

# **Improving Health-Care Statistics Through Electronic Medical Records and Health Information Exchange**

**National Center for Health Statistics, Centers for Disease Control and Prevention**

## **SUMMARY**

### **Goals:**

- Explore the potential for electronic medical records EMRs to enhance the nation's health-care statistics capabilities, and
- Identify ideas for further development and research

### **Vision for EMRs and Where We Are in Realizing That Vision**

Presenter: Karen Bell, Office of the National Coordinator for Health Information Technology  
Office of the Secretary, U.S. Department of Health and Human Services.

### **Vision**

A “patient-centric system” with health information technology (HIT) available anywhere, at any time in a secure and confidential manner and will include virtual care.

The achievement of this vision will:

- Streamline administrative costs
- Permit comparison of quality of care across providers
- Maximize health and medical information in a way that optimizes patient treatment
- Promotes preventive services such as cancer screenings

### **How to get there**

The federal government has the following tools at its disposal:

- Laws and regulations
- Purchasing clout, conditions of doing business
- Payment policy and incentives
- Influence/leadership

### **Other relevant issues**

Privacy, security, medical legal issues, culture change, workforce

Electronic health records (EHR)—what do we actually mean?

Possible meanings: Physicians within a specific system have access to an EHR, patients have access to EHR or to electronic EHI for their entire medical histories

Value of HIT—need to standardize ways to measure its value

## **The HealthCare Statistics Enterprise and Requirements for Successful Statistics**

Presenter: Catharine Burt, National Center for Health Statistics, Centers for Disease Control and Prevention

### Major Considerations

- What data elements are in EMRs?
- What is the extent of penetration of EMRs
- Who controls the data and how can we get access to it?
- What are the quality and consistency of data?
- What benefits may arise from using EMR data statistically?
- What privacy and confidentiality issues must be considered?

### HIT adoption

NCHS began tracking HIT adoption in ambulatory care settings in 2001. The original goal was to determine when adoption would be so common that our data collection systems would have to respond to it.

### Messaging standards

NCHS conducted a transmission study for the Emergency Department component of the National Hospital Ambulatory Medical Care Survey (NHAMCS-ED). The goal of the study was to determine which data elements in the survey were covered by “messaging standards” to ensure their interoperability. The ASC X12 837 Health Care Service Data Reporting Guide (HCSDRG) was used as a standard. The HCSDRG had similar or identical standards to the NHAMCS ED for patient demographics, diagnosis codes, e-codes, and expected source of payment. No messaging standards were available for most clinical variables collected on NHAMCS-ED. Some potentially valuable data such as charge data has messaging standards, although these data are not currently collected on NHAMCS.

### EMR data quality

Analysis of data from hospitals using and not using EMRs found that item non-response was similar for the different types of hospitals, but many questions remain unanswered about data quality.

Do EMR’s offer possibilities for improved data collection? For our surveys which sample individual physician encounters, the opportunity exists to get information beyond that encounter including clinical, financial, and provider data, as well as linkages to other data sources.

Privacy and confidentiality are also obviously important.

## **Data Elements in Electronic Health Records (EHRs)**

Presenter and discussion leader: Paul Tang, Palo Alto Medical Foundation and American Medical Informatics Association

### Presentation summary

For clinicians, the goal of using an EHR is to deliver high quality care. Use of data for other purposes (e.g., quality improvement, population reporting, clinical research) should be a by-product of its use in the care process. To produce data that can be used appropriately for health statistics, it is suggested that we re-focus on the population health questions that need to be answered first, then ascertain what data are needed to answer these questions and how best to capture these data in EHRs.

EHRs have data in a variety of domains that are standardized, but because not all the code sets are complete, use of local enhancements to the code sets prevents full interoperability among EHR systems without manual intervention (e.g., mapping of non-standard codes). For example, ICD-9-CM diagnosis codes do not include important clinical classifications of asthma that affect treatment decisions. On the other hand, ICD-10-CM does include the required codes for asthma.

There are many clinical situations, such as patient refusal of a recommended treatment, that do not have standard codes. Consequently, the providers use “dummy codes” which results in a lack of standardization. In addition, physicians may want to document information that is more specific than a given coding scheme allows. For example, they may wish to code drug allergies to specific ingredients rather than just the chemical class.

### Next steps

- Re-examine health statistics priorities to see if there are priorities that have not been implemented due to lack of accessible data that could now be served using EHR data
- Develop clinical measures that reflect population-health priorities
- Identify critical data elements for the desired measures
- Identify data gaps in critical elements
- Get data through EHRs as part of user workflow.

### Discussion

Re-examine health statistics priorities: The feasibility of using EHR data to assess population health and overall patient health was discussed. One related limitation is the lack of outcome data. Kaiser has made some progress in this area and envisions more in the future. Whether or not health care statistics priorities should be re-examined was questioned, as there are already federal processes in place to establish these priorities.

Quality of coding: Potential opportunities to improve upon coding include use of the National Library of Medicine’s RxNorm system for medications.

Reinforcement of good coding: Reinforcing good coding behavior by reusing data to improve user workflow improves accuracy and completeness of coding.

Because of interoperability limitations, EHRs rarely contain data across the continuum of care, but these data are important both to patient care and to obtaining population health statistics.

## **Accessing EMR Data: Where the data reside and how federal agencies could arrange access**

Presenter and discussion leader: Linda Kloss, American Health Information Management Association

### Presentation summary

Data from medical records are reported to many different federal and non-federal sources. Medical records data are put to many uses, fill many needs, and requirements from both public and private entities including for payment purposes and assessing health care quality.

There are numerous reporting requirements and measurement systems. They are not consistent or coordinated with each other.

To assess quality, bridging silos across patient settings is important. Trusted aggregators of data can pull data together across sources and supply it for purposes of assessing quality and creating transparency, which is of interest to public and private payers and for public health. These same data can be used for clinical research, policy research, and other types of performance evaluation.

A variety of activities are needed to produce health care statistics including data collection and capture, standards for interoperability, data exchange, aggregation, and analyses by different users. State health information exchange organizations take on the role of data capture and information exchange and sometimes data aggregation.

Tools available to achieve these goals are: laws and regulations, purchasing clout, conditions of doing business, payment policy and incentives, and influence

### Discussion

Health Information Exchange presents the opportunity to cross-settings to create patient centered data. The group affirmed the importance of having different types of data, including: patient-centered data, provider-centered data, cross-sectional data, and longitudinal data, limited standardized datasets that cover all patients, and more detailed data sets obtained from representative samples. The complementary nature of these data sources was discussed.

‘Trusted aggregators’ face legal barriers related to privacy and confidentiality that limit sharing with others, including the providers who contribute the data in the first place.

Factors that influence completeness and accuracy of data include the quality of the original documentation, the variability of performance measures and payer reporting, and availability of secondary databases at the state or local levels.

Data tend to be provider centric because the data are owned by providers. The proprietary interests and profit-motives of organizations that own data cannot be ignored. The market is changing rapidly. A set of requirements is needed to lay on top of the market, to see how the market responds.

Efforts to identify quality measures are occurring on a broad number of fronts at once, both public and private. The American Health Information Community (AHIC) has a workgroup to create interoperability standards for electronic quality measures.

## **Quality, completeness, accuracy, comprehensiveness, and timeliness of data**

### **Cost of converting data for use in health-care statistics**

Presenter and discussion leader: Mark C. Hornbrook, The Center for Health Research, Kaiser Permanente Northwest

#### Presentation summary

Electronic medical record (EMR) data are becoming common for health care delivery. EMR data are input by providers in the process of providing care. Health care statistics are derived from EMR data warehouses. The data requirements and idiosyncrasies for health statistics differ from those for patient care. EMRs are a boon to researchers, but using data designed for patient care for research purposes is challenging.

Access: Some businesses have adopted EMR systems for patient care but do not allow researchers access, while others allow access to varying amounts of EMR content.

Accuracy and completeness: EMR data are like hard copy charts. They are only as good as the people who input it. Many different people input it, with different levels of training and comfort with computers.

In capitated systems, providers have historically minimized diagnosis and procedure coding. Documentation incentives must be parallel between capitation and fee-for-services systems for semantic interoperability. HMO recording incentives include: Medicare compliance regulations, Medicare risk-adjusted payment formula, and quality assurance programs such as The Health Plan Employer Data and Information Set (HEDIS).

Rule out diagnoses are a problem for researchers. They may not be distinguishable from confirmed diagnoses unless the provider is assiduous in documenting diagnosis status.

Interoperability: Many data elements are not interoperable, even when the same vendor sets up the system. Home grown coding is typical; 97% of Kaiser Permanente clinics have home grown coding. The large volumes of information in text forms can be challenging given lack of standard coding. “Natural language processing” offers promising solutions.

Research infrastructure costs money. The cost of building data systems of value to research must be included into the cost of the provider system. Researchers cannot afford to pay for them, and the providers benefit from the researcher’s access to this data.

#### Discussion

Silos: Kaiser Permanente has not had much incentive to interface with systems outside their system. Collaborative work between VA and Kaiser would be helpful.

Within Kaiser, local knowledge is needed to interpret codes.

Each new component of a system requires specific knowledge and testing. For example, analysis of laboratory data requires knowledge of biologics.

## **Privacy and Confidentiality: How these concerns may affect use of data for health-care statistics**

Presenter and discussion leader: Marcy Wilder, Hogan & Hartson

### Presentation summary

Using EHR to produce health care statistics requires both technology solutions and policy solutions. The technology solutions are moving faster than the policy solutions, which is problematic.

The original purpose of the Health Insurance Portability and Accountability Act (HIPAA) was to establish standards for data exchange for the purpose of paying health claims. The legislation was not focused primarily on privacy or clinical concerns.

Congress had intended to enact privacy legislation in the future and at the time HIPAA was enacted simply added a few paragraphs requiring the Department of Health and Human Services (HHS) to develop regulations if in fact Congress did not enact privacy protections. Congress did not enact privacy legislation and HHS therefore developed the regulation.

The paramount goals of the regulations were to both protect privacy and ensure that the business of health care delivery continues unimpeded. Other concepts included public health purposes. When the regulations were first written, there was not much clinical data in electronic format, so there was very little focus on this aspect. In general, de-identified data can be used for many different purposes, but not identifiable data. HIPAA has very specific requirements for when data can be considered deidentified and often times date of birth and zip code must be removed. The regulation was amended to permit the use of a limited data set, which can include dates of birth and zip code, as long as there is a signed data use agreement ensuring privacy will be protected and limiting the uses to certain acceptable purposes such as public health or health services research.

Options for using identifiable data include for research:

- Get a waiver from the institutional review board (IRB) of the organization holding the data. This burdens the organization that holds the data.
- For public health purposes, state laws limit disclosure to specific people in the state who can get the data. These laws often don't account for the federal government getting these data.
- Federal rules were not made with a number of uses that we have in mind today. Policies should be re-visited and updated as appropriate.

### Discussion

International health regulations require the federal government to get local approval before reporting data to the World Health Organization. Some have suggested a federal law requiring local disclosure to the federal government in limited circumstances.

De-identified data are not useful for many of the purposes for which we can use identified data.

Non-covered entities are doing things that covered entities can't do. Policy-makers are examining what is not covered and what ought to be covered.

Federal working groups are looking at state and federal public health data requirements.

## Next Steps

Facilitator: Richard Kronick, Senior Service Fellow, NCHS, and Professor, University of California, San Diego

- Move to standardize data elements not now standardized, e.g., by working with relevant standard setting bodies to establish standards.
- Providers' adoption of EMR systems is currently too low to support their use for nationally-representative surveys of health-care facilities. At this stage of adoption, pilot studies and research are needed.
- Work with a few health systems that have EMR data, e.g., Kaiser Permanente, OCHIN, VA, and Geisinger, to obtain their data and test the process of using these data to complete the various DHCS surveys, at least NHAMCS, NAMCS, NHDS, NSAS, and perhaps the National Home and Hospice Care Survey. That is, imagine that we received EMR data instead of going through our abstraction process. How close could we come to gathering the information currently collected on DHCS surveys using these EMR data? We might also learn from this pilot project in what ways the DHCS surveys could be enhanced/improved if we had EMR data. Presumably, at a minimum, a great expansion in the number of sampled records would be possible, and, perhaps also, a substantial expansion in the number of data items that could be gathered.
- Develop a case for the need for data to support the health statistics enterprise. The United States spends \$2.2 trillion on medical care and knows little about what we produce for it. As the country works through changes in payment and other factors that are likely to be required to move to a more sustainable annual increase in health expenditures, better data on what we are producing in terms of the quantity and mix of services and health-related outcomes will lead to more informed public and private decisionmaking. As is the case with economic statistics, these statistics are much more powerful if available at the small area level.
- Start by defining priorities about what we would like to know, and then identify the consequent data that the health statistics enterprise should want to collect.
- Look at the international experience described in Dan Friedman's recent study for NCHS, e.g., in the U.K., where almost all general practitioners (GPs) have electronic systems, there is more reporting of performance measures. Most U.K. hospitals, which are on budgets, do not have such systems.
- Consider the next step in data availability. The situation could be very different in five years. A shorter term strategy could explore what clinical data could be added to claims data, e.g., from Medicare, Medicaid, and private payers, to achieve incremental improvements and meet up with data from EMRs when they become more generally available.
- Consider starting with AHRQ's DECIDE network, which includes Kaiser Permanente, United HealthCare, and VA, to put a large group together.
- Analyze uses of data already compiled, e.g., the data base on immunizations at NCHS. Separate what sense one can make of the data when have them vs. data collection and sources of data to aggregate.

- Implement ICD-10. Look at number of codes reported and reporting requirements, though this step is separate from EMR considerations.
- Engender trust in governance of data to build trust in data sharing. To do so, need to involve patients and providers; the situation is competitive.
- Pilot creating a portrait of health care and health in a geographic area through health systems that use EMR systems and have a large fraction of people in an area, such as Intermountain Health Care in some of its markets or the Marshfield Clinic.
- Adopt an incremental approach that requires legislative and regulatory change over a long period. First, translate our world into specifications others would understand, e.g., at OCHIN. Second, test use cases, including a market approach or working with an integrated delivery network. Take an iterative approach, to learn, test, etc., to move the field.
- Make and articulate well the business or public policy case for any of the above ideas. An NCVHS 2003 report on a 21<sup>st</sup> century vision for statistics did not lay out the case well.
- Think in terms of specific policy issues, perhaps by focusing on overall health systems issues. Without a frame of reference, one will get only platitudes. The whole population side is not well known.
- Keep focus on something, e.g., need to test systems, experiment, deal with sources of data, including pulling data from vaults.
- Consider SAMSHA approach of collecting and putting data into a common framework itself, rather than trying to forge and implement agreement among all the states.
- Consider enlisting organizations that have received recent Clinical and Translational Science Awards (CVTSA) from NIH as partners for pilot work with EMR systems. Each needed to have a community part. For example, the application from Kaiser-Permanente and Oregon Health Sciences University (OHSU) concerned the state of Oregon and proposed improving health of the Oregon population. Kaiser-Permanente, OHSU, OCHIN, Providence Hospital, and VA would cover about 99% of the population.
- Bring people together under NCHS auspices to talk about how to improve population's health data.
- Build collaboration across federal partners to undertake activities together instead of doing the same things separately and all going after the same data. ONC has ongoing processes on standardization, harmonization, and certification criteria through priority setting, to build to its vision for the country. ONC is to work on aggregating data and use, and BioSense at CDC to get real-time clinical data. The Population Health Committee of AHIC, chaired by John Lumpkin, expects to get to public health statistics in 5-6 months. NCVHS could make this a priority. The federal Data Council could, too.
- Use EMRs to drive health statistics down to the local level, e.g., states, counties, and localities. Test formulating information at those levels, so there would be more useful information to go along with clinical information.



- Apply methods to look inside the black box of each EMR system to assess the data in it, e.g., coding and data entry, so know the quality of data for possible uses, e.g., pay-for-performance systems. The field needs methods to assess whether the data are statistically sound.
- Certify users, not systems. One cannot ensure the quality of the data by certifying systems.
- Explore whether some sites funded by a federal program would be willing to participate in another federal program to look at data quality, e.g., AHRQ and RHIOs, CTSAAs, at DHHS or VA levels.
- Use EMR systems to think boldly. Price will not work as a lever. For example, regarding coverage decisions, payment and coverage policies could be drivers for a business case for EMRs, and could affect the variations in use that Wennberg has found. EMR cost will be huge. Think about EMRs not as being in the control of the provider: to shift the balance of power more to the payer and to have the provider control cost. Payment policy may be used to get access to EMRs, and to gather information from them.
- Undertake public relations campaigns.
- Incorporate in planning the RFP that NCPHI, CDC, will soon issue to engage health information exchanges on how they will respond to public health, e.g., reportable diseases. The ideas will be tested over 3-5 years.
- Incorporate in planning that NCVHS from June will look at the framework for secondary uses of clinical data, including the likely benefit and why. Privacy issues are more complicated. The research community has come to HHS about problems and possible solutions that generally may be perceived as giving less protection to patients' privacy. Congress may look at this, but over a longer period.