

Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M (2010) Genome-wide association studies in diverse populations. Nat Rev Genet 11:356–366. doi:10.1038/nrg2760

Siani M (2013) Mount Sinai Researchers will use $2.9 million nih grant to study disease in ethnically-diverse populations. Researchers will tap the Mount Sinai BioBank Program, BioMe™, which includes Electronic Medical Records on 29,000 patients. http://www.mountsinai.org/about-us/newsroom/press-releases/mount-sinai-researchers-will-use-29-million-nih-grant-to-study-disease-in-ethnically-diverse-populations. Accessed August 10, 2014

Sullivan PS, McNaghten AD, Begley E, Hutchinson A, Cargill VA (2007) Enrollment of racial/ethnic minorities and women with HIV in clinical research studies of HIV medicines. J Natl Med Assoc 99:242–250

Suther S, Kiros GE (2009) Barriers to the use of genetic testing: a study of racial and ethnic disparities. Genet Med 11:655–662. doi:10.1097/GIM.0b013e3181ab22aa

Tenorio SL, O'Donnell CI, Hernandez J, Rozjabek HM, Lynch D, Marcus PM (2014) Culturally sensitive approaches to recruitment and retention of Hispanics in the national lung screening trial. J Immigr Minor Health / Center Minor Public Health 16:761–764. doi:10.1007/s10903-013-9862-0

Toobert DJ, Strycker LA, King DK, Barrera M Jr, Osuna D, Glasgow RE (2011) Long-term outcomes from a multiple-risk-factor diabetes trial for Latinas: inverted exclamation markViva Bien! Transl Behav Med 1:416–426. doi:10.1007/s13142-010-0011-1

Yoon J, Zulman D, Scott JY, Maciejewski ML (2014) Costs associated with multimorbidity among VA patients. Med Care 52(Suppl 3):S31–S36. doi:10.1097/MLR.0000000000000061