Testimony To The National Committee On Vital And Health Statistics
~ Subcommittee on Privacy and Confidentiality~

**The Community as a Learning System for Health:
Using Local Data to Improve Community Health
Part II
Testimony and Public Comment**

**Denise Love
Executive Director, National Association of Health Data Organizations**

My name is Denise Love and I am the Executive Director of the National Association of Health Data Organizations (NAHDO). I want to thank you, on behalf of NAHDO for the opportunity to offer testimony about the issues related to the collection and release of identifiable health care data to stimulate improvements in health care cost, quality, and access.

The National Association of Health Data Organizations is a non-profit membership and educational organization, established in 1986, to promote the public availability of health care data and improve statewide health care surveillance systems.   NAHDO and its members have been at the forefront of establishing policies and processes for disseminating health care data for diverse uses ranging from research, public and community assessment and improvement activities.

The need for timely, comprehensive, and robust information has never been more critical. Today's applications demand more of our existing data, yet our data systems fall short; today's data sources do not capture complete information in a single system. Information gaps can be filled through strategic linkage across data sets, but there are numerous challenges posed by these linkages.   The need for standards in the collection of identifiable, granular data in uniform formats across data sources as well as standard release practices will facilitate the elimination of information gaps and promote the integration of data across the system. To do so, major barriers must be addressed. We believe that lessons learned throughout the NAHDO community can inform future data policies to improve the national information infrastructure.

My testimony will focus on the following three major areas:

■ The need for standardized patient and physician/provider identifiers
■ The need for improvement in data sharing and data exchange
■ NAHDO Recommendations to NCVHS

The need for standardized patient identifiers

We need identifiable data at the patient and provider levels to fully utilize our data sources. Most major public health data sets collect key fields in a slightly different format and the same data sets may vary in formats across states and jurisdictions. Because legal and political environments differ, there is variation in how patient fields are captured (or not captured), which hinders downstream analyses, linkages, and applications.

For example, many state hospital discharge data sets have relied on the patient's social security number (SSN) as a component of a suite of patient fields that, when combined with dates (birth, admission, discharge), create a unique record. As fewer payers and providers collect SSNs, there will be implications for the states that rely on this field.

For political or legal reasons, many of these same entities may not capture patient name and address; further diminishing the capacity to link data across data sources, time, and providers. Name fields formats vary across data sources and the standard may not accommodate hyphenated names. Further complicating matters, states that do capture patient address, have found noise in that field---no distinction between mailing and residential address is accommodated in the standard, yet this distinction is important when geocoding.

These examples heighten the need for patient demographic fields to be collected in a standardized format across data systems, to allow for a standard encryption method that will assist with linkages as well as provide for common applications and measures.

The need for improvement in data sharing and data exchange

If a state/data source does capture unique patient fields, such as address, the sensitivity and concern around these fields often results in the lack of sharing of the identifiable data with other data agencies/sources, even if the agency is a sister agency within a public health department. In addition many public health authorities in different states and jurisdictions have difficulty sharing identifiable data across borders. This is sometimes due to stipulations in law other times it may be due to lack of resources or the will to find innovative solutions that offer benefits to both parties.

Many agencies deploy statistical and management controls to release an anonymized data set. Some agencies will de-identify the data using encryption (one-way hash of SSN and/or other demographic fields) and aggregate other fields such as dates. De-identification methods have proven to be an effective first line of defense to protect patient identity. Combined with data board oversight and data use agreements that stipulate authorized uses, states have created work-arounds to legal and political concerns related to the release and exchange of data. But de-identification and data use agreements limit the repurposing of the data for data linkage, measurement, and public health surveillance.

The room for improvement in data integration and exchange is huge. The variation in collection and release practices have resulted in a fragmented patchwork of data bases, and limit the utility resulting in large gaps of information about the patients and the populations environment, and exposure risks. We must continue to find innovative solutions to these

barriers to promote the integration of clinical and administrative data which have the potential to support robust measurement and surveillance.

NAHDO Recommendations
NAHDO welcomes a national discussion, led by NCVHS, about the complex issues related to privacy and confidentiality. What is essential is greater cooperation across states and jurisdictions and data sources/agencies. Working together, we can maximize the utility of our large-scale data bases by linking and enhancing across sources, adding more intelligence in the combined data with the potential of reducing burden to the providers and payers collecting the data.

- Lead the effort to develop messaging to the public and policy makers about the need for identifiable data and how these fields can be captured without compromising the patient's privacy expectations, yet improve the public's health through evidence-based decision-making.

- Encourage uniformity and consistency of patient demographic fields across public health data sets in terms of both the definitions and format. For example, The Health Information Technology Standards Panel (HITSP) is working on a demographic model to define a standard for name (National standards organizations can't accommodate names that have more than three parts, e.g. hyphenated names). Aligning the HITSP effort with public health needs for enhanced data will accrue benefit to many stakeholders including providers being asked to supply various entities with data, and patients seeking healthcare.

- Promote more collaborative discussions across programs and data systems to align data needs and standards. One example is the project underway by the Center for Disease Control and Preventions' National Program of Cancer Registries (NPCR) and NAHDO to identify priorities for harmonizing discharge and cancer registry data bases, this effort needs to occur across more data sets. Harmonization across the two data sets will facilitate linkages between discharge data and cancer registries and also has the potential for reducing the reporting burden of providers.

- Provider and physician identifiers pose other challenges to states; states using the National Provider Identifier (NPI) have discovered that physicians often have multiple numbers, and providers and payers differ in their coding and assignment of the NPI. Public health agencies should work with the Centers for Medicare and Medicaid Services (CMS) to improve the assignment of a unique and stable physician identifier, perhaps developing two separate fields, one for the physician that is unique, and then a separate field for facility/location. This would allow easy aggregation of individual providers cases, while still retaining capacity to measure physician groups.

- De-identification will continue to play a role in data release/data exchange, but we need more robust public data sets. NCVHS can establish national expert panels to establish an analytic framework to identify and design best practices for enhancing de-identified data sets. For example, for hospital discharge data sets, what indicators/flags can be added by the data agency that will enhance the utility of the de-identified data set? Examples are creating keys that indicate a hospital readmission to the same or different hospital or adding mother's medical record to the newborn record to facilitate maternal outcomes studies. Another example, is to add parity to birth codes, so that measurement related to birthing can distinguish first births from others, a key indicator related to elective deliveries and cesarean sections. NCVHS can lead the effort to identify new methods to add intelligence in de-identified data sets.

- We need a common definition or definitions of Personal Health Information (PHI). How we define personal health information (PHI) is changing with the advancement of technology and the generation of genetic and bio-signature data, resulting in the handling of more sensitive data.

- NCVHS can help states and public health agencies synchronize privacy messages and create model exchange policies to facilitate data sharing across federal, state, and private sector agencies. More precise data will improve public health practice and health services research, but introduce new challenges that could result in inhibition of data flows. For example, Geo-coding of the data enable us to connect the dots, but this added power poses challenges to patients privacy as well as data exchange.

- We need help with defining a legal standard for inter-state transfer or exchange of our data sources containing PHI. State and federal laws overlap and the current process is complex and onerous. These data gaps inhibit measurement activities, especially in markets where there is a great deal of cross-border migration.

NAHDO's members have been innovators in the aggregation of large-scale health care data bases, the release of public data products, including the release of comparative performance reports on providers, and enhancement of data through data linkage. As our members expand their reporting systems to include all payer claims data bases (APCDs), the issues of identifiers, de-identification methodology, and data exchange will add more complexity. NAHDO welcomes a larger dialogue on these issues, led by NCVHS.

I thank you again for the opportunity to provide this testimony.