NCVHS
National Committee on Vital and Health Statistics

December 5, 2012

The Honorable Kathleen Sebelius
Secretary
Department of Health and Human Services
200 Independence Avenue, S.W.
Washington, D.C. 20201

Re: A Stewardship Framework for the Use of Community Health Data

Dear Madam Secretary,

The National Committee on Vital and Health Statistics (NCVHS) is the Department of Health and Human Service's statutory public advisory body on health data, statistics, and national health information policy. NCVHS has historically made recommendations regarding stewardship of health information collection, use, and disclosure. This letter addresses data stewardship in the context of communities using data to advance health. It is a follow on to NCVHS's 2011 Report, *The Community as a Learning System: Using Local Data to Improve Local Health* [2011 Report].[1]

The communities whose efforts were highlighted in the 2011 Report identified the need for resources for stewards of community data. Stewards of community data are those responsible for collecting, managing, using, safeguarding, and disclosing information about community health, whether the information was originally collected through public health activities, research, patient care, or in some other way. Thus, there may be multiple data stewards, each interacting with the data at different phases of the data life cycle, and each with differing concerns, expertise, resources, and responsibilities.

Communities today are using digital data to tackle important health issues in ways that were not even imagined a few years ago. Dramatic changes are occurring in the use of data by and about communities. Digital data are growing exponentially, and data formerly stored in static documents are now being made accessible in flexible digital formats. Communities are eager to leverage these resources to increase understanding about themselves and opportunities for improving their health. These developments should be cultivated and guided by responsible data stewardship practices. HHS and state governments are seeking to disseminate their health data more widely and make it available for the public good. These developments should be cultivated and guided by responsible data stewardship practices. Individuals who participate in community health initiatives and the communities themselves as a whole must trust that their information is being used appropriately.

[1] NCVHS, *The Community as a Learning System: Using Local Data to Improve Local Health*, December 2011 [2011 Report].

In this spirit, NCVHS offers elements of a useful framework for effective community health data stewardship. This letter reports findings from recent NCVHS work, identifies eight elements of a framework to guide community health data stewardship, and makes four recommendations as to how HHS could initially facilitate effective data stewardship at the community level while preserving local autonomy and creativity in the use of health related information.

Background

The NCVHS Subcommittee on Privacy, Confidentiality and Security held a hearing on April 17-18, 2012, on governance and data practices of community health data stewards. Wide ranging testimony addressed the unique issues of different types of communities, the challenges of protecting privacy in small groups, the appropriate use and safeguarding of results, and community attitudes about data use. The Subcommittee also studied available resources that address data stewardship and Fair Information Practices.

The 2011 Report defined a community as "an interdependent group of people who share a set of characteristics and are joined over time by a sense that what happens to one member affects many or all of the others."[2] Communities differ in many ways. They have different governance structures, different needs, and different values.

Communities use health information obtained from many sources—clinical care, claims data, research, and public health, among others—in order to understand factors affecting the health and wellness of community members and how these might be improved. These community efforts often combine health information with other types of data such as environmental measures, climate measures, food availability and transportation options.

Current structures for protecting individually identifiable data rely on ethical principles such as informed consent, de-identification, or, in the case of federally funded research covered by the Common Rule,[3] waiver of consent through an Institutional Review Board (IRB) process. These approaches may not always be applicable or practicable for community health data uses.

Based on the original Code of Fair Information Practice produced by a 1973 advisory committee to the Secretary,[4] stewardship frameworks consisting of a minimum set of principles or best practices have been proposed for the protection of individual health information, primarily information initially obtained in patient care.[5] These frameworks are a useful starting point for identifying principles to consider when applied to today's broader uses of data by communities to improve their health.

[2] 2011 Report, p.8.

[3] 45 C.F.R Part 46 "Protection of Human Subjects."

[4] Advisory Committee on Automated Personal Data Systems, *Records, Computers, and the Rights of Citizens* (U.S. Department of Health, Education & Welfare), 1973.

[5] For a discussion of these stewardship frameworks, *see* NCVHS, *Health Data Stewardship: What, Why, Who, How: An NCVHS Primer*, <http://www.ncvhs.hhs.gov/090930lt.pdf>..

The Committee’s prior work on stewardship is also valuable. In 2009, the Committee defined stewardship as “the responsibility of ensuring the appropriate use” of information, and included “data collection, viewing storage, aggregation and analysis” in its definition of appropriate use.[6] It is clear today, however, that stewardship responsibilities extend even more broadly. A complete definition would include ensuring the integrity of data because appropriate use presupposes that data are of reasonable quality to accurately measure what they are intended to measure. It also would extend to a variety of roles across the full life cycle of the data: a data steward may design a study or initiative, collect data from subjects, analyze data for public health or research, share data with others, report on results, or repurpose data for a new and creative use. Moreover, more than one data steward may have responsibility for the same data across the life cycle, so that the steward who uses the data today may not be the same as the one who collected or created the data originally. Stewards have the responsibilities appropriate to their roles at the time they handle the data. Therefore, in order to capture this broader vision, we update our definition of stewardship as “the responsibility of ensuring appropriate collection, management, use, disclosure, or safeguarding of information.”

HHS is leading efforts to encourage innovative uses of health data for consumers and communities. At the April hearing, and at earlier NCVHS hearings conducted for the 20ll Report, testifiers told NCVHS that stewardship guidance is needed and desired and that a useful stewardship framework should account for community differences in a flexible manner. [7]

The Committee offers the following as a starting point for further work to develop stewardship resources that will enable community researchers, data users, and data subjects to understand the chain of trust required of effective stewards, wherever in the data life cycle their roles arise.

Elements of a Useful Framework for Community Health Data Stewardship

One of the most important and overarching goals of effective stewardship is to enhance trust in the processes of data collection, management, use, disclosure, or safeguarding. In the words of one expert who testified before the Subcommittee, “trust is our most important resource.”[8] This framework advances the understanding of stewardship by presenting elements that communities should address to put them on the path to effective health data stewardship. NCVHS also emphasizes that while stewardship represents universally applicable values, communities are diverse and the specific implementation of stewardship elements will vary.

[6] *Id*.

[7] *See* Transcript of Hearing of the Subcommittee on Privacy, Confidentiality, and Security, NCVHS, April 17, 2012, *available at* < http://ncvhs.hhs.gov/120417tr.htm>.

[8] Testimony of Kelly Edwards, PhD, Associate Professor, Bioethics & Humanities, University of Washington Schools of Medicine and Public Health, NCVHS Subcommittee on Privacy, Confidentiality, and Security, Silver Spring, Maryland (April 17, 2012).

1. Openness, transparency, and choice

One of the most basic fair information practices, first identified by the 1973 HEW Advisory Committee, is that there should be no data systems the very existence of which is kept secret. That Committee also stated that data subjects should normally get notice, at or before the collection of data, as to what is being collected and why, and what will happen to that data. These practices, and similar ones adopted around the world,[9] require openness and transparency. When individuals know what information is being collected and why, they know what to expect, have an opportunity to ask questions, understand the goal of data collection and use, and are more comfortable participating in these endeavors. The same is true of communities as a group.

The Committee heard testimony that openness and transparency promote trust. For example, a study shows that individuals almost always will consent to research if asked, but they still want to be asked.[10] While we are not suggesting consent is always a proper or feasible step before undertaking a community health initiative, some kind of outreach to subjects—and the communities with which they are associated—can smooth the way for a successful project.

Community health data stewards at our hearing requested help in developing appropriate methods to publicize data collection and uses. There are a variety of ways to inform people about community health initiatives, including posting information on public web sites, making information available in libraries or community centers, staffing information booths at community gatherings, mailing information to households, or individually notifying actual participants. These or other methods for informing communities might be appropriate depending on the circumstances of data collection and anticipated uses.

During the testimony, NCVHS heard that, in addition to being informed before a project begins, after the work is done, communities want to know how their data have been used, what has been learned, and the impact it has had. The Subcommittee also heard that communities resent when their data are used frequently but without efforts to share information that might be beneficial to them. Openness and transparency relate not just to informing communities and individuals at the beginning of an initiative, but keeping them informed throughout the time it is being conducted, and at the end to "close the loop."

2. Purpose specification

Another long established fair information practice is purpose specification. This means that data stewards consider the purpose of data collection and the future use of data collected at the outset of a project's design, and they make both initial purpose and anticipated uses explicit. This may include designs that anticipate re-purposing from the outset, for example, an initiative to survey and examine children in a study of the incidence of asthma where the data will later be used in

[9] *See, e.g.,* Organization for Economic Cooperation and Development, *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (*23 Sept. 1980); Canadian Standards Association, *Model Code for the Protection of Personal Information*, CSA Standard CAN/CSA-Q830 *(Mar.* 1996), *available at* <http://www.csa.ca/cm/ca/en/privacy-code/publications/view-privacy-code>.

[10] Ludman, E.J. et al., "Glad you asked: participants' opinions of re-consent for dbGap data submission," J. Empir. Res. Hum. Res. Ethics. 2010 Sep 5(3):9-16.

combination with air quality data to explore potential allergens or toxins in a community. A corollary practice is that when a data steward proposes to make significant changes from the originally identified purposes, she must evaluate whether and how to undertake such a course. And this principle pertains whether she is the same steward who originally collected the data or seeks to use an existing data set. Purpose specification encapsulates the ideas that data stewards should avoid causing unwelcome surprise and should consider changes carefully. If due consideration is made at the time of a proposed change, the data steward has the opportunity to weigh potential adverse effects on individuals and communities, and to mitigate any problems or concerns.

Community data stewards should also consider the balance between establishing a more specific and narrow scope or a less specific and broader scope when data are collected initially. The advantages of narrow specification are that the purposes are easily defined and described, so that communities and individuals may be more likely to trust the data steward and allow the desired uses of their data. This approach, however, may afford less flexibility. A researcher who specifies a more open-ended or unknown purpose gains greater flexibility for future uses but with the risk that individuals or communities may be surprised by future discoveries based on their personal information without separate consent, communication, or other form of outreach. Stewards need help in understanding how to strike these balances.

3. Community Engagement and Participation

Community engagement addresses the need to consider whether and how it is appropriate to involve communities in decision making about data collection and use.

Beyond individual consent, there are circumstances in which community consultation or even prior approval is appropriate. Where communities have established governance structures, data stewards should identify, respect and utilize these structures in all instances where required or feasible. In particular, the Subcommittee heard testimony from experts in American Indian tribal affairs about the importance of consulting tribal governments.

However, many communities do not have an obvious governance structure. The Subcommittee heard testimony about a variety of methods for collaborating effectively with such communities in decision making, including consulting with advisory boards and community leaders, identifying and actively engaging respected spokespersons, convening focus groups, and making data stewards available in locations where community members may learn about and give input regarding data uses. The panelists at our hearing stressed the importance of getting members of the community involved who can appropriately represent community views. Data stewards and communities both need information about these methods and best practices for various contexts.

4. Data integrity and security

Data stewards should ensure the security of their data, including availability for their intended use, their integrity, and, when appropriate, their confidentiality. Responsible data security starts with an evaluation of risks that can be anticipated and a plan to mitigate those risks by including protections that prevent loss of confidentiality, alterations of data that would make the data unreliable, or the unavailability of data when needed. These risks should be re-evaluated periodically.

Mitigation techniques include physical, administrative, and technical safeguards to ensure the integrity, availability, and, where appropriate, the confidentiality of data. Implementation of these safeguards varies with the data in question and the context. For example, encryption is a technical safeguard that may be applied to health information of identifiable patients to protect both confidentiality and integrity, but may not be needed to protect the confidentiality of population level data. Nevertheless, even population level data may require protection against physical destruction or malicious alteration.

Data stewards should ensure that data are as accurate, complete, and up-to-date as needed for the intended use. This long-standing fair information practice aims to ensure that information is not used in ways that might be misleading or untrue, and that incorrect data are not used to make inferences or decisions that might harm individuals or communities. Data integrity may be accomplished by carefully defining data requirements, understanding data sources, accessing trusted sources, and using verifiable practices for data collection, among other methods.

Much use of data to improve community health involves data that are not identifiable as to individuals. Data users also increasingly aggregate or merge data sets when using community health data. Stewards should employ reliable methods to aggregate, de-identify or merge data sets, and to ensure that data analytics result in valid inferences about a group. They should also be able to audit the data and implement processes to manage errors.

5. Accountability

Another fair information practice is accountability. Accountability requires identification of the person or entity responsible for stewardship at each point in the life cycle of data, from initial collection and use, wherever it is safeguarded, and following through to dissemination of results and, if appropriate, the safe archiving or disposition of raw data. These responsible entities should consider mechanisms for enforcement and redress in cases of failure to meet data stewardship responsibilities.

Failure to identify and address concerns regarding proper stewardship, and to ensure someone is accountable at every step, may lead to a variety of downstream consequences, some mild, and some quite serious. For example, a community that participates in a study but whose concerns are ignored may just refuse to participate in future studies. However, there have been cases where communities responded through their legislatures or the courts.

of the Havasupai tribe." The tribe also objected to many other studies having been published

Even more tragically, public outcry over the retention and use of blood spots obtained from newborn screening in both Texas and Minnesota resulted in the public health agencies of both states destroying, under court order, these valuable resources. The consequence was a result of parents' concerns that they had not been informed about the retention and use of data about their children, and had not been asked to consent in accordance with a state law.[12] In Texas, the legislature passed a law that now requires, for new collections of blood spots, consent from the parents and an opportunity to opt out.[13] In a New York Times article, the lead plaintiff in the Texas case, Andrea Beleno, was quoted as saying, "[t]he irony is if you had asked me, I probably would have consented. I would love for there to be a cure for breast cancer which runs in my family. I would love for there to be a cure for diabetes. The way the state went about it just made me distrustful."[14]

In 2010 the Institute of Medicine convened a workshop to address the challenges and opportunities related to blood spot research specifically citing the need for greater accountability regarding the operation of state newborn screening programs.[15]

6. Protecting de-identified data

Much of the data used in community health initiatives are not identifiable to a particular individual. In contrast, formulations of fair information practices have generally focused only on personally identifiable information. Similarly, data that have been "de-identified" in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy and Security

[11] *See Tilousi v. Ariz. State Bd. of Regents, Havasupai Tribe v. Ariz. State Univ. Bd. of Regents*, 204 P.3d 1063 (9th Cir. 2008), *review granted, review denied by Havasupai Tribe v. Ariz. State Univ. Bd. of Regents*, 2009 Ariz. Lexis 82. In March 2010, consolidated parties entered into a settlement agreement. *See also* Amy Harmon, "Indian Tribe Wins Fight to Limit Research of Its DNA" N.Y. Times, Apr. 21, 2010, *available at* <http://www.nytimes.com/2010/04/22/us/22dna.html?pagewanted=all&_r=0>; Paul Rubin, "Havasupai Tribe Win Nice Settlement From ASU in Scandalous Blood Sample Case," Phoenix New Times, Apr. 22, 2010, *available at* <http://blogs.phoenixnewtimes.com/valleyfever/2010/04/havasupai_tribe_finally_win_ni.php>.

[12] *See Bearder v. State of Minn.*, No. A10-0101 (Minn. App. Nov. 16, 2011), *available at* <http://www.lawlibrary.state.mn.us/archive/supct/1111/OPA100101-1116.pdf>; *Beleno v. Tex. Dep't of State Health Servs.*, No. 5:09-cv-00188-FB (W.D.Tex., San Antonio Division, filed Mar. 12, 2009); *see also* Peggy Fikac, "State to destroy newborns' blood samples," Hous. Chron. Dec. 22, 2009, *available at* <http://www.chron.com/news/houston-texas/article/State-to-destroy-newborns-blood-samples-1599212.php>.

[13] Tex. Health and Safety Code Ann. §33 (2008), *available at* <www.statutes.legis.state.tx.us/sotwdocs/hs/htm/hs.33.29647.78775.htm>.

[14] Harmon, Amy, "Where'd You Go With My DNA?" N.Y.Times, Apr. 25, 2010, p. WK1, *available at* <http://www.nytimes.com/2010/04/25/weekinreview/25harmon.html?pagewanted=all>.

[15] Institute of Medicine, Roundtable on translating genomic-based research for health. *in* Olson S, Berger AC, Rapporteurs, *Challenges and Opportunities in Using Residual Newborn Screening Samples for Translational Research.* National Academies Press 2010, pp. 51–54.

Rules are outside of the scope of its regulation.[16] The Privacy Act of 1974 and the federal regulations for the protection of human subjects (the "Common Rule") also only apply to research when identifiable individuals are involved. However, the Committee heard testimony that complete confidence in de-identification[17] may be misplaced. Data may, in fact, be re-identified intentionally or unintentionally, especially when data are combined from multiple sources or derived from small subpopulations.

The Subcommittee also heard testimony that there is a lack of clarity as to how communities should use de-identified data. Communities may wish to analyze their data at levels as local as census tracts, enabling inferences to be drawn about small groups. The Subcommittee heard testimony about cases in which communities object to particular uses of their data, even when the data cannot identify individual community members. The Havasupai case, described above, is an example.[18]

There are also concerns that communities need guidance about how to follow up to ensure the proper management of de-identified data when these data are available for public use. When de-identified data are made available for public use, a common practice is the "data use agreement" for downstream data users. Communities need help in understanding what these agreements should include and what kinds of follow up are helpful in assuring adherence to their terms.

[16] De-identification is usually understood in the context of the HIPAA Privacy Rule, which defines de-identified data as that which has been stripped of 17 specific identifiers and "[a]ny other unique identifying number, characteristic, or code" where "the covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information." 45 C.F.R. § 164.514(b)(2)(ii). However, there are other related concepts that may be used to discuss data that are not identifiable to a particular individual.

HIPAA also defines a "limited data set," which may include more granular geographic data and dates than de-identified information but may only be used for research, public health, or health care operations. 45 C.F.R. § 164.514(e)(1).

In the public health arena, most of the data used are not obtained from covered entities and, therefore, are not subject to the HIPAA Rules. Each data set may have its own set of disclosure rules requiring data use agreements, or may have no rules governing use. The use of identifiable data by communities is not prevalent, but *usually* requires a signed agreement with the data collector to use. The use may also be covered by the Common Rule, 45 C.F.R Part 46, depending on the funding, and so require review by an Institutional Review Board (IRB) or other privacy board. It may or may not require the consent of the subject.

Public health officials consider data that are not identifiable in two categories, confidential (or "private") and non-confidential (or "public"). Confidential data are derived from individual records or aggregate data from public health or clinical systems and eliminate direct identifiers. They may also include specific variables identified as 'confidential' by regulation or policy governing the collector. The release and use of these data invariably requires review and a signed agreement. This kind of information is being used more frequently by communities.

Non-confidential data refers to individual level data that contain no "confidential" information *or variable values that, alone or in combination with other available data, could re-identify an individual. Non-confidential data also often refers to* aggregated data that, when cross-classified with appropriate cell suppression or other data perturbation techniques, cannot be used to re-identify individuals. *Non-confidential* data generally do not require any type of agreement for use. These data are the most accessible and widely used by communities.

[17] Here we use the term "de-identified" in a generic sense, not specifically with regard to the HIPAA Rules, to include data that are not personally identifiable to an individual, and that are not confidential or private.

[18] Mello M.M. & Wolf L.E., "The Havasupai Indian Tribe Case—Lessons for Research Using Stored Biological Samples," N. ENGL. J. MED. 2010; 363:204-207.

7. Attending to the Risks of "enhanced" data sets

Community data uses may call for merging data from several sources to enrich the information available for planning and decision-making. However, these data "mash-ups" pose new and potentially unanticipated risks. The data set created by merging data from several sources may allow individuals to be identified or inferences to be drawn about members of small groups. Communities need guidance about how these risks are to be identified and handled. For example, some data "mash-ups" may raise greater red flags than others, such as the intersection of geo-location data with health conditions or health information or the inclusion of data from social media. Data enhancement may be especially problematic when people would not expect the data to be combined—for example, correlations between prescriptions filled, food purchases, and method of payment for food—or when the analysis of the data thus combined might have negative consequences for them.

The 2012 Federal Trade Commission report, *Protecting Consumer Privacy in an Era of Rapid Change*,[19] recognizes this problem and urges that it be addressed in the development of privacy by design. Communities likewise need guidance about the risks and management of enhanced data sets.

8. Stigma and discrimination

Community health data stewards should be alert to data uses that might result in discrimination against the community or its members. An example would be the possibility of redlining in the housing loan market if there is robust data about poor health status for a particular community. Data stewards should address ways to communicate findings that do not stigmatize or result in negative attitudes or behaviors toward the community.

Recommendations

NCVHS recognizes that the Department is making unprecedented efforts to facilitate the use of data. The Department is also playing a leadership role in establishing public-private partnerships to advance data use.

NCVHS believes now is the right time for HHS to advance a stewardship framework for community health data uses. Communities are innovating in their uses of health data, and governments at all levels are increasing efforts to disseminate their data.[20] These efforts can help communities promote improvements in health, health care, and other meaningful quality of life issues.

The Committee recommends that HHS:

[19] The report is available at <http://www.ftc.gov/opa/2012/03/privacyframework.shtm>.

[20] Many different entities are exploring these important developments. *See, e.g.*, Institute of Medicine, National Academies, *Digital Data Improvement Priorities for Continuous Learning in Health and Health Care - Workshop Summary* (Sept. 28, 2012).

1. Facilitate the development and promulgation of models for stewardship of community health data. HHS might facilitate this development through pilot projects, demonstration projects, grants, case studies, and the like.

2. Support the development of dynamic guidance resources that compile best practices that experts, communities, and other data users are learning about stewardship. NCVHS has identified the following high priority areas for resource development:

- How-to examples and case studies about de-identified data;
- Data use agreements and their enforcement;
- Risks of disclosure and data reporting, including information about risks of data aggregation and methods for protecting small groups when data are analyzed in small cells;
- Methods and best practices for openness and transparency; and
- Community engagement and closing the loop with communities.

3. Compile case studies of results that communities achieve through their uses of data so other communities might learn and be inspired.

4. Promote the creation of training materials for researchers who collect and use community health data.

Data-supported health improvement is a nationwide concern and a priority for the Department, but stewardship underpinnings remain weak. NCVHS is committed to advancing this stewardship framework and is considering how it can continue to help the Department study and implement these recommendations.

Sincerely,

/s/

Larry A. Green, M.D.
Chairman, National Committee
on Vital and Health Statistics

Cc: HHS Data Council Co-Chairs