Building Data Research Resources from Existing Data Sets: A Model for Integrating Patient Data to Form a Core Data Set

Harvey A. Schwartz, Agency for Health Care Policy and Research
Selma C. Kunitz and Rene Kozloff, Kunitz and Associates, Inc.

Dedicated to the Memory of Gerald Stanley Cohen

Containing health care costs and improving the effectiveness of care are today's leading health issues. Although there is uncertainty as to the extent of the current restructuring of this country's health care system, there is considerable agreement regarding the need for automated longitudinal patient data that describe both an individual's demographic characteristics and the site, process, cost, quality, and outcomes of care. Such data are needed to analyze which procedures are efficient and produce satisfactory patient outcomes. In our rapidly changing health care system, this information is necessary for decision-making by practitioners, policymakers, health plans, consumers, and patients.

In response to the critical need for information, recent efforts are focused on computerized patient records and health information services. It is anticipated that computerized patient records will serve multiple purposes. They will improve patient care by eliminating redundant data collection and providing more facile access to a patient's history. Computerized paper records and health data systems will provide data to:

❏ facilitate consumers' and patients' comparisons of aggregate data across provider entities, to make informed decisions regarding the use and costs of care;

❏ allow consumers and payors to track health care costs and utilization and to assess the nature and quality of care delivered;

❏ allow researchers to conduct studies aimed at improving the effectiveness and efficiency of care; and

❏ allow policy staff to develop methods for adjusting aggregate data by severity and risk descriptors and address public health issues such as access to care.

Utilizing data for an array of purposes raises many questions regarding what data are needed by whom and in what format. Different levels of detail are required for patient care, research, health policy, public health issues such as access, health care organizations' internal accountability and continuous quality improvement (CQI), and for review of costs and quality across provider institutions. Detailed patient-level data are needed for care while aggregate data by provider type may suffice for some policy questions. Detailed data by well-defined patient subgroups are needed for research studies of the care process and outcomes.

Currently, several efforts are focusing on the development of computerized patient records systems, health data networks, and a health care information infrastructure. However, there does not appear to be an identifiable locus of authority responsible for guiding these efforts, so that data will be usable across health care entities for care, research, and policy purposes. The planning, methodology, and coordination required to develop valid and reliable information are not widely discussed.

Developing uniform and consistent patient data sets that will not only document and benefit care but will also benefit researchers and policy requires several ingredients. Appropriate technology, such as computer-based clinical information systems and intersystem linkages is mandatory. Policies for appropriate privacy and security safeguards are necessary and methodologies that focus on creating

comparable, reliable data are critical. It appears as if the technology issues and some of the methodology issues are being addressed by the emerging health data efforts. However, there are numerous additional methodology and policy questions that must be addressed in order to create usable health data systems:

- ❑ What data are needed?
- ❑ What are the potential sources for these data?
- ❑ Are the necessary data currently accessible?
- ❑ Who will generate automated patient records?
- ❑ Where will the automated patient record reside?
- ❑ Who will "own" the automated patient record?
- ❑ Will guidelines for privacy and confidentiality be implemented and enforced?

This paper identifies efforts recently initiated as part of the health data systems focus, attempts to clarify underlying assumptions and definitions, and illustrates the need for policies to guide the creation of health data systems. Its objectives are to illustrate the importance of developing an action plan that can guide, synthesize, and integrate emerging efforts and to suggest an approach to developing a model patient record from existing data sets.

■ Current Status of Patient Records

Detmer et al., (1991) describes the inadequacies of the paper patient record in a report, *The Computer-Based Patient Record: An Essential Technology for Health Care*, from the Institute of Medicine (IOM) Committee on Improving the Patient Record. The shortcomings of paper medical records are numerous: they are fragmented, located in numerous health care settings, poorly documented, variable across providers, and often illegible. Information about a single episode of care may reside in a physician record (history and symptoms); in the hospital (laboratory, surgical procedures); and in a rehabilitation or nursing home setting. Even within a single provider institution such as a hospital, information, if computerized, is likely to be contained in different, discrete departmental systems that are not integrated.

A recent Agency for Health Care Policy and Research (AHCPR) sponsored study assessed the status of existing automated ambulatory medical records systems and found that they are primarily limited to academic settings (Grady and Schwartz, 1993). Further, the study found that community-based ambulatory records systems supported billing but retained little demographic, clinical, and outcome data. In another recently completed study supported by AHCPR (CHIS Project Summary, 1994), KAI (Kunitz and Associates, Inc.) found that, in general, community hospital clinical information systems do not retain patient data after discharge, cannot easily link the diverse financial and clinical systems within an institution -- or across inpatient and outpatient settings -- and do not support facile retrieval. Moreover, the portions of the clinical data continuum that are automated reside in separate systems, such as the laboratory, pharmacy, inpatient order-entry, and outpatient ambulatory systems. In addition, there is a lack of standardized terms and data formats. Practitioners' vocabularies and documentation are variable and do not have consistent operational definitions and associated codes for data elements. Patient outcome data are limited to disposition and do not encompass functional status, satisfaction, and quality of life. Further, there are no accepted measures for these descriptors. Health care facilities measures of performance are not defined. These studies document the limitations in accessing existing automated ambulatory and hospital medical records systems for patient care, as well as for study of the costs and quality of care and for documenting hospital performance.

Even the ability of advanced academic hospital information systems to link automated patient level data is limited. As one example, the computerized medical record on the HELP (Health Evaluation through Logical Processing) System at Latter Day Saints (LDS) Hospital in Salt Lake, Utah permits linking cost data from the financial system

to the clinical data. Yet, in order to link these data, researchers must access the cost data by transferring text files from one computer to another. Future developments at LDS include an electronic interface between these two computers (Evans et al., 1994).

Despite the evidence that there is a lack of accessible, reliable longitudinal clinical data, recent health care initiatives indicate that information will be available to identify quality and cost-effective health care organizations and that individuals will be able to utilize this information to choose among health plans and providers. For example, recent testimony (Hunter et al., 1994) suggests that most of the technologies we need to address questions regarding the costs, quality, and outcomes of health care already exist in both the public and private sectors. The testimony also suggests that facilitating linkages among automated systems currently utilized by health care providers will take us a long way toward a national health information infrastructure.

■ Overview of Ongoing Efforts

In response to the array of needs for health care data, a variety of attempts directed toward integrating health information, developing computerized patient records (CPR), and creating health data networks have been initiated. Some efforts are directed toward defining the scope of health information infrastructures. For instance, the Committee on Regional Health Data Networks of the IOM has focused on the creation of regional health data organizations (HDOs), while the American National Standards Institute (ANSI) and others are focused on identifying and developing standards. The Computer-Based Patient Record Institute (CPRI) has focused on gaining support for the CPR. The visions and definitions inherent in these efforts differ, although they each argue that computerized systems are needed to facilitate storage and access of longitudinal patient data integrated across the con-

tinuum of care from birth to death, while protecting the individual's privacy.

A recent IOM HDO study focused on understanding and improving the performance of the "health system" to guide the development of health data networks (Donaldson and Lohr, 1994). The study's report points out that data are needed to describe the health of the public and patterns of illness and injury; to identify unmet regional health needs; to document patterns of health care expenditures on inappropriate, wasteful, or potentially harmful services; to identify cost-effective care providers; and to provide information to improve the quality of care in hospitals and practitioners' offices. The report asserts that comprehensive, population-based secondary data that are collected over time are necessary to describe and risk-adjust patients' outcomes and processes of care. The report envisions that the longitudinal record will contain summary information from primary patient records that provides a synopsis of clinically significant health information.

A major goal for a health information infrastructure described by the Work Group on the Computerization of Patient Records (1993) is administrative simplification. The Work Group envisions that such an infrastructure would provide an interconnected communication network that links all participants in the health care system to reduce cost and improve the quality of patient care. Each provider would be connected via an automated patient record system that would also include links to medical literature and guidelines to improve care. Moreover, the infrastructure would include decision support applications and linkages between provider systems and change the delivery of health care. The Work Group defines the information infrastructure as the computer hardware, information system software, and the communication networks which allow access and exchange of information among practitioners and administrators; and linkages or transfer of patient data from one care facility to

another to coordinate services. In contrast, the Group defines a CPR as the data repository or storage of a patient's paper health record in a computer.

Detmer et al., (1991) provides a different concept, defining the computerized patient record as an electronic patient record that resides in a system, specifically designed to support users by providing accessibility to complete and accurate data, alerts, reminders, clinical decision support systems, links to medical knowledge, and other aids. This definition encompasses the data repository and the software necessary to access and utilize information. Detmer points out that the CPR could improve the quality and efficiency of information-intensive health care services and the quality of patient care, assist in reducing health care costs, and support research (Detmer et al.,1991).

The committees and work groups involved in health data systems development recognize that health care information standards are necessary in order to build automated patient records and a health information infrastructure. Further, these "top down" or senior management approaches justify the need and expenditures for a national shift toward automated health information and delineate recommendations. However, they do not consider the array of patient care information needs, address the paucity of standards for coding demographic and clinical process information, or provide a plan of action for development. The development process must consider the current status of clinical and related health information, clarify definitions, and detail the vision from the "bottom up" or from the operational health care organizational staff that will create the actual computerized patient records. The process must also define the contents of the automated patient record, uses, constraints, and necessary policies.

The ANSI Healthcare Informatics Standards Planning Panel (HISPP) efforts focus on ensuring formal coordination among health care informatics standards efforts in the U.S. and providing mecha-

nisms for responding to European standards efforts. HISPP coordination efforts are in transmission and content standards. The National Library of Medicine is sponsoring the Unified Medical Language System (UMLS) project. The UMLS is developing a translator among existing coding structures, such as the ICD9 and SNOMED, and terms found in the literature for use with patient record systems. However, Bishop and Ewing (1992) argue that these structures are inadequate because they were developed for specific purposes, contain a great deal of overlap, and do not easily accommodate to the ongoing expansion that reflects change in and growth of knowledge. The authors argue, instead, for a common language structure and codes, based on natural language associations, that can serve common purposes.

Existing committees and work groups specify that standards must encompass message transmission; field definitions, including format and structure; coding standards, such as the ICD-9-CM for diagnosis; and content standards that cover broad categories of information, such as demographics. These efforts generally reflect broad visions of regional and national health information structures recently summarized by Manderscheid and Henderson (1994) as depicted in Figure 1.

One effort has recently been sponsored by the Subcommittee on Ambulatory and Hospital Statistics of the National Committee on Vital and Health Statistics (NCVHS) and the DHHS Interagency Task Force on the UACDS. These groups have recently put forth the Uniform Ambulatory Care Data Set (UACDS) for consideration as minimum data to be collected in ambulatory care settings. Comments were solicited (NCVHS, 1994) and provided by members of research and administrative agencies, both state and Federal; professional health care associations, business groups, and labor organizations. They depict the differing needs among researchers, providers, policy analysts, constituency groups, and others. The comments also illustrate the difficulty in establishing consensus

Figure 1: Reformed Health Care Data Requirements
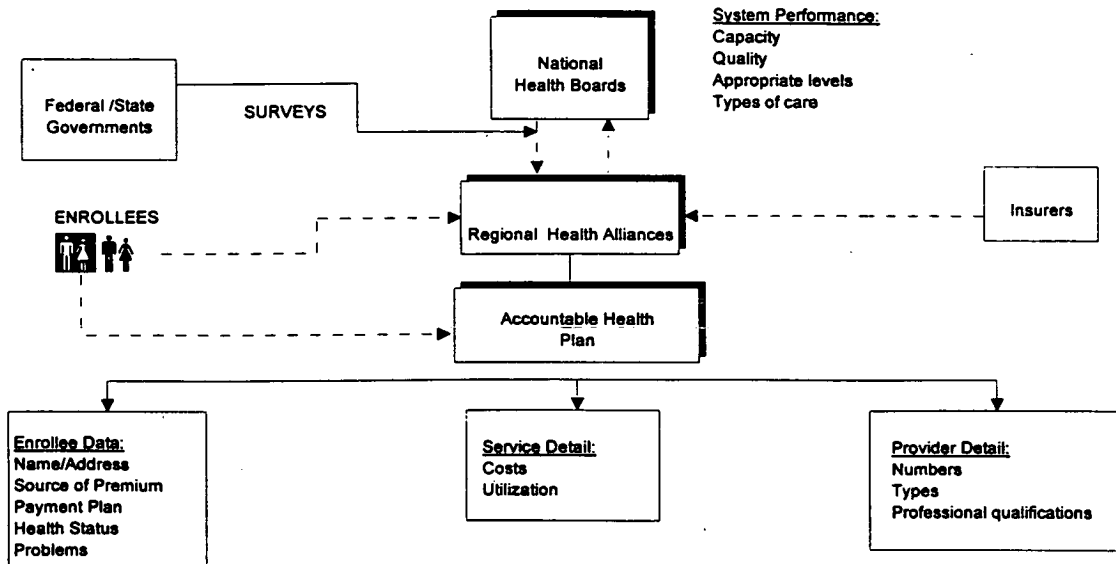
among groups with different needs and point to the need to set priorities in choosing a finite set of required data.

It is unclear whether or not there is a locus of authority that is leading and coordinating these discrete committee and work group efforts. Without coordination and direction efforts may be repetitive or even contradictory, rather than complementary.

■ Issues in Automating Health Data

The review of the above health system automation efforts reveals that there is little focus on the content requirements across health care settings. A conceptual framework has not yet evolved to integrate the content needs of automated patient records that document the care process, costs, and outcomes across the continuum of health care settings. Further, care data needs have not been meshed with policy and research data requirements, and data sources are not delineated.

The relationship between the detail of care that is contained within the paper medical record and the automated patient record has been discussed only in broad terms. As one example, the storage

of detail from computerized monitors and diagnostic technology (e.g., devices that continuously monitor intracranial pressure and magnetic resonance imaging [MRI]) has not been addressed. Simborg and Gabler (1992) recently recommended that the CPR should not become the automated repository of the paper record, but instead should be created from a systematic examination of the health care data collection process. An examination of the data collection process could identify areas in which efficiencies might be achieved by automation, and delineate data needs and uses.

Issues in Definition

Conceptual terms are also not yet consistently defined. For example, the terms, "minimum data set" and "core data set" are used interchangeably. It could be argued that a "minimum" data set refers to data that must be collected on every individual for "administrative" or payment purposes, while a "core" data set implies a set of well-defined and consistently collected data necessary for specific purposes or research questions.

Part of the confusion regarding available patient data may be caused by the lack of definition

of terms. Patient data in the form of **administrative secondary data**, or that required for reimbursement are available for Medicare patients and within some states. These data generally include the provider type, diagnosis, procedures, and charges for care and are used to describe mortality, costs of procedures, and geographic variations.

While these data may be imperfect and reflect biases because their primary use is for reimbursement, they are helpful for studies that focus on provider settings, reimbursement, and access to care and may be appropriate for many of the questions that the HDO regional efforts hope to address. Moreover, while the diagnosis and procedure codes contained in administrative or claims data sets are often referred to as, and are technically, "clinical" data, they do not describe the actual processes or details of care that must be examined to evaluate clinical practice and outcomes.

In contrast, patient care or **clinical data** are primary data collected as a routine part of patient care, in order to document a patient's demographic characteristics, medical history, symptoms, tests, procedures, treatments, and, sometimes, outcome. These data are contained in paper medical records rather than in an accessible automated format. They are often not readily retrievable, do not contain consistent data elements or codes, and are bound by privacy and proprietary concerns that often limit their use for research.

Patient level clinical data are necessary not only for patient care, but also for the continuous quality improvement and accountability activities of health care organizations. In addition, these data are essential for clinical studies, including medical effectiveness research and for the resulting health care policy. Clinical data can be used to describe and examine the decision processes regarding the choice and timing of tests, procedures, treatments, and outcomes of care.

Use of clinical data have contributed to decreasing mortality from over 19 percent to just under 10 percent in a four-year period, by reducing the time from admission for myocardial infarction to induction of thrombolytic therapy from an average of more than two hours after admission to 45 minutes (Jones, 1992). It is this level of detail that researchers and providers require to assess effectiveness and contain costs.

Technology

Database management systems (DBMS), optical storage technology, and record linkage software exist and could be used to automate the patient record and create health data systems. However, these technologies have not been widely applied within the health care setting. Thus, there are not systems in place that can support the CPR or that only need to be linked to provide a health data network (Detmer et al., 1991).

■ Automated Patient Records Model

In order to overcome the current limitations of clinical data and the lack of appropriate application of extant technology within the health care setting, an action plan is needed. We suggest building a model automated person-level record that captures data from existing sources within the health care system. The model-building process will identify:

❏ A **core data set** that contains socio-demographic characteristics, health status, encounters with the health care system, costs of care, and clinical symptoms, diagnoses, procedures, and outcomes;

❏ A **conceptual framework** that links patient care data across settings and organizational entities;

❑ **Mechanisms** for linking patient data, while considering the need for privacy and confidentiality;

❑ **Policies** that address data ownership, data linkage, the privacy and confidentiality of patient data, research usage, and other related issues;

❑ **Methodology** that helps to ensure accurate and reliable data; and

❑ **Technology** that must be explored to transform the model into a CPR that will provide functional requirements for health information systems and a data source for the health data networks.

The action plan for developing a model automated person-level patient record will:

❑ **Build on Available Data, Codes, and Technology** -- Although not optimal, some data in hospitals, laboratories, pharmacies, etc. are automated, accessible, and can contribute to the data contents of the model. Further, administrative systems contain accessible data and must be explored regarding the utility of the data and/or associated codes, for potential quality assurance contributions, and to test linkage mechanisms. For persons 65 and over, health related data exist in Medicare files. State vital statistics registries and national death files can likewise be used for their data, coding, validation, and linkage contributions.

❑ **Identify Gaps** -- The process of building the model will illuminate missing data elements, coding inadequacies, records linkage limitations, and automation requirements.

❑ **Conceptualize the Data Flow** -- The core contents of the patient-level data record can be used to stimulate discussion on the level of data needed for automated records that will support patient care and on the data that must flow into a regional and national health data system.

Model

An initial schematic of a person-level health or patient record is depicted in Figure 2. Data could come from the array of health care settings and, to

Figure 2: Patient Record Prototype
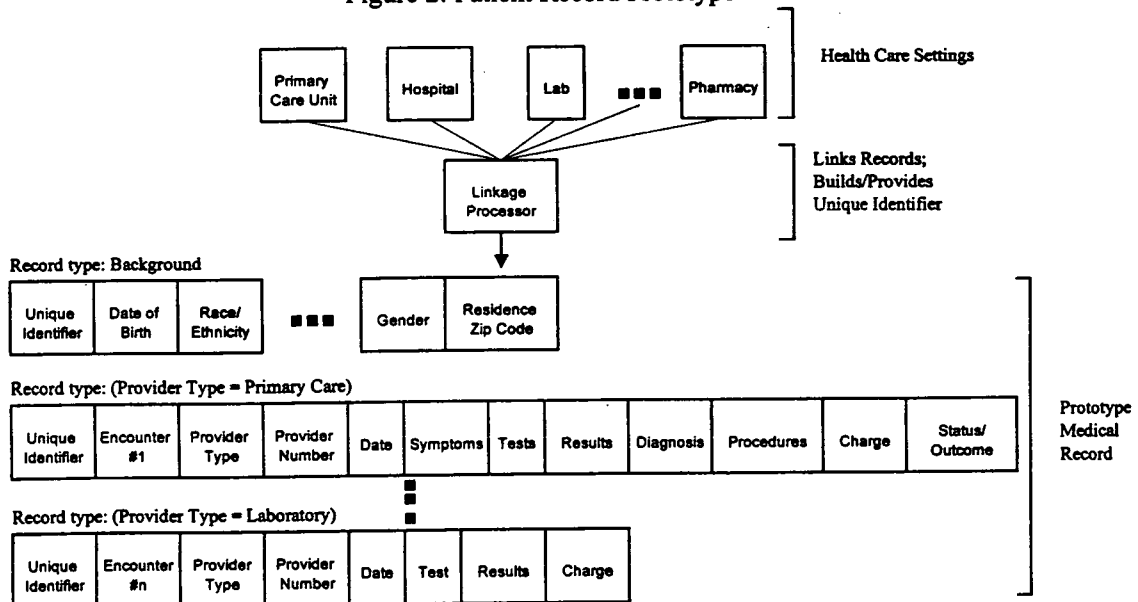
ensure privacy and confidentiality, data records would be linked using a processor that would also assign the unique patient identifier in the repository patient record. Record types would be coded, and would contain different information and differ by the type of provider from which they are derived.

The functions of the record linkage software program are outlined in Figure 3. It is anticipated that the patient identifying information would be housed in a person's primary care unit. The linkage processor would store the patient identifying data and create the unique identifier. It processes records from other providers and links the record as shown. The schematic can be used to delineate the necessary model-building steps:

❑ **Identify Core Data Set** -- The broad categories of data mirror the continuum of care and include person unique identifier, demographics, and history; encounter symptoms, problems, or diagnosis; tests, procedures, treatments, and costs; and patient outcomes. We suggest that for the model, specific data elements be derived from existing administrative data systems, such as the Medicare hospital discharge form (UB92, UB1500,) as it ad-

dresses many privacy issues and its strengths and weaknesses have been documented. For additional items that may benefit from a coding system (e.g., reason for encounter), suggested approaches could be derived from a review of critical literature that describes their development and use. It would be helpful to describe the core data set from two perspectives: those data that can be derived from existing data sets; and those that would be useful for research, but not currently collected. We suggest that expert-researchers review and comment on the proposed items.

❑ **Identify Existing Data Codes** -- Data codes are currently proposed or in use for numerous data elements. For example, ICD-9-CM is generally used for diagnosis and CPT-4 is used in ambulatory settings for procedures, while SNOMED III is used for laboratory findings.

❑ **Elicit Support** -- Work with "owners" of existing automated care data sets, such as vital statistics birth and death registries, and academic care settings to test linkage mechanisms.

Figure 3: Record Linkage Architecture

- 158 -
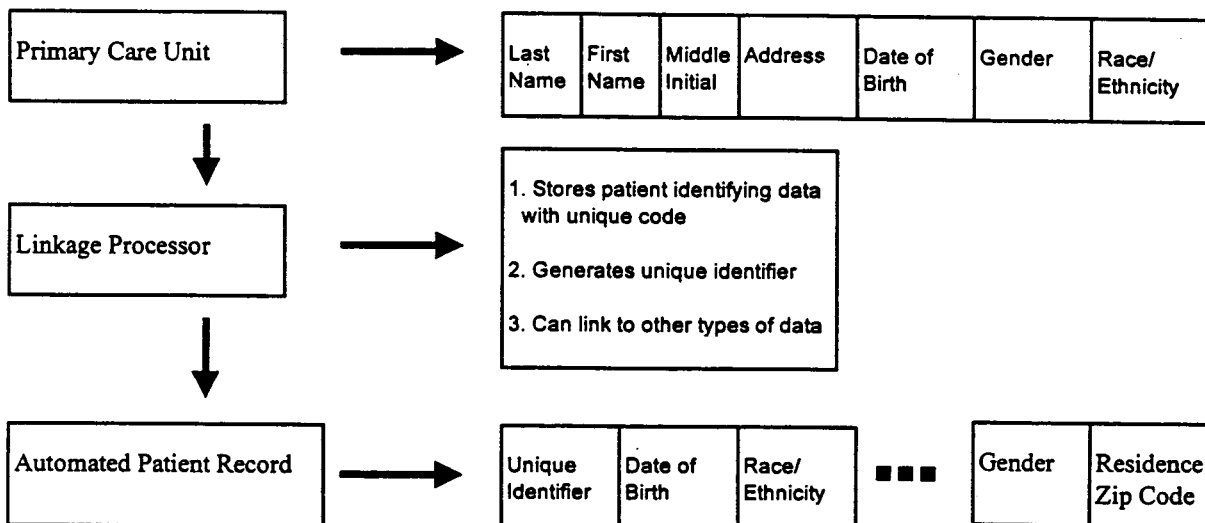
❑ **Utilize Linkage Mechanisms** -- Link data within institution and across health care organizations and data sets utilizing a unique patient identifier that masks an individual's identity in the core data patient record. Technical and process issues have been addressed for other applications, such as vital statistics (Roos et al., 1992), the census and the military.

■ Policy Issues

Developing an initial automated patient record model will help to refine policy issues. Policies must be in place to ensure that the integrity, accuracy, security, privacy, and confidentiality of the data are maintained. Over the last few years, efforts to explore policy issues and propose solutions have been undertaken by the Task Force on Privacy (DHHS), the Office of Technology Assessment, the CPRI, and various work groups in the Federal and state governments and private sector associations. The following policy areas and approaches to policy development synthesize these efforts.

Legislation

Legislation is needed in order to establish uniform and comprehensive standards to protect confidentiality and privacy. Preemptive legislation could cover all health information and records, except when information is provided for other than direct care services. Federal privacy protection should cover health care information, regardless of form, location, or user/holder, as the transition is made from paper to automated records. Preemptive legislation will eliminate the current problems magnified by patchworked state regulations and promote the protection of records as they cross state lines, and allow for improvements in health care/insurance transactions, in particular, and personal privacy, in general.

Legislation could set the stage for the establishment of uniform data standards, collection of a core data set, and guidelines for electronic data interchange, which in turn would enhance the development of a longitudinal patient record and the health information infrastructure.

Fair Information Practices

Legislation based on a code of fair information practices would ensure that individuals about whom data are collected have the right to know about the collection, the purpose of the collection, and the intended uses of the data; the right to review and correct any personal data; and the right to control and authorize the use of the data for any purposes other than originally intended. Additionally, the legislation would prohibit the existence of secret data systems and allow data to be collected and used only for legitimate purposes. As the data move to various parts of the health care system and beyond, they will remain subject to fair information practices.

For example, a health researcher may have a project that requires identifiable health information about patients. The plan must be submitted to the IRB, that will determine the necessity of using identifiable information. If approved, the researcher then becomes subject to the rights and obligations associated with protecting the privacy and confidentiality of the data.

This legislation should clearly delineate the rights of the individual about whom the data are collected and conditions under which disclosures by data collectors are allowable. It should also identify instances in which information is not protected by Federal legislation and is publicly accessible. A structure of penalties and an effective mechanism for enforcement of them should be developed and clearly defined.

Informed Consent

Inherent in a code of fair information practices is the ability of an individual to know about and approve the use of data about him or her. In order

to give informed consent for the use of data, the individual needs information about the content of the health records, the purpose of disclosure, to whom it will be made, during what period of time, and the safeguards against further, unauthorized disclosure. Informed consent does not preclude the use of data for research, but, along with other protections, such as a code of fair information practices and uniform protections across institutions and states, it will identify the rights and obligations of data subjects and data users.

Unique Identifiers

The use of a unique label (i.e., identifier) for identifying health related information pertaining to a particular individual is central to ensuring the smooth operation of the health care system; accuracy of information and data; reduction of burdens and costs; and improvement in the quality and coordination of patient care. Universal identifiers will support the development of a comprehensive, longitudinal computer-based patient record containing clinical, financial and research data. Unique identifiers must be thoughtfully developed and protected, so as not to become a threat to personal privacy or automation in the health care system.

While there is considerable agreement that each person enrolled in the health care system will need a unique identifier, the specifics of that number have not yet been determined. One policy concern is the way in which the number will be used to connect individually identifiable information from within a health care system with information outside the system. It has been suggested that when an individual's information is exchanged among health care providers or health plans, only the identifying number, and not the name will be transferred.

The choice of the unique identifier presents a challenge. While the social security number (SSN) is said to be the most cost effective and timely method, it has several important shortcomings. It is not unique and there are duplicate numbers and multiple users of a single number. It is difficult to

determine the validity of the SSN, because there are no check digits or other security measures applied to the number. Furthermore, the SSN is widely disseminated and used for a large variety of non-health related purposes, thus making it accessible for linkage with non-health related information. This potential ability to link many aspects of a person's life may, both in reality and in the perception of the public, allow for the creation of dossiers about individuals and growing threats to personal privacy and confidentiality.

In order to be considered an effective and safe number, a unique identifier would:

❑ **Contain check digits** to protect against accidental inaccuracy of the identifier;

❑ **Use encrypted digits** to allow for appropriate linkages, but not individual identification;

❑ **Have protection** through sanctions on its use and enforceable penalties for its misuse.

Because it is already so widely used, adoption of the SSN as the unique health care identifier would require the application of highly effective protections in order to control linkages with other non-health data systems. Developing clearly stated allowable linkages would complement the application of these protections if the SSN is used. However, developing a new number and restricting its use to the health care sector would ensure that each person's health number is of little use for linking health and non-health information but would impede the legitimate flow of information for research and decision-making. What appears to most important in selecting an identifying number is the development of a privacy framework that will protect the individual prior to implementation of any numbering system. Attention to the cost-effectiveness and time efficiency of developing and implementing a numbering system must also be considered.

Minimum and Voluntary Data Sets

Another policy issue is whether minimum and voluntary data sets can be established. While the NCVHS and Medicare provide examples of minimum data sets, incentives for voluntary data sets are unclear. Although researchers and policy staff argue that there are questions that can only be addressed by voluntary data collection, some provider constituency groups point out that these data must also have a "business need."

Data Linkages

To develop a CPR, links of an individual's medical records across care sites and over time must be established. **Record linkage** techniques utilize names, date of birth, social security number, gender, address, and other demographic characteristics to "match" an individual's discrete records. It is important to note that the characteristics used to match an individual's records do not need to be housed in an accessible automated patient record file. Instead, the unique patient identifier can be housed with identifying information in a "locked" or local file which is used for matching. The accessible CPR would contain the unique identifier rather than identifying characteristics such as name or address. Ownership and governance of this file of personally identifiable information and the unique patient identifier must be determined, and appropriate policy developed.

To further protect an individual's privacy for data that are transmitted from the health care setting to a regional or national health data system, age can be substituted for date of birth, and a location code that indicates a person's socioeconomic status, as derived from zip codes, can be utilized.

It is anticipated that in order to meet some research purposes, non-health data are needed to evaluate the effects of life style, income, and participation in governmental programs on health status. Policymakers will need to address the issues, approach to and wisdom of using linkage software,

and unique identifiers to link external files to CPR data at the individual level.

Data Elements and Codes

AHCPR has been sponsoring pilot studies of the ability to link data from discrete patient care data sets. One such study has attempted to integrate automated ambulatory patient record data from existing data collected in academic settings and document the relevance for investigating treatment differences for patients of differing race and ethnicity (KAI, 1994). Initial findings indicate that data are retrievable, can be compared, and contain some similar fields, such as race. However, the range of responses for race differs across the data sets from "white" and "black" to a more robust set of possible choices. Files from other sites indicate that the field does not always exist, while others may contain it but have little data entered in specific patient records. Integrated episodes of care or visits, procedures, tests, etc., linked for a particular diagnosis are not distinctly present. Policy should address the data contents and common coding structure in the CPR.

Data Ownership

As data proliferate in many settings, questions of ownership of both patient records and data arise. Current standards are ambiguous and create opportunities for data to be shared, sold, and linked without express consent from the patient. Several levels of ownership have been considered: ownership as the data subject, data collector and coder, and the possessor and holder of data.

The patient, as data subject, "owns" the data contained in a record, but does not own the record itself. Generally, record ownership is often state determined, with most states agreeing that the facility or provider owns the physical media in which the data reside and that it is subject to the patients' interest in the information contained in the record (American Medical Record Association, 1985). However, the legal concept of ownership is not

considered to be relevant where technology makes the creation of a record and its dissemination instantaneous. Instead, the rights and responsibilities of both the record subject and the record keeper with respect to the information must be delineated. In addition, protection of individually identifiable health information must safeguard the data as it moves within the health care system and to the various entities that will use it for care, policy making, and research.

Oversight and Monitoring Body

An oversight and regulatory body or panel could ensure that privacy goals are met and regulations are enforced. This body could be comprised of government staff, industry, academic, and health care representatives that are led by the Secretary of the Department of Health and Human Services. This body would set privacy and security standards; monitor and evaluate the implementation of these standards; sponsor related research; advise all health care system stakeholders, including consumers; support and participate in the development and implementation of informed consent forms and collection practices across all arenas; and work with the health care community to develop appropriate regulations and to satisfy all needs. This group can also serve as an independent judicial body, hearing grievances and allegations and adjudicating health information disclosure cases.

Education and Training

Privacy regulations can only be as effective as the bodies implementing them. Thus, the Federal government should take a major role in promoting the education of all participants in health care settings. The government should also take an active part in the review of current statutes, requiring educational programs, the determination of appropriate curricula, and the development of the necessary teaching materials for educating and training employers, employees and consumers. This effort should include sensitizing Federal and private sector leaders to the issues of privacy and the ethical

obligation to maintain confidentiality. In addition, large scale consumer education programs should be conducted, in order to increase general awareness of the customary uses of health care information and the civil rights of the individual with respect to personal information and privacy.

■ Summary

In response to the health information needs, there are emerging computer-based patient record and health information system efforts. There has been an array of related activities initiated that raise numerous issues regarding the uses of the data, the data that may be contained in automated patient records and health data systems, the sources of the data, and the current automated systems' capabilities to provide and link data in order to create automated longitudinal patient records and a health data infrastructure. Proposed policy and legislation suggest that automated person-level patient records and a health infrastructure must be a part of any reformed health care system.

Further, a conceptual framework that incorporates current data collection efforts and the continuum of data needs with health care, policy and research would clarify assumptions and increase the likelihood that the continuum of health data needs and efforts will be integrated into a successful and efficient automated health data system. Further policy needs would also be illuminated. In order to develop a framework that will encompass assumptions and the policies necessary for an automated longitudinal patient record and health data system, we suggest that current efforts be synthesized and that a model patient record be developed from existing automated systems that link data from various health care organizations and existing health-related records systems.

This paper has described various efforts, currently underway, intended to provide an automated patient record, health data system, and address policy issues. It also suggests an action plan that

relates the research needs for data with the resulting policy needs for information.

■ Notes

The views expressed in this paper are those of the authors and do not necessarily represent the views of the Agency for Health Care Policy and Research or the Federal government.

This paper draws on material from the authors' paper by the same title presented at the 1995 Joint Statistical Meetings in Orlando, Florida on August 15, 1995.

■ References

American Medical Record Association (1985). *Confidentiality of Patient Health Information*, American Medical Record Association, Chicago, IL.

Bishop, C.W and Ewing, P. (1992). Representing Medical Knowledge: Reconciling the Present or Creating the Future, M.D, *Computing*, 9(4), 218-225.

Detmer, D.E. et al. (1991). *The Computer-Based Patient Record: An Essential Technology for Health Care*, Committee on Improving the Patient Record, Dick, R.S.and Steen, E.B. (Eds.), Institute of Medicine, National Academy Press: Washington, D.C.

Donaldson, M.S. and Lohr, K.N. (1994). *Health Data in the Information Age: Use, Disclosure, and Privacy*, Institute of Medicine, National Academy Press, Washington, D.C.

Evans, R.S.; Classen, D.C.; Stevens, L.E.; Pestotnik, S.L.; Gargner, R.M.; Lloyd, J.F.; and Burke, J.P. (1994). Using a Hospital Information System to Assess the Effects of Adverse Drug Events, *Seventeenth Annual Symposium of Computer Applications in Medical Care*, C. Safran (Ed.), AMIA, 161-165.

Grady, M.L. and Schwartz, H.A. (1993). *Literature Review: Automated Data Sources for Ambulatory Care Effectiveness Research*, U.S. Department of Health and Human Resources, AHCPR Publication No 93-004, Rockville, MD.

Hunter, N.D.; Lasker, R.D.; Pelovitz, S.A.; Silva, J.; and Wallman, K. (1994). *Statement: Committee on Post Office and Civil Service Subcommittee on Census, Statistics and Postal Personnel*, U.S. House of Representatives, U. S. Government Printing Office, Washington, D.C.

Jones, F.G. (1992). Hospital's Continuous Improvement Campaign Reduces Heart Attack Deaths, *Medical Guidelines and Outcomes at Work*, 9-11.

Kunitz and Associates, Inc. (1994). *Evaluation of A Comprehensive Health Information System*, Project Summary, Unpublished Final Report.

Manderscheid, R.W. and Henderson, M.J. (1994). The New Informatics of National Healthcare Reform, *Behavioral Healthcare Tomorrow*, 11-15.

National Committee on Vital and Health Statistics, Subcommittee on Ambulatory and Hospital Care Statistics (1994). *Report on Uniform Ambulatory Care Data Set*, Unpublished Report.

Roos, L.L., et al. (1992). Record Linkage: An Overview. In Grady, M.L., and Schwartz, H.A., (Eds.), *Summary Report: Medical Effectiveness Research Data Methods*, AHCPR Pub. No. 92-0056, Rockville, MD.

Simborg, D.W. and Gabler, J.M. (1992). Re-engineering the Traditional Medical Record: The View from Industry. M.D., *Computing*, 9(4), 198-200.

Work Group on Computerization of Patient Records (1993). *Toward a National Health Information Infrastructure*, Report to the Secretary of the U.S. Department of Health and Human Services, Washington, D.C.

■ Bibliography

Brook, R.H. (1994). Health Care Reform Is on the Way: Do We Want To Compete on Quality? *Annual Internal Medicine*, 120(1), 84-86.

Gostin, L.O, et al. (1993). Privacy and Security of a Personal Information in a New Health Care System, *Journal of the American Medical Association*, 210(20), 2487-2494.

Hartz, D.R.; Kuhn, E.M.; Pryor, D.B.; and Krakauer, H. (1992). Mortality after Coronary Angioplasty and Coronary Artery Bypass Surgery (the National Medicare Experience), *American Journal of Cardiology*, 70, 179-185.

Houston Center for Quality of Care and Utilization Studies (1992). *Database Sources for Research in Quality of Care and Utilization of Health Services*, Houston, TX.

Kunitz and Associates, Inc. (1995). *Exploring the Race and Ethnic Data Available in AAMRS for Use in Medical Effectiveness Research Studies*, Unpublished Final Report.

Law, M. and Letts, L. (1989). A Critical Review of the Scale of Activities of Daily Living, *The American Journal of Occupational Therapy*, 43(8), 522-527.

Luft, H.S. and Romano, P.S. (1993). Chance, Continuity and Change in Hospital Mortality Rates, *Journal of the American Medical Association*, 270(3), 331-337.

Martin, K.S.; Scheet, N.J.; and Stegman, M.R. (1993). Home Health Clients: Characteristics, Outcomes of Care, and Nursing Interventions, *American Journal of Public Health*, 83(12), 1730-1734.

Mushlin, D.R.; Detsky, A.S.; and Phelps, C.E. (1993). The Accuracy of Magnetic Resonance Imaging in Patients with Suspected Multiple Sclerosis, *Journal of the American Medical Association*, 269(24), 3146-3151.

Pearlman, D.R.; Cain, K.C.; and Patrick, D.L. (1993). Insights Pertaining to Patient Assessments of States Worse than Death, *The Journal of Clinical Ethics*, 4(1), 33-41.

Spector, W.D. (1990). Functional Disability Scales. in B. Spilker, (Ed.). *Quality of Life Assessments in Clinical Trials*, Raven Press Ltd, New York, 115-129.

U.S. Congress, Office of Technology Assessment (1993). *Protecting Privacy in Computerized Medical Information*, U. S. Government Printing Office, Washington, D.C., OTA-TCT-576.

U.S. Department of Health and Human Services. (1993). *Health Care Financing Administration Fiscal Year Legislative Proposal on Computer-based Patient Records*, Washington, D.C.

U.S. Department of Health and Human Services. (1994). *Public Use Files Catalog as of January 1, 1994: Medicare/Medicaid Data Files*.

Waegemann, P.C. (1994). Toward an Electronic Patient Record, *Tenth International Symposium on the Creation of Electronic Health Record Systems and Sixth Global Congress on Patient Cards*, Medical Records Institute.

Ware, J.E. and Sherbourne, D. (1992). The MOS 36-Item Short-Form Health Survey (SF-36), *Medical Care*, 30(6), 473-483.

Weiner, J.M.; Hanley, R.J.; Clark, R.; and Van Nostrand, J.F. (1990). Measuring the Activities of Daily Living: Comparisons Across Na-tional Surveys, *Journal of Gerontology*, 45(6), S229-237.

Wennberg, J.E. (1985). Variations in Medical Practice and Hospital Costs, *Connecticut Medicine*, 49, 444-453. ∎