Journal of Clinical Epidemiology 70 (2016) 214–223

Journal of Clinical Epidemiology

Million Veteran Program: A mega-biobank to study genetic influences on health and disease

John Michael Gaziano[a,b,1], John Concato[c,d,*,1], Mary Brophy[a,e], Louis Fiore[a,e], Saiju Pyarajan[a], James Breeling[a], Stacey Whitbourne[a], Jennifer Deen[a], Colleen Shannon[a], Donald Humphries[a], Peter Guarino[c,d], Mihaela Aslan[c,d], Daniel Anderson[c], Rene LaFleur[c], Timothy Hammond[f], Kendra Schaa[f,2], Jennifer Moser[f], Grant Huang[f], Sumitra Muralidhar[f], Ronald Przygodzki[f], Timothy J. O'Leary[f]

[a]*Massachusetts Area Veterans Epidemiology Research and Information Center (MAVERIC), VA Cooperative Studies Program, VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130, USA*

[b]*Department of Internal Medicine, Harvard Medical School, Boston, MA, USA*

[c]*Clinical Epidemiology Research Center (CERC), VA Cooperative Studies Program, VA Connecticut Healthcare System, 950 Campbell Avenue, 151B, West Haven, CT 06516, USA*

[d]*Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA*

[e]*Department of Internal Medicine, Boston University School of Medicine, Boston, MA, USA*

[f]*Office of Research and Development, Veterans Health Administration, 810 Vermont Avenue N.W., 10P9, Washington, DC 20420, USA*

Accepted 22 September 2015; Published online 9 October 2015

Abstract

Objective: To describe the design and ongoing conduct of the Million Veteran Program (MVP), as an observational cohort study and mega-biobank in the Department of Veterans Affairs (VA) health care system.

Study Design and Setting: Data are being collected from participants using questionnaires, the VA electronic health record, and a blood sample for genomic and other testing. Several ongoing projects are linked to MVP, both as peer-reviewed research studies and as activities to help develop an infrastructure for future, broad-based research uses.

Results: Formal planning for MVP commenced in 2009; the protocol was approved in 2010, and enrollment began in 2011. As of August 3, 2015, and with a steady state of ≈50 recruiting sites nationwide, $N = 397{,}104$ veterans have been enrolled. Among $N = 199{,}348$ with currently available genotyping data, most participants (as expected) are male (92.0%) between the ages of 50 and 69 years (55.0%). On the basis of self-reported race, white (77.2%) and African American (13.5%) populations are well represented.

Conclusions: By helping to promote the future integration of genetic testing in health care delivery, including clinical decision making, the MVP is designed to contribute to the development of precision medicine. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Genomics; Cohort studies; Genotyping; Exome sequencing; Whole-genome sequencing; Veterans

1. Introduction

Progress in biomedical research, informatics, and medical care is merging in the 21st century, providing an enhanced ability to understand how genes affect health and disease. Activities such as the Human Genome Project [1], Hap Map project [2,3], single nucleotide polymorphism (SNP) Consortium [4], and 1,000 Genomes [5] establish a framework for understanding the human genome across populations. Technological advances in genotyping and sequencing have improved the quality, and lowered the cost, of analyzing genetic data. In addition, the emergence of ''personalized'' or ''precision'' medicine has increased interest in integrating genetic testing within health care delivery [6], and a US Precision Medicine Initiative [7] was recently announced.

Conflict of interest: The authors declare no conflict of interest related to this work, supported by the US Government.

Funding: The research was funded by the Office of Research and Development, Department of Veterans Affairs.

[1] Contributed equally to this work.

[2] Current address: Medstar Washington Hospital Center, 110 Irving St N.W., Washington, DC 20010, USA.

* Corresponding author. Tel.: 203-932-5711x2993; fax: 203-937-4932.
E-mail address: john.concato@va.gov (J. Concato).

http://dx.doi.org/10.1016/j.jclinepi.2015.09.016

What is new?

Key finding

- An observational cohort study design can be combined with an electronic health record system and genomic laboratory testing to create a mega-biobank.

What this adds to what was known?

- An epidemiologic research infrastructure and database, currently including almost 400,000 enrollees, is embedded within the national health care system operated by the US Department of Veterans Affairs.
- Two alpha-test projects are ongoing; four solicited beta-test projects have been approved.

What is the implication and what should change now?

- Researchers and clinicians should be prepared to understand, evaluate, and interpret genomic studies that will inform the practice of "precision medicine."

1.1. Epidemiologic context

Initial research efforts in this field have included studies that use a prospective observational cohort study design. Examples include the population-based Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium [8], as well as the National Cancer Institute Cohort Consortium [9]. Such consortia can analyze collected data among subsets of patients with and without a given disease, and then study gene-health associations, often using case-control analytic frameworks. Limitations exist, however, related to available health outcomes, the scope of medical data collected, and even the numbers of "case" patients with a specific disease or trait.

The emergence of large repositories of data on health and disease within electronic health record systems provides another opportunity for studying the genomic-health association. This approach can include cross-sectional or longitudinal health data, yet at lower costs per participant compared to primary data collection. A cursory review of large genomic initiatives around the world—such as the UK Biobank [10–12], similar efforts at Vanderbilt University [13], the Kaiser Permanente Research Program on Genes, Environment, and Health [14], the China Kadoorie Biobank [15], as well as others—reveal several key features for establishing these large-scale projects. For example, to increase sample size and manage costs, BioVU at Vanderbilt University links medical record data with residual clinical specimens deposited in a repository, using an "opt-out" strategy for patients to actively remove themselves from the study. This strategy, although efficient, has limitations regarding longitudinal data and information on lifestyle factors. The UK Biobank and Kaiser Permanente projects combine questionnaires (as in traditional cohorts) with electronic medical record data, using an "opt-in" strategy requiring active consent. The UK Biobank included in-person study visits and blood specimens; the Kaiser Permanente effort recruits, consents, and collects data via the Web, with saliva specimens sent in by mail.

1.2. Health system context

The UK Biobank and Kaiser Permanente initiatives exist within larger health care systems that—as with the Veterans Health Administration (VHA)—involve the integration of health care delivery. Although large-scale genomic research holds promise for any health care system, only a fraction of systems possess key elements for successful execution, such as (1) available study population; (2) infrastructure for carrying out scientific, regulatory, technological, and administrative needs of research; (3) capacity for longitudinal collection of data (with history, physical examination, laboratory, diagnostic information); (4) follow-up for clinical outcomes; (5) availability of a genetic and serum biobank that can centrally store and retrieve specimens for use in future research projects; and (6) retention of participants in the program over time.

The Department of Veterans Affairs (VA) is an appropriate setting to construct a mega-biobank, as well as to conduct genomic research. The VHA is the largest integrated national health system in the country, and most of the health care experience of the veterans who use the system has been captured electronically for many years. Other strengths of the VA include an altruistic Veteran population, over 100 research-ready medical centers, a state-of-the-art biorepository, a bioinformatics infrastructure to enable secure access to genetic and medical data, and an intramural clinical research network is embedded in the health care environment, serving to support the overall health care mission of VHA.

As of the early 2000s, genomic research activities in the VA had already included the creation of a DNA Bank, managed by the VA Cooperative Studies Program (CSP), to support genetic studies among participants enrolled in multisite clinical trials [16]. Subsequent discussions focused on the creation of a primary resource for genomic analyses. Preliminary planning included two surveys that solicited stakeholder input on establishing such a resource [17,18]. The results indicated that Veterans tended to be supportive of a VA-based initiative involving research on genomics, preferred a direct consent ("opt-in") approach, and wanted assurances that their data would be kept secure.