Published in final edited form as:
Mayo Clin Proc. 2013 September ; 88(9): 952–962. doi:10.1016/j.mayocp.2013.06.006.

The Mayo Clinic Biobank: A building block for individualized medicine

Janet E. Olson, Ph.D.[1,#], **Euijung Ryu, Ph.D.**[1,#], **Kiley J. Johnson, M.S., CGC**[2], **Barbara A. Koenig, Ph.D.**[3], **Karen J. Maschke, Ph.D.**[4], **Jody A. Morrisette, M.B.A.**[1], **Mark Liebow, M.D.**[5], **Paul Y. Takahashi, M.D., M.P.H.**[5], **Zachary S. Fredericksen, M.S.**[1], **Ruchi G. Sharma, B.S.**[1], **Kari S. Anderson, B.S.**[1], **Matthew A. Hathcock, M.S.**[1], **Jason A. Carnahan, B.S.**[2], **Jyotishman Pathak, Ph.D.**[1], **Noralane M. Lindor, M.D.**[1], **Timothy J. Beebe, Ph.D.**[1], **Stephen N. Thibodeau, Ph.D.**[6], and **James R. Cerhan, M.D., Ph.D.**[1]

[1]Department of Health Sciences Research, Mayo Clinic

[2]Center for Individualized Medicine, Mayo Clinic

[3]University of California, San Francisco

[4]The Hastings Center, Garrison, NY

[5]Department of Medicine, Mayo Clinic

[6]Department of Laboratory Medicine and Pathology, Mayo Clinic

Abstract

OBJECTIVE—To report the design and first three years of enrollment of the Mayo Clinic Biobank.

PATIENTS AND METHODS—Preparations for this Biobank began with a 4-day Deliberative Community Engagement with local residents to obtain community input into the design and governance of the biobank. Recruitment, which began in April 2009, is ongoing with a target goal of 50,000. Any Mayo Clinic patient who is 18+ years, able to consent, and a US resident is eligible to participate. Each participant completes a health history questionnaire, provides a blood sample and allows access to existing tissue specimens and all data from their Mayo Clinic medical record (EMR). A Community Advisory Board provides ongoing advice and guidance on complex decisions.

RESULTS—After three years of recruitment, 21,736 subjects have enrolled. Participants were 58% female, 95% of European ancestry, and median age of 62 years. Seventy-four percent lived in

Reprints and correspondence: Janet E. Olson, Ph.D; Mayo Clinic, 200 First Street SW, 6-241 Charlton Building , Rochester, MN 55905; olsonj@mayo.edu.
[#]The listed authors are co-first authors.

Conflicts of interest: None

Minnesota, 42% from Olmsted County where the Mayo Clinic Rochester is located. The five most commonly self-reported conditions were hyperlipidemia (41%), hypertension (38%), osteoarthritis (30%), any cancer (29%), and gastroesophageal reflux disease (26%). Among self-reported cancer patients, the five most common types were non-melanoma skin cancer (14%), prostate cancer (12% in men), breast cancer (4%), melanoma (3%), and cervical cancer (2% in women). Fifty-six percent of participants had at least 15 years of EMR history. To date, over sixty projects and over 69,000 samples have been approved for use.

CONCLUSION—The Mayo Clinic Biobank has quickly been established as a valuable resource for researchers.

Objective

The Mayo Clinic has a long, rich history of biological specimen and data banking. Since 1907, Mayo Clinic has archived all tissue slides and formalin-fixed, paraffin-embedded blocks and made them available to clinicians to assist in the care of the patient and to researchers with appropriate institutional approval to further understanding of disease processes.[1] This rich history of sample collections has been enhanced over the years with the addition of many collections of disease-specific samples, including samples from patients with various cancer types, neurological diseases, and various cardiovascular conditions. Although these clinical specimens have enabled countless research studies and advanced the understanding of disease pathogenesis, there is increasing scientific need for large biobank collections so that new discoveries can be made and validated.

Many large-scale biobanking efforts are underway worldwide. When linked with subject data from questionnaires and/or medical records, biobanks can serve as valuable resources in the study of complex diseases.[2] Examples of these biobanks include the Personalized Medicine Research Project at Marshfield Clinic, Wisconsin [3], BioVu at Vanderbilt University School of Medicine, Nashville, Tennessee[4], Kaiser Research Program on Genetics, Environment and Health, Oakland, California[5], the Icelandic database by deCODE genetics[6], the Generation Scotland Biobank[7] and the UK Biobank.[8]

Following several years of planning, in 2009, the Mayo Clinic Center for Individualized Medicine initiated a large scale biorepository named the Mayo Clinic Biobank. The goal of the Mayo Clinic Biobank (http://mayoresearch.mayo.edu/mayo/research/biobank/index.cfm) was to support a wide array of health-related research studies, especially those with the potential to improve patient care, by recruiting 50,000 subjects. Unlike the clinical biobank paradigm described by Zielhuis[9], the Mayo Clinic Biobank is not focused on a specific disease category. Its original goals were to provide a source of controls for disease-specific registries at Mayo Clinic, create a cohort that could be studied, and support nested case-controls studies with pre-disease specimens. We report here the design and implementation of the Mayo Clinic Biobank, including data from subjects recruited in the first three years of this ongoing project.

Patients and Methods

Deliberative Community Engagement

Before initiating the Mayo Clinic Biobank, we conducted a 4-day Deliberative Community Engagement Project (DCE) in September of 2007. The DCE was designed to allow citizen input into biobank design and to assure that community values were taken into account. To achieve diverse perspectives, we selected 20 lay members of the Olmsted County, Minnesota community who varied by age, sex, social and economic status, race, ethnicity, and employment. Deliberants were provided a briefing book presenting background information on biobanking issues. At the DCE event, patient advocates, potential biobank researchers, experts on biobanking procedures, human subjects protection staff, and privacy experts made presentations and answered citizens' questions. With the aid of professional facilitators, deliberants then met in small and large groups to develop recommendations. The community members made recommendations about biobank procedures and suggested guiding principles, including: the need for strong privacy protections, convenient recruitment, the importance of data sharing, limited options for return of research results, the importance of long-term community oversight, and an easy-to-understand consent document. The DCE followed an established format that has been used previously in developing biobank design and governance informed by community values and perspectives[10–13]. A key outcome of the DCE was the deliberant's recommendation that Mayo Clinic establish the Biobank Community Advisory Board (described below). Additional documentation concerning the DCE can be found on the Mayo Clinic Biobank website.

Development of the informed consent process

The Mayo Clinic Biobank is an opt-in biobank for which we obtain written informed consent. The informed consent document was developed by the Mayo Clinic Biomedical Ethics Research Unit based on guidance obtained during the DCE. A core group of community deliberants offered advice and reviewed drafts. Open communication, simplicity, and brevity were the primary values espoused. Final details of the informed consent process were developed in collaboration with the Biobank Community Advisory Board (described below). Final review and approval was obtained from the Mayo Institutional Review Board to ensure that the informed consent process met all legal requirements. Biobank participants agree to permit use of samples and/or data in multiple studies, provide access to questionnaire and data from the medical record (including past and future), provide blood samples drawn specifically for the Mayo Clinic Biobank, access to stored clinical samples, and permit to share de-identified data with other researchers through secured computer databases such as dbGAP (http://www.ncbi.nlm.nih.gov/gap). Participants agree to future contact for additional studies. The consent document discusses privacy protections and risks involved in participating and discusses the potential for receiving results from projects that use the Mayo Clinic Biobank. It also provides two checkboxes (included at the suggestion of the community members) allowing the participant the option of: 1) not allowing access to stored clinical specimens for research, and 2) not allowing family members access to samples after their death. The current consent document is available at the Mayo Clinic Biobank website.

Eligibility and Recruitment

Patients at Mayo Clinic, who are 18 years of age or older, able to communicate in English, have mental capacity to consent, and are residents of the United States are eligible for the Mayo Clinic Biobank. Study materials are provided in English only. Recruitment to the Mayo Clinic Biobank is primarily conducted via mailed invitation to patients who are scheduled for an appointment in a subset of Mayo Clinic departments and divisions. These currently include General Internal Medicine, Primary Care Internal Medicine, Family Medicine, and Preventive Medicine and the specialty areas of Obstetrics/Gynecology, and Executive Health. These departments were chosen to enhance selection of people who have Mayo Clinic Rochester as their source of primary or secondary care and to include those with a wide range of medical issues. While recruitment was initiated in Mayo Clinic Rochester, it was expanded to Mayo Clinic Florida in June of 2012.

Biobank staff mail a packet two to three weeks before a scheduled medical appointment that includes general information about the Biobank, an informed consent document, a baseline health history questionnaire, a choice of $20 incentives, and directions on how to provide a blood sample. Those who do not respond are telephoned after 14 days, unless the date of the appointment has already passed. Individuals not specifically invited through the mail may also enroll in the Biobank if they meet eligibility criteria. Blood samples are obtained by venipuncture at any Mayo Clinic outpatient lab during a visit or at a time preferred by the participant.

Baseline Health History Questionnaire

The baseline health history questionnaire was developed after a systematic review of similar questionnaires from institutions across the U.S. The general domains represented in the exemplar instruments were reviewed and considered for inclusion by the study team and key stakeholders. The final questionnaire includes general health questions from the SF-36[14], a series of linear analogue self-assessment well-being questions[15], a subset of the LOT-R measures of optimism[15], social support[16], depression[17], personal and family medical history questions, reproductive history, health behaviors, a limited series of dietary questions, physical activity[18], tobacco use, vitamins/supplements, environmental exposures, and demographics questions. A copy of the current version is available at the Mayo Clinic Biobank website. Questionnaires are returned via the U.S. mail or at a visit to Mayo Clinic. All questionnaires are visually scanned for errors and omissions when received from participants. Those with more than 10 errors are returned to the participant for correction. Data are programmatically evaluated for logic errors, incorrect skip patterns and other errors and visually compared to the original document for verification.

Biological Specimens

Participants in the Mayo Clinic Biobank are asked to provide a blood sample. The standard blood collection includes three 10 milliliter (ml) tubes with EDTA, a 10 ml tube without additives, and a 4.5 ml tube with sodium citrate (Na Cit). In a 10% random sample, a sodium heparin tube is substituted for an EDTA tube to permit processing into slow-frozen white blood cells (WBC). Blood samples are centrifuged at 2000 Relative Centrifugal Force (rcf) for 10 minutes and fractionated by automation (Perkin Elmer Janus robotic liquid handlers).

DNA is extracted from 4 ml of EDTA blood on an AutoGen (Holliston, MA) STAR DNA extraction machine and quantitated by Trinean (Belgium) instrumentation and cDrop software to determine both double stranded and total DNA amounts. Finally, DNA samples are volume checked to determine precise quantities extracted. All sample attributes are recorded. Unique identification numbers are assigned to each individual container received or created, and barcoded labels are adhered to each container. Patient demographics, sample attributes and storage locations are recorded and can be exported via reports or to databases. All demographics are disassociated from the sample record after 3 days to make patient information secure.

Output from the standard blood collection includes three aliquots of WBC (two in heparin and one in citrate), one to two DNA extractions (median 183 ug), 12 aliquots of EDTA platelet poor plasma (6 of 1000 ul, 6 of 500 ul), six aliquots of serum (1 of 1000 ul, 5 of 500 ul), and three aliquots of citrated plasma (1 of 1000 ul, 2 of 500 ul). Once created, samples are stored in a Nexus robotic −80°C high capacity freezer in 1.4 ml Matrix vials (Thermo Fisher Scientific, Hudson, New Hampshire). Samples can then be retrieved robotically upon demand and organized in a plating layout to be used with Perkin Elmer (Waltham Massachusetts) plating liquid handlers. Quality control procedures are available upon request.

Linkage to electronic patient data

A major source of data on Mayo Clinic Biobank participants is the Mayo Clinic medical record. A unified outpatient and inpatient medical record system has been in place at Mayo Clinic since 1907 [19]; the format transitioned from paper to electronic beginning in 1994 and has been completely paperless since 2004. Included in the electronic medical record (EMR) are all data from clinician notes (free text or annotated by natural language processing techniques[19]), diagnoses (using ICD-9 codes), laboratory measurements (including values and reference ranges), medical procedures (using CPT-4), medications (using RxNorm[20,21]), medical images (using DICOM[22]), vital signs, registration data and pathology (using SNOMED[23]). These data are available to the Mayo Clinic Biobank via the Mayo Clinic Enterprise Data Trust [19]. Paper records can be retrieved for approved studies.

As mentioned above, routinely reported medication data are available through the EMR for Mayo Clinic Biobank participants. These data include prescribed medications, over-the-counter medications and dietary supplements. To describe drug intake pattern for Mayo Clinic Biobank participants, data were extracted from all patient visits from the date of enrollment through July 2012. Medication groups were created mainly based on the National Drug File – Reference Terminology (NDF-RT) classification system. The details of the grouping and application to the 2009 Rochester Epidemiology Project (REP) can be found elsewhere[20,21].

Community Advisory Board

The Mayo Clinic Biobank has always sought advice from the local community. After the completion of the Deliberative Community Engagement Project described above, half of the participants agreed to become members of a standing Community Advisory Board (CAB)

for the Mayo Clinic Biobank. Supplemented by additional members, the current CAB provides advice on management and operation of the Biobank, reviews policies, evaluates participant materials, and provides input on complex policy decisions such as data sharing and return of research results. Examples include discussions about sharing Mayo clinic data with the federally mandated dbGaP resource housed at NIH, and detailed discussions about when and how to offer research findings discovered in the course of research in the hemoglobinopathies. The CAB, which meets quarterly, has 20 members and is co-chaired by a community member.

Access to Biobank

Requests from investigators who wish to access Mayo Clinic Biobank samples and data are reviewed by the Biobank Access Committee. All requests are considered, including those internal or external to Mayo Clinic, as well as both academic and commercial requests. Related fees vary by type and source of request and can be obtained by contacting the Mayo Clinic Biobank. The Access Committee is made up of two general internists representing areas from which patients are recruited, two clinical geneticists, a genetic counselor, a bioethicist, the community co-chair of the CAB, a statistician, several basic science researchers, an epidemiologist and the Biobank leadership. All proposals are reviewed by the Access Committee to ensure adherence to our principles, including scientific excellence, alignment with institutional goals, and potential to improve patient care. Because of the early stage of the Biobank and the large quantity of DNA extracted at time of collection, we have not yet had to deny requests for lack of samples. This may change over time, especially if there are requests for large quantities of serum and/or plasma, as our supply of those compared to the minimum needed for many tests is more limited in than is our supply of DNA.

Return of Research Results to Participants

The Biobank Access Committee, in collaboration with the CAB, has developed a policy for determining whether research results, including incidental findings, warrant being offered to Biobank participants. Potentially clinically meaningful results (e.g., genetic tests results with known clinical utility) are reviewed by an ***ad hoc*** panel of experts to determine whether knowledge of the research results would affect clinical care of Biobank participants. If determined by the expert panel to be valuable for clinical care, Biobank participants are offered the opportunity to discuss with a genetic counselor whether to learn their test results. Those who chose to learn their results are offered a second appointment with a genetic counselor at which time the research results are disclosed. If results will be used for clinical decision-making, the genetic counselor facilitates confirmation of the research results in a CLIA-approved lab.

Statistical analyses—Data on age at consent, sex, and residence at enrollment were collected from Mayo Clinic patient registration data, while body mass index (BMI, (height in cm)/(weight in kg)^2) data were obtained from the EMR. Self-reported questionnaires were used for obtaining education and smoking status for Biobank participants. Demographic characteristics of Mayo Clinic Biobank participants were summarized using percentages. The medians and 25th and 75th percentiles were calculated. Continuous variables were

categorized before calculating percentages. Counts and percentages were used to summarize the top 15 self-reported diagnoses, the 15 most commonly prescribed medication groups since enrollment, and the top 15 laboratory tests measured within 5 years prior to enrollment. The characteristics of Mayo Clinic Biobank participants were compared to the participants of the Center for Disease Control's 2011 Behavioral Risk Factor Surveillance System Survey (BRFSS: http://www.cdc.gov/brfss/technical_infodata/surveydata/2011.htm). The 2011 BRFSS data were obtained through telephone surveys from more than 500,000 adults who were selected as representatives within each state. Prevalence estimates from the BRFSS have been shown to be comparable to other national surveys such as the National Health Interview Survey and the National Health and Nutrition Examination Survey.[24–26] Both US total and upper Midwest (Minnesota, Iowa, North Dakota, South Dakota, and Wisconsin) BRFSS survey results were used for comparison.

Results

Over the first three years of enrollment, 21,736 Mayo Clinic patients consented to participate and provided both blood and questionnaire data. Among all invited Mayo Clinic patients, 29% chose to participate, 15% refused, and the remainder did not respond to the two invitations. For the Mayo Clinic Biobank participants, 58% were female, median age was 62 years, 36% were classified as overweight (BMI of 25–29.9) and 33% were obese (BMI > 30) (Table 1). We compared our data to national (US total) data from the 2011 CDC's BRFSS survey. Mayo Clinic Biobank participants were older (46% aged 65 or older in Biobank vs. 18% in BRFSS), were more likely to be female (58% in Mayo Clinic Biobank vs. 51% in BRFSS), were more likely to be of European ancestry (95% in Mayo Clinic Biobank vs. 77% in BRFSS), were more likely to be obese (33% in Biobank vs. 27% in BRFSS), and were more highly educated (81% associate's degree or higher in Biobank vs. 55% in BRFSS). The percentage of participants who never smoked was higher for Mayo Clinic Biobank participants (59%) than was reported nationally (55%). The percentage of current smokers was similar in the Mayo Clinic Biobank compared to national data (13% vs. 14%). Adjusting for age and sex did not materially change these results.

Mayo Clinic Biobank participants mainly reside in the upper Midwest states with more than 74% from Minnesota (MN); 42% of the total were from Olmsted County where the Mayo Clinic is located (Table 1). Compared to the BRFSS population from the Upper Midwest states, the differences between participants in the Mayo Clinic Biobank were somewhat smaller than when comparing to national US data (Table 1). For example, the Upper Midwest BRFSS participants were 90.8% European which is more similar to the 95% in the Mayo Clinic Biobank participants than the national average of 83.1%.

Mayo Clinic Biobank participants were recruited through 11 different practice locations (Table 2). Among all 21,736 participants, 22% had a clinic visit to Primary Care Internal Medicine at the time of enrollment, while 6% of them were volunteered. Patients having an appointment to the General Internal Medicine – Regional Practice had the highest participation rate when invited (33%), followed by patients seen in Executive Health (31%). Participation rate for Primary Care Internal Medicine was 26%.

Table 3 displays the top 15 self-reported diagnoses in the Mayo Clinic Biobank among participants in the first three years of enrollment. The five most commonly self-reported conditions were hyperlipidemia (41%), hypertension (38%), osteoarthritis (30%), any cancer (29%), and gastroesophageal reflux disease (GERD) (26%). Among 29% of participants with any self-reported cancer, the five most common cancer types were non-melanoma skin cancer (14%), prostate cancer (12% in men), breast cancer (4% in both men and women), melanoma (3%), and cervical cancer (2% in women).

Fifty-six percent of Mayo Clinic Biobank participants had at least 15 years of EMR data (Figure 1A). Seventy-seven percent had at least 2 clinic visits per year and an EMR length greater than one year (Figure 1B). Medication data were pulled from the EMR for visits from the date of enrollment through July 2012 (Table 4). The most common types of medications included laxatives (48%), opioid analgesics (41%), antilipemic agents (40%), topical anti-infectives/anti-inflammatory agents (35%), and anti-microbials (32%). The top five laboratory tests in the five years prior to enrollment were a lipid screen (total cholesterol, triglycerides, HDL cholesterol, LDL cholesterol), glucose, creatinine, aspartate aminotransferase (AST), and thyroid stimulating hormone (TSH) (Table 5).

Currently, the Mayo Clinic Biobank stores DNA, plasma, platelet poor plasma, serum, and WBC samples from all participants (Table 6).The total number of processed and stored sample vials on patients recruited in the first three years in the Mayo Clinic Biobank is 458,000, with 56,000 DNA, 116,000 serum, 215,000 plasma, 64,600 WBC, and 5,900 slow frozen cells. Median DNA available per patient is 182 μg (5^{th} percentile to 95^{th} percentile 74 μg to 343 μg). Median serum available per patient is 3.5 mL (5^{th} percentile to 95^{th} percentile 3 mL to 4 mL).

Since the establishment of the Mayo Clinic Biobank, over 60 studies have been approved to use Mayo Clinic Biobank biological specimens and/or self-reported questionnaire data. The most common use of the Mayo Clinic Biobank among investigators was as a control group for their disease-specific cases (N=34), while a few studies used the Biobank to increase the number of disease-specific cases by abstracting the medical records (N=13). Ten requests were to use existing data within the Biobank (i.e., whole-exome sequence data on 89 subjects); four were for normal samples for laboratory test development; three were to develop exposure-specific cohorts as well as a few other miscellaneous types of studies. Over 69,000 aliquots, mostly DNA samples for genotyping, have been approved for distribution. Data generated from samples and/or medical record abstractions are returned to the Biobank.

Discussion

Since being established in 2009 by the Mayo Clinic Center for Individualized Medicine, the Mayo Clinic Biobank grew to over 21,000 participants within three years. The goal of the Mayo Clinic Biobank is to support a wide array of health-related research studies, especially those with the potential to improve patient care. A key strength of the Mayo Clinic Biobank is that it is embedded in a thriving clinical practice and participants have agreed to allow researchers access to their past and future Mayo Clinic medical records. Many of the

participants have over 15 years of data in the EMR and many have many additional years of data available for manual abstraction from the paper medical records, if needed. EMR data allow us to use natural language processing (NLP) tools to extract other non-indexed data from the EMR text. By requiring all projects using the Mayo Clinic Biobank to return the data obtained from our participants we continually build the value of the institutionally-managed Biobank resource. Because of these strengths, over 60 projects have requested samples and/or data from the Mayo Clinic Biobank since the start of enrollment. Over 69,000 samples have been approved for distribution. Data incorporated into the Mayo Clinic Biobank include those from written questionnaires, EMR and administrative data. Blood samples and their derivatives have been made available for research. Finally, we have a robust, active, CAB with strong leadership from the community that has provided meaningful guidance on a wide array of topics.

Although there are myriad strengths associated with the Mayo Clinic Biobank, it has limitations. First, the Mayo Clinic Biobank, with a goal of 50,000, will be much smaller than many recently established biobanks such as the UK Biobank[27] (N=500,000+); the China Kadoorie Biobank[28] (N=500,000+); the BioVu Biobank[4] at Vanderbilt University School of Medicine in Nashville, Tennessee (N=150,000 plus) and the Kaiser Permanente Biobank[5] (N=100,000). Although it is small compared to some existing cohort studies, it can be used to assess common and moderately common diseases. It can also serve as a source of controls to case-series.

Second, the Mayo Clinic Biobank is not population-based and characteristics of its participant pool diverge from the population parameters offered by the BRFSS. For example, 81% of Biobank participants reported having an Associate's degree or higher compared to the national rate of 55%. Based on 2010 US Census Data (http://repweb.mayo.edu/population-information/rep-population-information), residents of Olmsted County, where the Mayo Clinic is located, had higher level of education (39.1% with a Bachelor's degree or higher compared to the national average of 27.9%). However, we found that residence did not fully explain the high rates of education in Biobank participants: 51.5% of Biobank participants from Olmsted County reported having a Bachelor's degree or higher compared to 39.1% in Olmsted County residents in general. We concluded that our greater than average rate of higher education may be influenced by multiple factors, including: a higher probability of recruitment of Mayo Clinic employees and dependents (who are more likely to be well educated than the rest of the local residents) considering our recruitment from primary care practices where most employees and dependents obtain their health care, and a perhaps higher than average probability of participation among the well-educated. In addition, comparisons between the Mayo Clinic Biobank participant pool and the BRFSS should be made with caution as the question-asking, sample frame, time frame, and method of data collection used by the BRFSS differs significantly from those associated with the Mayo Clinic Biobank. Such differences might introduce error that may artificially over- or under-estimate concordance and therefore bias.

Third, our Biobank may not completely represent the general population. However, the majority of the approved studies also recruit their disease cases from clinic patients, not

from the general population. Therefore, in studies utilizing the Biobank, we try to minimize selection bias by selecting matched controls to be comparable to the case distribution.

A fourth limitation is that the observed participation rate of 29% can be perceived as low. In a separate investigation, we found that sex, age, region of residence, and race/ethnicity were significantly associated with participation and that refusers more often cited privacy concerns while non-responders more often identified time constraints as the reason for non-participation[29]. It is interesting to note that an even lower response rate (5.5%) was reported in the UK Biobank, a similar opt-in biobank [30]. How these findings affect the inferential value of the Mayo Clinic Biobank is unclear.

Finally, for reasons of practicality and budget, the study materials were designed as English-only. The impact of this is likely to be only minimal as only 11% of the Minnesota population speaks a language other than English at home and more than 60% of those falling into this category can speak and read English very well (http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml). Similar data pertain to the Jacksonville, Florida region where Mayo Clinic Florida is located; only 5% of Jacksonville residents were born in a foreign country. Thus, we do not expect any higher impact of our English-only materials in the Mayo Clinic Florida and do not plan to translate our study materials into languages other than English. Nonetheless, the lack of translated materials may have contributed to the under-representation of minority participants in the Mayo Clinic Biobank.

Conclusion

In conclusion, the Mayo Clinic Biobank is an established valuable resource for researchers at Mayo Clinic. Biobanks like the one described herein - with biospecimens and risk factor data on a large number of participants embedded in a clinical practice and followed longitudinally - will be critical to the successful translation of individualized medicine.

Acknowledgments

Financial support and disclosure: The Mayo Clinic Biobank is supported by the Mayo Clinic Center for Individualized Medicine. This publication was also supported by Grant Number UL1 TR000135 from the National Center for Advancing Translational Sciences (NCATS) and Grant Number R01CA14517 from the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI).

We gratefully acknowledge the support acknowledge the efforts of our staff: Study coordinators, including: Maro Andayo, Tom Bitz, Lisa Hines, Kelly Lyke, Mary Lynch, Brenda Maringer, JoAnn Peterson, Olga Richter, Pam Vega; Programmers, including: Edward Asmann, Brandon Dallman, Aaron Kurtzhals, Shaun Maloney, Jonathan Schmidt; Administrative Assistants: Cathy Devine, Malinda Woodward. We also acknowledge the contributions of the Community Advisory Board.

Abbreviations

BRFSS	behavioral risk factor surveillance system survey
CAB	community advisory board
DCE	deliberative community engagement project
HR	hazard ratio

CI	confidence interval
EMR	electronic medical records

References

1. Giannini C, Oelkers MM, Edwards WD, et al. Maintaining Clinical Tissue Archives and Supporting Human Research: Challenges and Solutions. Archives of Pathology & Laboratory Medicine. 2011 Mar 01; 135(3):347–353. [PubMed: 21366459]
2. Gaskell G, Gottweis H. Biobanks need publicity. Nature. 2011; 471(7337):159–160. [PubMed: 21390108]
3. McCarty CA, Wilke RA, Giampietro PF, Wesbrook SD, Caldwell MD. Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. Personalized Medicine. 2005; 2:49–79.
4. Roden DM, Pulley JM, Basford MA, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. Sep; 2008 84(3):362–369. [PubMed: 18500243]
5. Scott CT, Caulfield T, Borgelt E, Illes J. Personal medicine--the new banking crisis. Nat Biotechnol. Feb; 2012 30(2):141–147. [PubMed: 22318029]
6. Gulcher J, Stefansson K. Population genomics: laying the groundwork for genetic disease modeling and targeting. Clin Chem Lab Med. Aug; 1998 36(8):523–527. [PubMed: 9806453]
7. MacLeod AK, Liewald DCM, McGilchrist MM, Morris AD, Kerr SM, Porteous DJ. Some principles and practices of genetic biobanking studies. European Respiratory Journal. Feb 1; 2009 33(2):419–425. [PubMed: 19181915]
8. Ollier W, Sprosen T, Peakman T. UK Biobank: from concept to reality. Pharmacogenomics. Sep 01; 2005 6(6):639–646. [PubMed: 16143003]
9. Zielhuis GA. Biobanking for epidemiology. Public Health. 2012; 126(3):214–216. [PubMed: 22325670]
10. O'Doherty KC, Hawkins AK, Burgess MM. Involving citizens in the ethics of biobank research: informing institutional policy through structured public deliberation. Soc Sci Med. Nov; 2012 75(9):1604–1611. [PubMed: 22867865]
11. O'Doherty K, Gauvin FP, Grogan C, Friedman W. Implementing a public deliberative forum. Hastings Cent Rep. Mar-Apr;2012 42(2):20–23. [PubMed: 22733326]
12. Avard D, Bucci LM, Burgess MM, et al. Public Health Genomics and Public Participation: Points to Consider. Journal of Public Deliberation. 2009; 5(1):Article 7.
13. Secko DM, Preto N, Niemeyer S, Burgess MM. Informed consent in biobank research: a deliberative approach to the debate. Soc Sci Med. Feb; 2009 68(4):781–789. [PubMed: 19095337]
14. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. Jun; 1992 30(6):473–483. [PubMed: 1593914]
15. Coates A, Fischer Dillenbeck C, McNeil DR, et al. On the receiving end—II. Linear analogue self-assessment (LASA) in evaluation of aspects of the quality of life of cancer patients receiving therapy. European Journal of Cancer and Clinical Oncology. 1983; 19(11):1633–1637. [PubMed: 6315445]
16. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med. 1991; 32(6):705–714. [PubMed: 2035047]
17. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care. Nov; 2003 41(11):1284–1292. [PubMed: 14583691]
18. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci. Sep; 1985 10(3):141–146. [PubMed: 4053261]
19. Chute CG, Beck SA, Fisk TB, Mohr DN. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. Journal of the American Medical Informatics Association. Mar 1; 2010 17(2):131–135. [PubMed: 20190054]

20. Pathak J, Murphy SP, Willaert BN, et al. Using RxNorm and NDF-RT to classify medication data extracted from electronic health records: experiences from the Rochester Epidemiology Project. AMIA Annu Symp Proc. 2011; 2011:1089–1098. [PubMed: 22195170]
21. Zhong W, Maradit-Kremers H, StSauver JL, et al. Age and Sex Patterns of Drug Prescribing in a Defined American Population. Mayo Clin Proc. 2013
22. Clunie DA. DICOM structured reporting and cancer clinical trials results. Cancer Inform. 2007; 4:33–56. [PubMed: 19390663]
23. Ruch P, Gobeill J, Lovis C, Geissbuhler A. Automatic medical encoding with SNOMED categories. BMC Med Inform Decis Mak. 2008; 8 (Suppl 1):S6. [PubMed: 19007443]
24. Hurt, Rd; WSAEJO, et al. MYocardial infarction and sudden cardiac death in olmsted county, minnesota, before and after smoke-free workplace laws. Archives of Internal Medicine. 2012; 172(21):1635–1641. [PubMed: 23108571]
25. Nelson DE, Powell-Griner E, Town M, Kovar MG. A comparison of national estimates from the National Health Interview Survey and the Behavioral Risk Factor Surveillance System. Am J Public Health. Aug; 2003 93(8):1335–1341. [PubMed: 12893624]
26. Li C, Balluz LS, Ford ES, Okoro CA, Zhao G, Pierannunzi C. A comparison of prevalence estimates for selected health indicators and chronic diseases or conditions from the Behavioral Risk Factor Surveillance System, the National Health Interview Survey, and the National Health and Nutrition Examination Survey, 2007–2008. Prev Med. Jun; 2012 54(6):381–387. [PubMed: 22521996]
27. Collins R. What makes UK Biobank special? Lancet. Mar 31; 2012 379(9822):1173–1174. [PubMed: 22463865]
28. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol. Dec; 2011 40(6):1652–1666. [PubMed: 22158673]
29. Ridgeway JL, Han LC, Olson JE, et al. Potential Bias in the Bank: What Distinguishes Refusers, Non-responders and Participants in a Clinic-based Biobank? Public Health Genomics. 2013
30. Swanson JM. The UK Biobank and selection bias. Lancet. Jul 14.2012 380(9837):110. [PubMed: 22794246]

Figure 1A

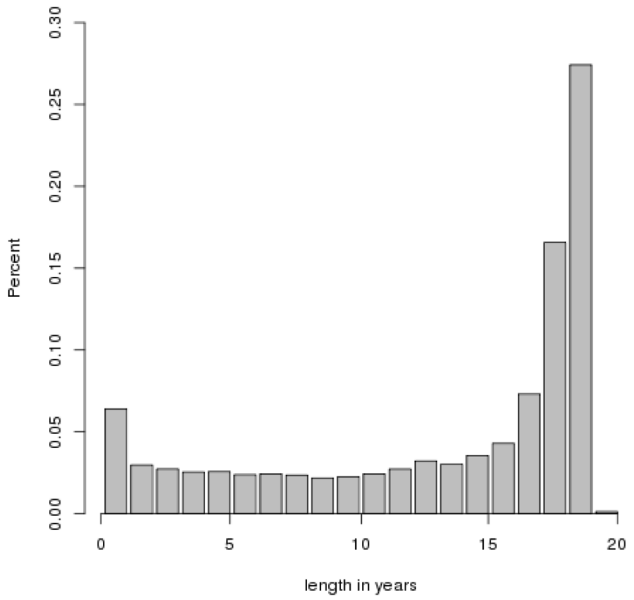

Figure 1B

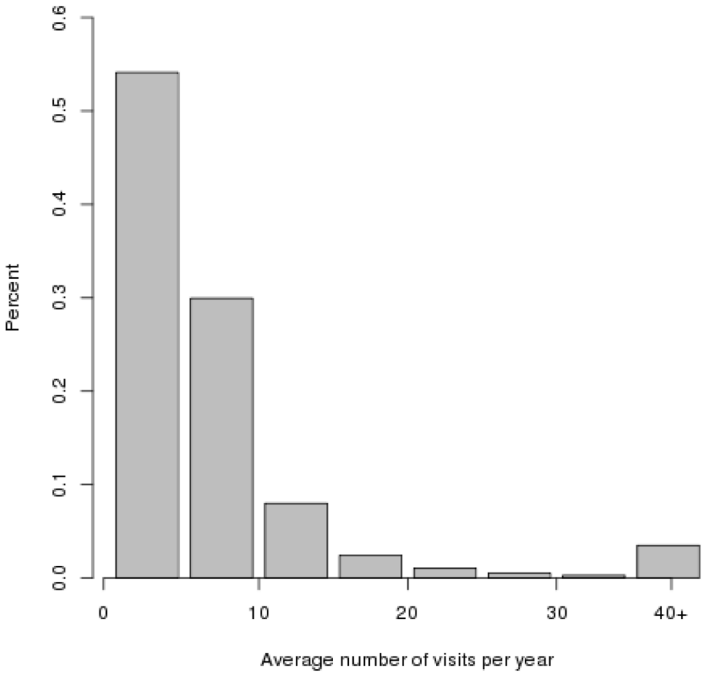

Figure 1.

Depth of the electronic medical records (EMR) for Mayo Clinic Biobank participants, using the length (panel A) of the EMR and the average number of visits per year (panel B). The maximum length of EMR available in Mayo Clinic is 19 years.

Table 1

Characteristics of Mayo Clinic Biobank participants enrolled between April 2009 and March 2012 (n=21,736), compared to the Center for Disease Control's 2011 BRFSS* participants. Data were collected at consent for Mayo Clinic Biobank and at survey for BRFSS.

	Mayo Clinic Biobank (N=21,736)	BRFSS Upper Midwest[a] (N=41,622)	BRFSS US Total (N=504,408)
Age in years, %			
18–44	16.0%	46.5%	48.3%
45–54	16.7%	19.5%	18.9%
55–64	21.9%	15.6%	15.2%
65 or older	45.5%	18.4%	17.7%
Female, %	57.5%	50.7%	51.3%
Residence location			
MN - Olmsted County only	42.1%	0.7%	0.04%
MN – other South East MN counties	19.2%	2.4%	0.1%
Remainder of MN	13.0%	30.9%	1.6%
Iowa	5.8%	19.6%	1.0%
Wisconsin	4.2%	36.9%	1.8%
North Dakota	0.7%	4.4%	0.2%
South Dakota	0.8%	5.2%	0.3%
Other US states	14.2%		95.0%
Race			
European Ancestry	95.0%	89.4%	76.7%
Black/African American	0.5%	3.8%	12.2%
Asian	1.2%	1.9%	4.1%
Native American/Alaskan Native	0.1%	1.7%	1.7%
Other	0.3%	2.1%	3.5%
Mixed races	3.0%	1.5%	1.8%
BMI, %[b]			
Less than 18.5	3.2%	1.6%	1.9%
18.5 – 24.9	27.8%	34.7%	34.9%
25.0 – 29.9	35.9%	36.3%	35.8%
30 or higher	33.1%	27.3%	27.4%
Education, %			
Less than high school	2.1%	10.3%	15.4%
High school graduate	16.6%	31.3%	29.3%
Associate degree or higher	81.2%	58.4%	55.3%
Smoking, %			
Never	58.7%	53.6%	55.0%
At least 100 cigarettes	41.3%	46.2%	44.8%
Current smoker [b]	12.5%	14.5%	14.2%

* BRFSS: Behavioral Risk Factor Surveillance System

[a] Upper Midwest includes Minnesota (MN), Iowa (IA), North Dakota (ND), South Dakota (SD), and Wisconsin (WI).

[b] Current smoker was defined as a subject who has been smoking at least 100 cigarettes so far and currently smokes.

Table 2

Summary of recruitment locations for all 21,736 enrolled patients and participation rate within each location.

Location	Number of participants	Percent†	Participation rate within location
Primary Care Internal Medicine	4858	22.4%	26.4%
General Internal Medicine - Regional Practice	3524	16.2%	32.6%
General Internal Medicine - National Practice	3075	14.2%	29.8%
Preventive, Occupational and Aerospace Medicine	2400	11.0%	25.2%
Family Medicine	2329	10.7%	24.5%
Family Medicine (NE Clinic)	1650	7.6%	23.0%
Volunteer/no appointment	1298	6.0%	--
Family Medicine (NW Clinic)	1003	4.6%	22.0%
Family Medicine (Kasson)	949	4.4%	18.0%
Executive Health	528	2.4%	31.4%
Gynecology*	101	0.5%	19.4%
Obstetrics*	21	0.1%	5.1%

†Percent of patients recruited from each location among all 21,736 MC Biobank participants.

*Recruitment had been ongoing for less than two months in these departments at the time these data were collected.

Table 3

Frequencies of the top 15 self-reported diagnoses at time of enrollment among Mayo Clinic Biobank participants enrolled between April 2009 and March 2012 (n=21,736).

	Counts	Percent
Hyperlipidemia	8979	41.3%
Hypertension	8174	37.6%
Osteoarthritis	6448	29.7%
Cancer [a]	6224	28.6%
Non-melanoma skin cancer [b]	2950	13.6%
Prostate (% in men) [b]	1107	12.0%
Breast cancer (% in both gender) [b]	941	4.3%
Melanoma [b]	692	3.2%
Cervical (% in women) [b]	240	1.9%
GERD [c]	5669	26.1%
Cataracts	5277	24.3%
Depression	4700	21.6%
Abnormal distance vision	4547	20.9%
Migraine headaches	3824	17.7%
Anxiety	3609	16.6%
Sleep apnea	3195	14.7%
Non-melanoma skin cancer	2950	13.6%
Asthma	2673	12.3%
Hyperthyrodism/hypothyroidism	2647	12.6%
Type II diabetes	2025	9.5%

[a] Counts for cancer are the number of Mayo Clinic Biobank participants with at least one of cancer types.

[b] Top 5 most common cancer types.

[c] GERD: Gastroesophageal reflux disease

Table 4

Frequencies of the top 15 most common medication groups for Mayo Clinic Biobank participants after their enrollment between April 2009 and March 2012 (n=21,736)[*], [†].

	Counts	Percent
Laxatives	10529	48.4%
Opioid analgesics	8987	41.4%
Antilipemic agents	8664	39.9%
Topical anti-infective/anti-inflammatory agents	7528	34.6%
Penicillins and beta-lactam antimicrobials	6929	31.9%
Gastrointestinal medications	6213	28.6%
Antidepressants	6092	28.0%
Genitourinary agents	5818	26.8%
Vaccines/toxoids	5612	25.8%
Topical nasal and throat agents	5526	25.4%
Diuretics	5201	23.9%
Beta-blockers and related medications	5176	23.8%
Sedatives/hypnotics	4848	22.3%
ACE inhibitors	4807	22.1%
Quinolones	4686	21.6%

[*] Medication data were extracted from electronic medical records using methods outlined in Pathak et al.[20]

[†] Participants who received multiple medications within the same medication group were counted once.

Table 5

Frequencies of the top 15 laboratory tests measured within 5 years prior to enrollment for Mayo Clinic Biobank participants (n=21,736).[†]

	Counts	Percent
Lipid Screen [a]	18751	86.3%
Glucose (Plasma)	18604	85.6%
Creatinine	18177	83.6%
Aspartate Aminotransferase (AST or SGOT)	16990	78.2%
TSH [b], Sensitive, (Serum)	16341	75.2%
Sodium, (Plasma/Serum)	15586	71.7%
Estimated Glomerular Filtration Rate (eGFR)	13083	60.2%
Calcium	12677	58.3%
Alkaline Phosphatase	12620	58.1%
Alanine Aminotransferase (ALT GPT) (Serum)	11071	50.9%
Bilirubin,Total	6684	30.8%
Hemoglobin A1C, B	6496	9.9%

[†]Participants who had a given laboratory test multiple times were counted only once.

[a]Lipid Screen included cholesterol (Serum), Triglycerides (Serum), High-Density Lipoprotein Cholesterol (Serum), and Low-Density Lipoprotein Cholesterol (calculated).

[b]TSH: thyroid stimulating hormone

Table 6

Current quantities of samples processed and stored in the Mayo Clinic Biobank, separately by sample type on all 21,736 Biobank participants.

Sample Type	Number of unique participants (%)	Median	5th Percentile to 95th Percentile
DNA	21715 (99.9)	182889 ng	(74016 ng, 343009 ng)
Serum	21713 (99.9)	3500 uL	(3000 uL, 4000 uL)
Plasma - EDTA - Platelet Poor	21649 (99.6)	6500 uL	(4500 uL, 9500 uL)
Plasma - Na Cit*	21695 (99.8)	2000 uL	(2000 uL, 2000 uL)
White Blood cells (buffy coat)	21727 (99.9)	20000 uL	(2000 uL, 25000 uL)
Slow Frozen white blood cells	1961 (9.02)	5400 uL	(5400 uL, 5400 uL)

*Na Cit: Sodium Citrate