EDITOR'S CHOICE

The population health record: concepts, definition, design, and implementation

Daniel J Friedman,[1] R Gibson Parrish II[2]

[1]Population and Public Health Information Services, Brookline, Massachusetts, USA
[2]Peacham, Vermont, USA

Correspondence to
Daniel J Friedman, Population and Public Health Information Services, 12 Gorham Avenue, Brookline, MA, USA 02445; friedman.daniel.jay@gmail.com

Received 19 October 2009
Accepted 20 April 2010

ABSTRACT

In 1997, the American Medical Informatics Association proposed a US information strategy that included a population health record (PopHR). Despite subsequent progress on the conceptualization, development, and implementation of electronic health records and personal health records, minimal progress has occurred on the PopHR. Adapting International Organization for Standarization electronic health records standards, we define the PopHR as a repository of statistics, measures, and indicators regarding the state of and influences on the health of a defined population, in computer processable form, stored and transmitted securely, and accessible by multiple authorized users. The PopHR is based upon an explicit population health framework and a standardized logical information model. PopHR purpose and uses, content and content sources, functionalities, business objectives, information architecture, and system architecture are described. Barriers to implementation and enabling factors and a three-stage implementation strategy are delineated.

PURPOSE

The purpose of this article is threefold: first, to provide a conceptual description of the population health record (PopHR), including preliminary definitions, design, purpose and uses, intended users, content, sources of content, functionalities, business objectives, and system and information architectures; second, to discuss key issues in developing a population health record for the US, including barriers to and enabling factors for development and implementation; and third, to recommend a PopHR implementation strategy.[i]

In this article, we argue that the PopHR constitutes a valid concept for the US and that the development and implementation of the PopHR could potentially correct basic problems in the comprehensiveness, accessibility, communication, timeliness, and use of population health information. By assimilating information and statistics from diverse data sets and sources, the PopHR would provide a more comprehensive view of population health than is currently available and support exploratory and other analyses of health and factors that influence it.

Recent developments in health informatics could facilitate the design, development, and implementation of the PopHR. This article draws upon relevant US lessons learned from healthcare (electronic health records (EHRs) and health information exchange); consumer health (personal health records); and public health (web technologies for reporting population health data, conducting surveillance, and communicating information to professional and public audiences). In addition, new opportunities for the development of PopHR in the US are presented by recent web developments in medical informatics and more generally, such as grid and cloud computing, web services, wikis, user-generated content, data mashups, and social networking.

The PopHR could prove useful for enhancing comparative effectiveness research, providing population context and benchmarks for clinical trials, and supporting the development of learning healthcare systems as clinicians learn more about the populations served than is possible through EHRs alone.[1] The PopHR could also serve as a step toward the goal of 'healing the schism' between medicine and public health.[2] However, this article will focus primarily upon the PopHR for population and public health purposes.

BACKGROUND

Genesis of the population health record

In 1997, the Board of Directors of the American Medical Informatics Association (AMIA) proposed a national health information strategy that included computer-based health records as one of eight major objectives for improving quality, increasing efficiency, and expanding access in the US healthcare system.[3] The AMIA Board postulated the PopHR as one type of computer-based health record, along with those for healthcare institutions, records and systems for primary care and ambulatory care, and personal health records (PHRs).[ii] In 1998, the National Committee on Vital and Health Statistics (NCVHS) also described population records as examples of computer-based health records, in addition to patient and personal records.[4] According to NCVHS, the PopHR would include non-identifiable data on healthcare, behaviors, patient monitoring, risk assessment, and healthcare spending. In 2000, the NCVHS concept of the PopHR evolved into the concept of the "community health dimension", and in 2001 it further evolved into the "population health dimension".[5–7] In 2005, the International Organization for Standardization (ISO) described the PopHR as one of several terms "commonly used to

---

[i] The PopHR to be described should not be construed as necessarily being a single PopHR for the United States. There could be multiple PopHRs for different geographical areas within the US.

[ii] The Institute of Medicine also described a "secondary patient record", that (a) "contains selected data elements to aid nonclinical users" and (b) is "often combined to form. . . a secondary database (eg, an insurance claims database).[52]

describe different types of health records in an electronic form". According to the ISO,

> a population health record contains aggregated and usually de-identified data. It may be obtained directly from EHRs or created de novo from other electronic repositories. It is used for public health and other epidemiological purposes, research, health statistics, policy development, and health services management.[8, iii]

Recent developments

In the more than 10 years since AMIA's 1997 and NCVHS' 1998 inclusion of the PopHR as part of a national health information strategy, the US and other countries have witnessed substantial development of the EHR, including conceptualization; delineation of definitions, standards, and content; implementation and penetration in integrated delivery systems, hospitals, and physician offices; and facilitation through public policies, the private market, and health information exchange. Similarly, since 1997, various iterations of electronic personal health records (PHR)—free-standing, tethered, and integrated—have emerged in the US and other countries.[9] In contrast to US and international advances in EHRs, PHRs, and health information exchange, minimal progress has occurred since 1997 in the US on the PopHR. Neither the PopHR nor its successor concept of the population health dimension has been further conceptualized or developed in the US since its mentions by AMIA in 1997, by NCVHS in 1998, 2000, and 2001, and ISO in 2005.

Need for a population health record

The lack of attention to the PopHR in the US does not result from any large scale improvement in the accessibility of population health data or any diminution of basic need for the PopHR since its initial conceptualization in 1997. The shortcomings in US population health data and related systems detailed by NCVHS in 2002 continue, including "insufficient attention to developing consensus approaches and standards that would simplify collecting, protecting, and accessing data"; "existing data [that] are unnecessarily difficult to locate, access, and use"; "lack of timeliness in making data available"; and "lack of geographic detail."[10]

US population health data are scattered widely at various agencies and web sites, in various forms, at various geographical levels, and with various statistical and reporting conventions, and require various levels of user statistical and computing expertise. There is no single, easily accessible source that provides comprehensive information on population health across regions, states, metropolitan areas, and counties, much less cities, towns, and neighborhoods. As a result, US federal, state, and local government agencies lack a single common and easily accessible source for basic population health data, as do legislators, community-based organizations, the media, and the public. Population-based data on the social determinants of health, needed for improving policy-making, program design,[11] clinical care, and health professional education,[12] remain dispersed and difficult to locate on multiple government and private organization websites. Assimilating statistics from diverse data sets and sources for a given geographic area, and comparing one set of statistics (such as county income quartile and median education) to others (such as smoking prevalence and chronic obstructive pulmonary disease hospitalizations and mortality) can be needlessly difficult and time-consuming.

CONCEPTUALIZING THE US POPULATION HEALTH RECORD

Defining the PopHR

Building upon the ISO definition for electronic health records,[8] we define the PopHR and the PopHR system.

1. The *PopHR* is defined as a repository of statistics, measures, and indicators[iv] regarding the state of and influences on the health of a defined population, in computer processable form, stored and transmitted securely, and accessible by multiple authorized users. The PopHR has a standardized and agreed upon logical information model, which is independent of PopHR systems. The primary purpose of the PopHR is the support of continuing, efficient, and effective public health practice; it contains information that is retrospective (historical), current, and prospective (predictive of future trends). To achieve this purpose, the PopHR should be based upon an explicit population health framework and include a schematic representation of all factors that potentially influence the health of a population, as well as those measures that define population health.[11 13–15]

   A *population* is "all the inhabitants of a given country or area considered together".[16 17, v] The area can be a nation, a region within a nation, a state or province, or a local area within a state or province, such as a county, a city, a town, or a neighborhood. A population may include demographically—or otherwise—defined subpopulations within its geographic bounds, such as those inhabitants belonging to specific age, sex, income, educational attainment, or race groups. *Population health* is the level and distribution of disease, functional status, and wellbeing of a population.[18] *Influences on population health* include the place and time under consideration, and the population's context and attributes.[18] See *PopHR content* below for a fuller description of influences on population health.

   Recent advances in public health informatics center on rapid collection of data on individual patients, or of tabulated data on patients, from EHRs, hospitals, and clinical laboratories,[19–23] and exchange of these data among health departments and between health departments and the Centers for Disease Control and Prevention (CDC).[24] These systems, focusing on data collection and exchange from small numbers of clinical data sets, do not meet the definition of a PopHR, as they do not use a population health framework and are not usually population-based. In addition, the CDC, the National Cancer Institute, and multiple state health departments currently maintain population health information systems, some of which provide the ability to perform customized web-based queries of one or more population-based data sets. This latter function is known as a 'web-based data query system' (WDQS). Current WDQS systems, lacking explicit population health frameworks and associated

---

[iii] See also Hammond 2004;[53] Häyrinen, Saranto, Nykänen 2008;[54] Humphreys 2000.[55] All referred to but did not further develop the PopHR beyond the brief descriptions by the NCVHS and the ISO. Without using the term PopHR, Connecting for Health proposed a related concept.[56]

[iv] Health *statistics* are "numerical data that characterize the health of a population and the factors that influence its health."[10] Health statistics include—but are not limited to—frequency counts, numerical tabulations, rates, ratios, proportions, and measures of dispersion. A population health *measure* is a unit, or a system or scale of such units, for expressing the amount or degree of health of persons in a defined population. A health *indicator* is "a measure that reflects, or indicates, the state of health of persons in a defined population, eg, the infant mortality rate".[57]

[v] Alternative definitions of *population* include the "whole collection of units from which a sample may be drawn"[16] and a "group of individuals, in contrast to the individuals themselves, organized into many different units of analysis, depending on the research or policy purpose".[58] These alternative definitions, which do not require a population to be geographically bounded, would allow the patients in a given healthcare practice, or admitted to a specific hospital within the past year, to serve as a population.

logical population health information models, also do not meet the definition of PopHRs.[25–27]

2. A *PopHR system* is defined as a mechanism for recording, retrieving, and manipulating information in population health records.[8] The PopHR itself is essentially an information container, while the PopHR system includes the people, rules, data standards, processing and storage devices, software applications, and computer networks that facilitate access to, and use of, the PopHR. A PopHR system would provide user access, communication among nodes, and analytic, visualization, and display functionalities.

PopHR purpose and uses

The primary purpose of a PopHR is to document the state of and influences on the health of a defined population, in support of public health practice. Uses of a PopHR and an accompanying PopHR system(s) include monitoring population health status and outcomes; conducting community health assessments and health impact assessments; identifying population health disparities; designing public health interventions, programs, and policies; targeting interventions and programs to specific populations; evaluating the impacts and outcomes of interventions, programs, and policies; supplying feedback to providers of information; and supporting public health and healthcare personnel. The PopHR facilitates public health practice and public policy discussions by allowing authorized users to look at (1) multiple factors impacting health within a defined population in ways that are not possible when data are not brought together for analytic purposes, and (2) multiple aspects of population health itself, such as the levels and dispersion of disease, functional status, and wellness.

A national PopHR, containing data on multiple aspects of population health over time, would efficiently provide comparable information to support evaluation of the effectiveness of healthcare and public health programs on population health in multiple geographies and at multiple geographical levels, and maximize accountability of healthcare providers. In a recent solicitation by the Institute of Medicine to 20 000 stakeholders, respondents ranked comparative effectiveness research on health delivery and disparities as the two highest combined primary and secondary research areas,[28] both of which would be facilitated by the PopHR.

Although a PopHR would focus on population health, information in a PopHR could also be available for, and useful in, guiding the delivery of individual healthcare, including the diagnosis and treatment of specific health conditions. For example, having readily available information on the population prevalence of specific conditions would provide a context for the diagnostic process. Having information on community conditions and the prevalence of risk factors could assist healthcare providers in counseling patients about the prevention of future disease and injury and in recommending community-based interventions for current conditions. The PopHR could also provide information for comparative effectiveness research on the use and benefits of different therapies among different populations.

PopHR intended users

Primary intended users of the PopHR would be health departments, community-based organizations with responsibilities for public health, public health clinics and practitioners, and some healthcare providers. Secondary intended users of the PopHR would include researchers, the media, legislators and other policy makers, educators, community-based organizations, members of the public, employers, and payers.

PopHR content

A PopHR would usually contain information about the health of a single population, such as the population within the geographic bounds served by a state or local health department. Although many of the data collections and their resulting data sets serving as sources of content for the PopHR include identifiable information on individual members of a population, such identifiable data would not be stored in the PopHR or provided by the PopHR system. The PopHR would provide only aggregated data, such as statistics, measures, and indicators.

The content of a PopHR should ultimately reflect the full range of measures that describe the health and the factors influencing the health of a defined population (see online figure 1). Statistics, measures, and indicators describing the health of the defined population would be quantitative, including the level and distribution of diseases, functional status, and wellbeing.[vi] Information on factors influencing the health of the defined population would include the place and time under consideration for the population, its context (natural environment, cultural context, political context), and its community attributes (biological characteristics, built environment, health services, economic resources, population-based health programs, collective lifestyles and health practices).[18] Most of the information describing the factors influencing the health of the defined population would be quantitative, but some might be ordinal or nominal (such as the presence or absence or extent of enforcement of a particular policy).

Sources of PopHR content

PopHR content should ultimately derive from a wide range of information sources on population health and factors influencing it, such as ongoing population surveys, vital registration, public health surveillance, environmental sampling, Medicare and other payer claims, population censuses, and public health practice-related programs.[vii] Other population-based information sources include state-based hospital discharge and all payer claims data sets.

The extent and richness of nationally available information describing both population health and factors influencing it should not be underestimated. Many national data sets provide data for multiple geographic levels, using common data elements[viii] with standard definitions and controlled vocabularies across those geographic levels. Sources for such data sets include, but are not limited to, the Centers for Disease Control and Prevention, the US Census Bureau, the Health Resources and Services Administration, the National Oceanic and Atmospheric Administration, and the Centers for Medicare and Medicaid Services. Data sets describing state populations and

---

[vi] These measures and statistics encompass summary measures of characteristics within a population, such as mean blood pressure and median income, measures of incidence and prevalence, such as breast cancer incidence rate and proportion of families in poverty, and measures of dispersion. These quantitative measures require knowledge of the size of the population (denominator) and either the number of members of the population with a given characteristic (numerator) or the distribution of the characteristic within the population. Both the size of the population and the number of members with a given characteristic are frequently estimated by sampling the population. Populations with known or statistically estimated denominators and numerators are *defined populations*.[59]

[vii] Few, if any, US health departments currently use an overarching logical information model for collecting, storing, and retrieving their population health information, and instead process and store data in individual data sets. As a result, the population health data currently maintained by most health departments do not meet the basic requirements of a PopHR. Nevertheless, a health department could use a logical information model for creating a PopHR from non-standardized data sets so that the resulting PopHR would conform to the requirements.

[viii] In statistics the term ‘variable’ is often used instead of ‘data element’. We use the term data element to represent both terms.

Table 1 High-level health system objectives for an Electronic Health Record (EHR), as specified in ISO/DIS 18308.2, and for a proposed Population Health Record (PopHR)

EHR ID*	EHR objective	Proposed PopHR objective†
HSR1	The EHR should enable the consistent capture, processing, retention, protection and communication of health information, such that interoperability is achieved in support of shared care, improved quality of care, effective resource management, providing evidence of actions taken in health(care), and in support of the uses of anonymized information for health system management.	The PopHR should enable the consistent capture, processing, retention, protection and communication of health information, such that interoperability is achieved in support of shared *public health practice and clinical care*, improved quality of care, effective resource management, providing evidence of actions taken in *public health practice*, and in support of the uses of anonymised information for *public* health system management.
HSR2	The EHR should enable authorized users to access health information that is relevant, intact, and appropriate to their permissions and within a timeframe that is appropriate to the context.	See footnote‡
HSR3	The EHR should enable authorised users to access health information seamlessly and as originally organised, independently of the EHR systems and of the physical formats in which it was originally stored.	The PopHR should enable authorised users to access health information seamlessly, independently of the *PopHR* systems and of the physical formats in which it was originally stored.
HSR4	The EHR should enable the communication of all health information between care settings, subject to appropriate consent and access rights, to a sufficient quality to support safe shared clinical care.	The PopHR should enable the communication of all health information between *public health jurisdictions*, subject to appropriate consent and access rights, to a sufficient quality to support safe *integrated public health practice and* shared clinical care.
HSR5	The EHR should ensure that subjects of care receive the most appropriate care as quickly as possible.	The PopHR should ensure that subjects of *public health services* receive the most appropriate *services* as quickly as possible.
HSR6	The EHR should enrich audit & research activities within healthcare organisations.	The PopHR should enrich audit & research activities within *public health and* healthcare organizations.
HSR7	The EHR should support strategic planning decisions.	See footnote‡
HSR8	The EHR should support continuing health professional learning.	See footnote‡
HSR9	The EHR should support the workflow of clinical teams and care settings.	The PopHR should support the workflow of *public health* teams and *practice* settings.
HSR10	The EHR should help society move toward the practice of personalised or individualised medicine.	The PopHR should help society move toward the practice of *population-wide healthcare*.

The text taken from ISO/DIS 18308.2—Health Informatics—Requirements for an electronic health record architecture, is reproduced with the permission of the International Organization for Standardization, ISO. This standard can be obtained from any ISO member and from the website of the ISO Central Secretariat at the following address:www.iso.org. Copyright remains with ISO.

*These codes are internal unique identifiers used in ISO/DIS 18308.2 for the business objective statements, to assist in referring to them in other documents. They convey no specific meaning and bear no relation to identifiers used in any other publication.

†Changes in wording of the PopHR objectives from the wording of the EHR objectives are shown in italics.

‡The wording of the PopHR objective is the same as that of the EHR objective except for the substitution of *PopHR* for *EHR*.

collected and maintained by states sometimes have minor state-to-state variations in data element definitions and vocabularies; they may also have greater geographic granularity than similar data sets maintained for the nation, due to differing policies regarding confidentiality for the federal government versus individual state governments.

The PopHR information architecture (see below) should be sufficiently flexible to enable future inclusion of indicators, statistics, and measures from new content sources for defined populations, such as EHRs, PHRs, and data from health information exchanges, especially once basic demographic data elements are included and standardized.[29]

PopHR functionalities

Core functionalities of PopHR systems include analyzing, visualizing, and reporting information; querying information; generating alerts; generating reports on specific diseases or factors influencing health; identifying disease outbreaks and geographic clusters of disease or influencing factors; displaying temporal trends of diseases; and benchmarking. Once PopHR systems incorporate core functionalities, richer enhanced functionalities, such as social networking, wikis, user-generated content, and data mashups, could be added for particular PopHR uses.[30] [31] Providing context-sensitive help and supporting materials about retrieved information would be another important PopHR functionality. Such information would allow users to better understand and interpret population health information, such as the difference between counts and rates, and the importance of statistical stability and significance when comparing rates for different populations.

The PopHR content could be indexed and presented for users in multiple ways, with a particular user's view based upon user preference, role, or the particular population health issue of interest. For example, views of PopHR content could be generated for user-selected population subgroups, for a particular influence on health, or for a particular health or policy issue. Visualizations would include maps, graphs, and tables. The PopHR's analysis and display features could use geographic information system functionalities, as well as other interactive visualization tools already present on some websites.

PopHR business objectives

Business objectives for a PopHR would be similar to those specified for an EHR.[32] High-level health system objectives for a PopHR to support public health practice include providing timely—but controlled—user access to population health information, facilitating communication among public health jurisdictions, supporting strategic planning efforts, supporting training and continuing education of public health professionals, facilitating the evaluation of public health activities, and enriching research activities (see table 1). High-level public health practice objectives for a PopHR focus on safe and effective delivery of public health practice and enough flexibility to accommodate variations in practice methods, the advent of new population health problems, and evolution in the understanding of existing problems, which could lead to different approaches to practice.[32] The high-level citizen inclusion objective for a PopHR encourages citizen participation by providing access to information in the PopHR.

PopHR information architecture

In contrast with business objectives, fewer information architecture requirements for an EHR would be applicable to a PopHR. Applicable requirements include those concerning kinds of health record entries, representation of data values within health record entries, data retrieval and views, and communication and interoperability.[32]

Many information architecture requirements for an EHR would not be applicable to a PopHR, or would require extensive modification to be applicable, including those concerning the structure of health record entries, intra-record links, the representation of clinical process and workflow, ethical and legal issues, and confidentiality.[32] Because an EHR provides health information about, and is used to treat, individuals, its information requirements emphasize the workflow of clinical care, attestation, and maintaining information security and confidentiality. Although confidentiality must be kept in mind, a PopHR would provide aggregated health information about populations, and the risk of breaching confidentiality would be much lower than for an EHR, as long as basic safeguards were in place.[33 34]

PopHR system architecture

The specific purpose and intended uses of a PopHR will determine its overall system architecture. Specific components of the architecture of a PopHR system will be further determined by the population and time period that the PopHR will cover, and the information content and functions needed to meet the intended uses of the particular PopHR.

Several models for a PopHR system architecture can be conceptualized (see table 2; see figures 2 and 3, available online). These models differ according to: (1) the method(s) for compiling and processing information content for the PopHR; (2) the storage location(s) for PopHR information content; and (3) the location of applications and services[ix] that retrieve, process, and present PopHR information in response to user queries.[x] Because much of the content of a PopHR would be derived from other population-based data sets, a PopHR system would not typically engage in primary data collection, but rather in the analysis of primary data sets, followed by the compilation and processing of the results of these analyses to create the information content of a PopHR. This compilation and processing could be performed either: (1) prior to user requests for information from a PopHR; (2) 'on-the-fly' in response to user requests; or (3) using a combination of prior and on-the-fly compilation and processing. Because some population-based data sets that could serve as sources of information for a PopHR are quite large (eg, birth registration and census data sets), in order to increase efficiency and decrease the amount of time needed to retrieve information, a PopHR system might use intermediate data sets in which one or more large data sets is reduced in size by either selectively removing infrequently used data elements to form an 'abstracted data set', or pre-tabulating and indexing the data set on frequently retrieved data elements to form a 'pre-tabulated data set' on which statistical calculations could be performed on-the-fly as necessary in response to user requests (see figure 3 online). In addition to retrieving PopHR information, PopHR system applications and services would control access to PopHR information to maintain its confidentiality and security; generate statistical calculations, tables, maps, and graphs; and provide documentation about, and help in using, the PopHR and the PopHR system.

The PopHR could include measures and statistics stored and maintained at one node,[xi] such as a state health department, or at multiple nodes in multiple geographical areas in multiple organizational settings, such as a health department, a social services agency, an economic bureau, and an environmental protection department, or multiple state health departments. Table 2 illustrates four different approaches to information storage and retrieval for PopHR systems. The first model describes a system in which information storage and retrieval are centralized (see figure 2 online). An example of the first model would be an individual state health department or federal agency that controls its own data and computing resources, and centrally compiles, processes, stores, and provides access to health statistics and measures for the population it serves. Most federal agencies, and state and some local health departments, currently follow this model for their population health information systems. The second model differs from the first in allowing information to reside at multiple nodes, but is similar to the first in handling information retrieval through a single access point such as an individual state health department, federal agency, or independent organization (see figure 2 online). Although some health departments have more than one physical location and their system architectures might appear to resemble the second model, data sets at remote sites are usually processed to form abstracted or pre-tabulated data sets, which are then transferred to a central site where information is retrieved.[xii] The third and fourth models use distributed applications and services for retrieving information with either centralized or distributed information storage, respectively. An example of the third model would be an organization that compiles and stores information for one or more PopHRs in a centralized location, such as a single federal agency or a single state health department, but provides access to the PopHR content to other organizations that provide, in turn, their users with the ability to retrieve PopHR content through their own remote web and application servers.[xiii] An example of the fourth model would be information content stored in multiple state health departments, with applications and services for retrieving information distributed among those state health departments as well. The second and fourth models, with information storage distributed among multiple nodes, may be viewed as 'virtual' PopHRs; these models would require either standardization of information content or methods—such as the use of external standardized terminologies—to map and convert information into the local data elements used by each participating organization or jurisdiction.[8]

In any of these models, user access to information would be controlled by the systems and applications that process information requests and return results. Depending on the sensitivity of the information, access could be unrestricted for all users, restricted based upon specific roles of users and particular types of information (eg, access to preliminary disease measures for a given

---

[ix] Services includes web services "to support interoperable machine-to-machine interaction over a network" and middleware services that allow processes on one or more machines to interact and coherently function.[60]

[x] ISO 20514 states the importance of functional and semantic interoperability for EHRs and the need to distinguish between the EHR and the EHR system.[8] A similar need for functional and semantic interoperability, and for independence of the PopHR information model from the PopHR system implementation technology, exists for the PopHR.

[xi] A PopHR node is a single physical location where PopHR content is stored and maintained. Examples of possible nodes include state health departments and national data holders such as the CDC, the US Census Bureau, and the National Oceanic and Atmospheric Administration.

[xii] An exception to this is the DataFerret/DataWeb system maintained by the US Census Bureau in collaboration with the CDC. The DataWeb system constructs joined data sets 'on the fly' using geographic or demographic data elements that are common to the joined data sets. The data sets are maintained and stored in multiple locations throughout the United States.[61]

[xiii] If an organization's website simply points to another website that provides the actual PopHR compilation, storage, and retrieval function, it does not meet the definition of the third model.

Table 2 Models for Population Health Record (PopHR) system architecture

Model	Information storage	Information retrieval
1	Centralized	Centralized
2	Distributed	Centralized
3	Centralized	Distributed
4	Distributed	Distributed

time period might be restricted to authorized users within the organization producing that information), or restricted based upon the likelihood that a user could use the information to identify one or more persons and compromise their privacy.

IMPLEMENTING THE US POPULATION HEALTH RECORD

Barriers

Barriers exist to the development, implementation, diffusion, and maintenance of the PopHR in the US. These barriers are comparable in type, but often different in content, to those confronting implementation and diffusion of EHRs, PHRs, health information exchange, and health information technology generally.[35–39] Barriers in the US to the PopHR include data, financing, and policy and politics. The extent to which these barriers actually impede the PopHR will depend upon the specific information and system architectures for a particular PopHR, its enabling policies, and its strategies for phased development and implementation.

Data

Data timeliness and the periodicity of data collection and data release vary among population health data sets. In addition, standards for collection and coding, even of basic demographic data elements such as race and ethnicity, often differ among data sets and among the same or similar data sets held by multiple agencies and organizations.[40]

Financing

At least one and sometimes multiple WDQSs currently operate in each of 28 states, CDC, and other federal agencies with responsibilities for population health data.[25–27 41–43] These WDQSs provide access to a variety of population health data sets, and reflect the fragmentation of population health data collection and stewardship responsibilities among federal, state, and local governments. Existing WDQSs are not based upon explicit population health frameworks and standardized logical information models, do not provide comprehensive information content or functionalities, and do not support all the potential uses, of a PopHR. The PopHR(s) could replace these unconnected and uncoordinated WDQSs, and ultimately should decrease total investment in WDQSs. Nonetheless, the benefits of implementing a PopHR shared among states or a national PopHR might be viewed as favoring the federal government, with benefits accruing less to states and more to federal agencies and other users requiring comparative population health information across states and at multiple geographical levels.

Policy and politics

Perhaps the most basic barrier to the PopHR in the US is fragmentation of population health data collection and data stewardship responsibilities among federal, state, and local governments. Most data pertaining to population health are collected and maintained by state health departments and the CDC (see appendices 1 and 2 online). However, much data pertaining to the influences on population health are collected and maintained by a variety of state and federal agencies, such as the US Census Bureau, the Agency for Healthcare Research and Quality, the Centers for Medicare and Medicaid Services, the Environmental Protection Agency, and others (see appendix 1). Additionally, due to categorical funding of public health programs by Congress, fragmentation of data collection and maintenance within CDC and state health departments also exists, leading to multiple data silos in multiple governmental agencies. The consequences of this fragmentation include added difficulties in obtaining permissions to use data, standardizing data, and collaborating on the development of web services. Similarly, neither semantic interoperability nor standardized terminologies currently exist across population health data sets, and neither overarching models of population health nor logical information models exist across US health departments. Constructing connecting augers between data silos to facilitate a PopHR may be impeded by organizational cultures and restrictive policies relating to release of aggregated data that discourage sharing of data and increasing public access to data. Such policies and practices often vary among data sets within individual agencies and also between agencies. Finally, governmental concerns about the lack of a clear mandate in public opinion for sharing even aggregated and anonymized data via the web could also impede a PopHR.

Enabling factors

At least three factors could facilitate and enable implementation of a US PopHR.

The first factor would be a concerted and coordinated effort to portray to key stakeholders the PopHR as a public good that would enable monitoring, comparing, and evaluating national, state, and local health interventions and policies. The PopHR would also facilitate understanding of the social determinants of health at multiple geographical levels, as recently emphasized by the Blue Ridge Academic Health Group and the WHO.[11 12] The full value of the PopHR as a public good would derive most fully from its flexible inclusion of a wide range of population health information held by federal government agencies, state and local agencies, and ultimately by organizations responsible for the stewardship of EHRs. Providing easy and integrated access to a wide range of health data, as indicated by Booz Allen Hamilton and the Federation of American Hospitals, will increase data ‘liquidity’ and could ultimately enhance population health and healthcare.[44]

The second factor would be establishing clear and transparent governance structures for a PopHR(s), similar to those needed for public health informatics generally and for PHRs.[45 46] To the extent that the information content, services, and applications of a PopHR(s) are distributed, an explicit governance structure including key stakeholders will be needed; linkage or integration of PopHRs across different organizations or geographic jurisdictions would require either: (1) organizational agreements or government regulations to standardize population health frameworks and logical information models, and information content, processing and retrieval; or (2) flexible models and implementations that would allow the use of distributed applications or services to access and process information from multiple PopHRs. Models for governance structures already exist, including government led efforts such as the National Electronic Disease Surveillance System and the Public Health Information Network, government-led collaborative efforts such as caBIG, and consortia such as MedBiquitous.[24 47–49]

The third factor would be providing incentives for the development, implementation, and maintenance of PopHR(s). As indicated earlier, many federal agencies with population health data and at least 28 state health departments already maintain partially duplicative and overlapping WDQSs.[25–27 41–43] In fact, several federal and state agencies maintain multiple WDQS, sometimes separately for individual data sets. The cumulative federal and state costs of maintaining these WDQS—including hardware, software licenses, data cleaning and refreshing, applications development, and user support—could be transferred and conceivably reduced through constructing a PopHR(s). Using existing mechanisms, four types of incentives could enable development of PopHRs. First, an existing health agency, such as CDC, could serve as a national model by reorganizing its existing overlapping WDQS with identical geographical granularity into a single national PopHR based on an explicit population health framework and a logical information model. Second, current public health federal funding programs for states could provide incentives for collaborative development of a state or multistate PopHR and disincentives for continued development or enhancement of state-based WDQS not meeting the minimum definition of a PopHR. Third, new funding programs such as the Health Information Technology and Clinical Health Act (HITECH), part of the American Recovery and Reinvestment Act of 2009, could eventually fund state health information exchanges to provide anonymized EHR data for aggregation into a PopHR, and define 'meaningful use' to include provision of anonymized EHR data into a PopHR; similarly, the Key National Indicator System[xiv] could include development of a national PopHR. Fourth, voluntary certification of state and local health departments by the Public Health Accreditation Board could also mandate state health department use of a PopHR.

Phased development and implementation

Barriers to a US PopHR(s) could be minimized through a phased approach to development and implementation, with near (1–5 years), intermediate, and long term steps.[50] Near term steps would center on the development of a PopHR population health framework and logical information model, and the development and implementation of PopHR systems based on the first model of a PopHR system (see above), with centralized information storage, information processing, and system functions. Initial PopHRs and their supporting PopHR systems could build upon existing web-based data query systems in federal agencies and state health departments. Necessary changes to these WDQSs would include use of logical information models, overarching population health theoretical frameworks and, depending upon the particular WDQS, additional population health information. Near term development of the PopHR could either focus on nationally-held data that includes state and substate data, or on state-held data with eventual scaling to the nation.

Use of the first model for near term implementation of PopHRs would minimize political, policy, data, technical, and financing barriers, because information storage would be centralized for each PopHR, existing WDQSs could be further developed into PopHRs, and inter-organizational agreements, if needed at all, would relate solely to inter-organizational sharing of de-identified and probably aggregated data. The near term implementation of the PopHRs would encompass core functionalities only.

Intermediate term steps would center on the third model, with centralized information storage and distributed information and system functions. Depending upon the organizations participating in the PopHR system under the third model, political, policy, and financing barriers may exist relating to the need for inter-organizational agreements on logical information models, overarching population health theoretical frameworks, and technical standards for the construction of applications and services. However, centralized information storage would eliminate the marginal costs of distributed information storage and might minimize data standards issues relating to formatting. Planning for the distributed information and system functions for the intermediate term would begin during the near term implementation of the first model. The intermediate term implementation of the PopHR would encompass both core and enhanced functionalities.

Long term steps would center on the second model, which entails distributed information storage and centralized information processing and systems functions, and the fourth model, which entails distributed information storage and distributed information processing and systems functions. Both the second and the fourth models may enable the greatest flexibility in information sources for PopHRs, given distributed information storage. Distributed information storage could be especially helpful for use of aggregated data from EHRs, since in 10 to 15 years population-based EHRs may exist in some geographic areas and not in others. Given the complexities of distributed information storage, whether with centralized or distributed information processing and systems functions, barriers to the second and especially the fourth model would probably be greatest. The long-term implementation of the PopHR would encompass both core and enhanced functionalities.

CONCLUSIONS

National progress in fostering the development and implementation of EHRs and PHRs has not been accompanied by any concerted efforts to develop the PopHR. Due to increased recent recognition of the importance of a broad range of influences on population health[11 12] and passage of the Health Information Technology for Clinical Health Act and the Patient Protection and Affordable Care Act, factors and funding that could support development and implementation of the PopHR now exist. This viewpoint paper serves solely as an initial attempt to identify conceptual, definitional, design, and implementation issues relating to the PopHR. An important next step in the conceptualization, design, and development of the PopHR could be formation of a national expert panel, perhaps using a process similar to that used for a national framework for the secondary use of health data, to enumerate and reach expert consensus on major PopHR issues.[51]

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Institute of Medicine (US).** Roundtable on evidence-based medicine. In: Olsen L, Aisner D, McGinnis JM, eds. *The learning healthcare system: workshop summary*. Washington, DC: National Academy Press, 2007.
2. **White KL.** *Healing the schism: epidemiology, medicine, and the public's health*. New York NY: Springer-Verlag, 1991.
3. **American Medical Informatics Association, Board of Directors.** A proposal to improve quality, increase efficiency, and expand access in the US health care system. *J Am Med Inform Assoc* 1997;**4**:340–1.
4. **National Committee on Vital and Health Statistics.** *Assuring a health dimension for the national information infrastructure*. Washington, DC: Department of Health and Human Services (US), 1998.
5. **National Committee on Vital and Health Statistics.** *Toward a national health information infrastructure*. Washington, DC: Department of Health and Human Services (US), 2000.
6. **National Committee on Vital and Health Statistics.** *Information for health: a strategy for building the national health information infrastructure*. Washington, DC: Department of Health and Human Services (US), 2001.

[xiv] Mandated by Section 5605 of the Patient Protection and Affordable Care Act (Public Law 111-148).

7. **Detmer DE.** Building the national health information infrastructure for personal health, health care services, public health, and research. *BMC Med Inform Decision Mak* 2003;**3**:1.
8. **International Organization for Standardization.** *ISO/TR 20514: Health informatics—electronic health record—definition, scope, and context*. Geneva: ISO, 2005.
9. **Detmer DE,** Bloomrosen M, Raymond B, *et al*. Integrated personal health records: transformative tools for consumer-centric care. *BMC Med Inform Decision Mak* 2008;**8**:45.
10. **Friedman DJ,** Hunter EL, Parrish RG. *Shaping a health statistics vision for the 21st century*, Washington, DC: Department of Health and Human Services (US), 2002.
11. **World Health Organization.** *Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the Commission on the Social Determinants of Health*. Geneva: World Health Organization, 2008.
12. **Blue Ridge Academic Health Group.** *The role of academic health centers in addressing the social determinants of health. Report 14*. Atlanta, GA: Emory University, 2010.
13. **Evans RG,** Stoddart GL. Consuming research, producing policy? *Am J Public Health* 2003;**93**:371—9.
14. **Friedman DJ,** Starfield B. Models of population health: their value for US public health practice, policy, and research. *Am J Public Health* 2003;**93**:366—9.
15. **Krieger N.** Ladders, pyramids, and champagne: the iconography of health inequities. *J Epidemiol Community Health* 2008;**62**:1098—104.
16. **Last JM,** Spasoff RA, Harris SS, *et al*, eds. *A dictionary of epidemiology*. 4th edn. New York, NY: Oxford University Press, 2001:136—7.
17. **Young TK.** *Population health: concepts and methods*. New York, NY: Oxford University Press, 2005.
18. **Parrish RG,** Friedman DJ, Hunter EL. Defining health statistics and their scope. In: Friedman DJ, Hunter EL, Parrish RG, eds. *Health statistics: shaping policy and practice to improve the population's health*. New York, NY: Oxford University Press, 2005:3—23.
19. **Centers for Disease Control and Prevention.** *About bioSense*. Atlanta, GA: CDC. http://www.cdc.gov/biosense/ (accessed 24 Sep 2009).
20. **Diamond CW,** Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. *Health Aff* 2009;**28**:454—66.
21. **Lazarus R,** Yih K, Platt R. Distributed data processing for public health surveillance. *BMC Public Health* 2006;**6**:235. doi:1186/1471-2458-6-235.
22. **McDonald CJ,** Overhage JM, Barnes M, *et al*. The Indiana Network for Patient Care: a working local health information infrastructure. *Health Aff* 2005;**24**:1214—20.
23. **New York City Department of Health and Mental Hygiene.** *Primary care information project*. New York, NY: New York City Department of Health and Mental Hygiene. http://www.nyc.gov/html/doh/html/pcip/pcip.shtml (accessed 3 Apr 2009).
24. **Centers for Disease Control and Prevention.** *Detailed definition of PHIN*. Atlanta, GA: CDC. http://www.cdc.gov/phin/about.html (accessed 24 Sep 2009).
25. **Centers for Disease Control and Prevention.** *CDC WONDER*. Atlanta, GA: CDC. http://wonder.cdc.gov/ (accessed 15 Sep 2009).
26. **Friedman DJ,** Parrish RG. Characteristics and desired functionalities of state web-based data query systems. *J Public Health Manag Pract* 2006;**12**:119—29.
27. **National Cancer Institute.** *SEER cancer query systems*. Rockville, MD: National Cancer Institute (US). http://seer.cancer.gov/canques/ (accessed 15 Sep 2009).
28. **Institute of Medicine.** *Initial national priorities for comparative effectiveness research, report brief*. Washington, DC: Institute of Medicine, 2009.
29. **Institute of Medicine.** *Race, ethnicity, and language data: standardization for health care quality improvement, report brief*. Washington, DC: Institute of Medicine, 2009.
30. **Eysenbach G.** Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. *J Med Internet Res* 2008;**10**:e22.
31. **Hardey M.** Public health and web 2.0. *J R Soc Promot Health* 2008;**128**:181—9.
32. **International Organization for Standardization.** ISO/DIS 18308.2: health informatics—requirements for an electronic health record architecture. Geneva: ISO, 2009.
33. **Rudolph BA,** Shah GH, Love D. Small numbers, disclosure risk, security, and reliability issues in web-based data query systems. *J Public Health Manag Pract* 2006;**12**:176—83.
34. **Stoto MA.** Statistical issues in web-based data query systems. In: Friedman DJ, Parrish RG, eds. *Issues in evaluating web-based data query systems*. Princeton, NJ: Robert Wood Johnson Foundation, 2008:56—61.
35. **Blumenthal D,** DesRoches C, Donelan K, *et al*. *Health information technology in the United States: the information base for progress*. Princeton, NJ: The Robert Wood Johnson Foundation, 2006.
36. **Congressional Budget Office.** *Evidence on the costs and benefits of health information technology*. Washington, DC: Congressional Budget Office (US), 2008.
37. **Halamka J,** Overhage JM, Ricciardi L, *et al*. Exchanging health information: local distribution, national coordination. *Health Aff* 2005;**24**:1170—9.
38. **Middleton B.** Achieving US health information technology adoption: the need for a third hand. *Health Aff* 2005;**24**:1269—72.
39. **Middleton B,** Hammond EW, Brennan PF, *et al*. US EHR adoption: how to get there from here. Recommendations based on the 2004 ACMI retreat. *J Am Med Inform Assoc* 2005;**12**:13—19.
40. **Friedman DJ,** Cohen BB, Averbach AR, *et al*. Race/ethnicity and OMB directive 15: implications for state public health practice. *Am J Public Health* 2000;**90**:1714—19.
41. **Centers for Disease Control and Prevention.** *Welcome to WISQARS*. Atlanta, GA: Centers for Disease Control and Prevention (US). http://www.cdc.gov/injury/wisqars/index.html (accessed 15 Sep 2009).
42. **Centers for Disease Control and Prevention.** *Behavioral risk factor surveillance system*. Atlanta, GA: CDC. http://www.cdc.gov/BRFSS/ (accessed 15 Sep 2009).
43. **Centers for Disease Control and Prevention.** *Data access: health data interactive*. Atlanta, GA: CDC. http://www.cdc.gov/nchs/hdi.htm (accessed 15 Sep 2009).
44. **Penfield SL,** Anderson KM, Edmunds M, *et al*. *Toward health information liquidity: realization of better, more efficient care from the free flow of health information*. Rockville: Booz Allen Hamilton, 2009.
45. **Reti SB,** Feldman HJ, Safran C. Governance for personal health records. *J Am Med Inform Assoc* 2009;**16**:14—17.
46. **Yasnoff WA,** Overhage JM, Humphreys BL, *et al*. A national agenda for public health informatics. *J Am Med Inform Assoc* 2001;**8**:535—45.
47. **Centers for Disease Control and Prevention.** *National Electronic Disease Surveillance System*. Atlanta, GA: CDC. http://www.cdc.gov/nedss/ (accessed 2 Apr 2010).
48. **National Cancer Institute.** *Welcome to the caBIG® community website.* Bethesda: National Cancer Institute. http://cabig.nci.nih.gov/ (accessed 2 Apr 2010).
49. **MedBiquitous.** *MedBiquitous about us*. Baltimore: MedBiquitous Consortium. http://www.medbiq.org/about_us/index.html (accessed 2 Apr 2010).
50. **Stead WW,** Kelly BJ, Kolodner RM. Achievable steps toward building a National Health Information Infrastructure in the United States. *J Am Medical Informatics Assoc* 2005;**12**:113—20.
51. **Safran C,** Bloomrosen M, Hammond WE, *et al*. Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. *J Am Medical Informatics Assoc* 2007;**14**:1—9.
52. **Institute of Medicine (US).** Committee on Improving the Patient Record. In: Detmer DE, Steen EB, Dick RS, eds. *The computer-based patient record: an essential technology for health care*. Revised edn. Washington, DC: National Academy Press, 1997.
53. **Hammond WE.** *Electronic medical records—getting it right and going to scale*. New York, NY: The Commonwealth Fund, 2004.
54. **Hayrinen K,** Saranto K, Nykanen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform* 2008;**77**:291—304.
55. **Humphreys BL.** Electronic health record meets digital library: a new environment for achieving an old goal. *J Am Med Inform Assoc* 2000;**7**:444—52.
56. **Connecting for Health, Markle Foundation.** *Decision-making for population health—"first principles"*. New York, NY: Markle Foundation. http://www.connectingforhealth.org/resources?first_principles.pdf (accessed May 2010).
57. **Centers for Disease Control and Prevention.** *Reproductive health: glossary*. Atlanta, GA: CDC. http://www.cdc.gov/ReproductiveHealth/EpiGlossary/glossary.htm (accessed 19 Aug 2009).
58. **Kindig DA.** Understanding population health terminology. *Milbank Mem Fund Q* 2007;**85**:139—61.
59. **Friedman DJ.** *Assessing the potential of national strategies for electronic health records for population health monitoring and research. Vital and Health Statistics, Series 2(143)*. Hyattsville, MD: National Center for Health Statistics, 2006.
60. **W3C Working Group.** *Web services glossary: W3C working group note 11 February 2004*. W3C. http://www.w3.org/TR/ws-gloss/ (accessed 26 Sep 2009).
61. **US Census Bureau.** *Centers for disease control and prevention. dataferret: for the dataweb*. Washington, DC: Census Bureau (US). http://dataferrett.census.gov/TheDataWeb/index.html (accessed 6 Oct 2009).