

Mining the CPR (and Striking Research Gold)

[Save to myBoK](#)

by Judith J. Warren, PhD, RN, FAAN, Marcelline Harris, PhD, RN, and Edward O. Warren, MS

How can HIM professionals help clinical researchers meet their goals? The answers may lie in the CPR. Here are some pitfalls to watch for -- and opportunities that may be knocking.

As the membership organization of health information management professionals, AHIMA fosters the professional development of its members (to promote) ...quality health information to benefit the public, the healthcare consumer, providers, and other users of clinical data. -- AHIMA mission statement

Consider all the possible users of clinical data. Primarily, you might immediately think of clinicians, reimbursement specialists, and payers. But this "user group" also includes researchers who access patient records for a variety of uses -- such as screening for subjects to enroll into clinical trials, determining best practices, and describing healthcare resource utilization.

How do health information management (HIM) professionals meet the needs of this particular group of users? By maintaining patient records, designing chart forms, providing patient records containing variables of interest, coding clinical data using administrative or other classification schemes, and abstracting requested information from the record.

These services have been very successful in the paper-based patient record environment -- and they will continue to be in demand. The advent of the computer-based patient record (CPR), with its capability of querying large databases, retrieving data on variables being studied, and downloading data into statistical analysis programs, means that if anything, researchers will increase their demands for patient data and information. In other words, the use of this data for research purposes will be driven by its simple availability. In turn, the ways the data is used will multiply -- to include areas such as patient outcomes research, complex epidemiological studies, and data mining.

The CPR, as described by the Computer-based Patient Record Institute, is "electronically maintained information about an individual's lifetime of health status and health care."¹ CPR systems expand clinical research capability by giving authorized researchers access to large quantities of health information while maintaining the confidentiality of patients and providers. This electronic clinical database supports the continuity of patient care, communicates patient plans of care, serves as a management resource, and extends clinical knowledge.

The promise of the CPR is that clinicians will record patient data at the point of care, use standardized terminology, and adhere to healthcare informatics standards. However, the rigor of clinical research will require more care with data and the way it is recorded. HIM professionals can meet this need if they understand the needs of researchers.

What Can Go Wrong: Understanding Threats

Clinical research studies determine the relationships between variables of interest, such as age, gender, risk factors, physiological parameters, and therapeutic interventions. The CPR offers the promise of conducting clinical research in a less expensive, less time-consuming manner than has been possible with paper-based records. It also makes it easier to study large, defined populations rather than smaller samples, which may provide new insights.

While recent advances in computer technology and networking capabilities have spurred the development of research using the CPR, experts are becoming increasingly aware of some important issues. Foremost among these are threats to the conclusion validity of the study. If not controlled, these threats contribute to errors in analyses, conclusions, and decision making. While many of the surrounding issues apply to observational and large database studies, they are particularly relevant when using the CPR for clinical research.

A researcher must determine threats to the validity of the study and account for them both in the design phase and the analysis strategies of the data. There are two general types of threats:

- threats to internal validity refer to the scientific adequacy of the study. In controlling threats to internal validity, researchers are trying to control potentially confounding effects of extraneous variables that may influence the events being studied
- threats to external validity refer to the "generalizability" of the study. Consumers of research findings examine external validity threats to make decisions about populations, settings, and treatments to which the findings of the study can be correctly applied

Significant validity threats related to the CPR include the sampling frame, the adequacy and accuracy of measures, instruments, and data, sample size, the appropriateness of statistical methods and accuracy of results, control of extraneous variability, and the comparability of groups.

The Sampling Frame

A sampling frame is a listing of every member of a population using the sampling criteria of the study. Threats to internal validity arise if:

- processes used to identify subjects within the study do not provide sufficient coverage of the range of possible values on important variables
- case identification strategies are such that all potential subjects are not available for inclusion in the study
- there is the potential to bias the conclusions of the study because chance differences in groups have influenced the effect observed

Investigators using CPRs will need to access data dictionaries, in effect, at the time of data acquisition to know how specific variables were defined. Similarly, they will need to review computer algorithms that identify cases. (How are the algorithms determined? By patient characteristic? By problem list? By standardized terms? Who determines and enters the information underlying the sampling strategy -- the clinician or a coder?) Investigators must carefully determine whether the sampling frame is adequate for the purposes of the study.

Adequacy of Measures, Instruments, and Data

How do investigators determine the adequacy of the instruments they use to measure variables? One way to measure the reliability of instruments is to look at characteristics such as accuracy, precision, validity, reliability, and responsiveness. These may be more difficult to evaluate in using data from the CPR, but electronic data acquisition methods do not alter the need to know this information. It's also important to understand how software handles patient data to ensure accuracy of its use and interpretation.

Ensuring data accuracy using electronic data acquisition processes requires the development of methods that detect, eliminate, and repair faulty data, ideally at the time of data acquisition. To date, there are no widely acceptable methodologies to either accomplish data entry at the point of care or to estimate the accuracy of data acquisition technologies.

When data is entered into the CPR using text, the provider, in essence, becomes the measuring instrument. For this reason, consistency in recording information becomes a significant reliability issue. For example, it is largely unknown whether socially undesirable variables (e.g., drug and alcohol use during pregnancy) are similarly recorded in CPRs and paper-based records. ICD-9-CM and CPT-4 codes lack the precision and accuracy for defining outcomes at the level of clinical detail required for outcomes analysis. What's more, the use of coding systems requires timely updates as new interventions and technologies are introduced.

A related issue is the completeness of data elements (sometimes referred to as core data) within a CPR. While there is no agreement on what constitutes complete data within the CPR, there seems to be consensus that resolution will require

agreement on two levels: essential data needs that apply to all health records and data needs specific to various end users.

This presents a moving target for several reasons. As knowledge development progresses, perspectives change on what constitutes essential data. Different clinicians require different types and views of clinical data. Furthermore, "business rules" often direct the determination and use of selected data for billing codes, utilization review, process improvement, and other administrative purposes.

Current database technology requires data modeling techniques that reconcile different views of data (e.g., relational databases). Everyone involved in the design and use of these databases needs to understand how and why different end users search patient records and make that data easily available and accurate.

Investigators who only occasionally use electronic databases and data warehouses will need clear navigation support and protections against events such as the loss of data. More experienced users will likely want to customize data retrieval processes to meet their specific needs. Futurists suggest that clinical data will eventually be linked to a wide range of virtual reality and bioinformatics applications (e.g., virtual reality models in surgery and the human genome project).

Sample Size

CPRs offer the potential for investigators to study large, defined populations rather than the smaller samples that are feasible in paper-based patient record studies (mainly because of the labor intensiveness of data abstraction from paper records). The challenge is to determine the right sample size. While small sample sizes may not adequately allow for detection of statistical significance, samples that are too large will find significance no matter what. Similarly, the size of different groups under study must be defined.

Appropriateness of Statistical Methods and Accuracy of Results

Classical statistical testing supports the analysis of data that meets assumptions of true experiments, such as randomness of error and normal distributions of data. While some statistical tests are fairly robust to violations of these assumptions, others are not.

An additional analysis issue is whether the multiple variables that may impact study conclusions are retrievable in the first place. Complex models with possible interdependence and interactions between variables are appropriate when analyzing large data sets. However, if the data is not in the database, it cannot be analyzed.

In addition to the need to use an appropriate statistical test, it is of utmost importance that the data being analyzed is accurate. Using inaccurate, biased, or invalidated data can result in misleading conclusions.

While many technologies support direct data entry in the CPR, there are few established methods to monitor and report the accuracy of data. Different data sources are likely to have different underlying data structures. There is a need for methods that detect, eliminate, and repair faulty data. The knowledge of domain experts is essential in developing validation methods, from testing algorithms used in data retrieval to creating software that integrates these data into accessible and usable information for research use.

Extraneous Variability

"Confounding variables" are factors that affect subjects or settings not addressed by subject selection or statistical adjustment. An advantage of randomized trials is that investigators can control for the effects of unmeasured factors through random selection and assignment. Few databases allow for analysis of differences in case mix on factors such as comorbidity or severity of illness. This is important because differences between groups may be due to case mix rather than the intervention or policy being examined.

Researchers can only use data that is collected, stored, and retrievable. Measures of outcomes such as functional and cognitive status, subclinical disease, and symptoms are noted to be particularly lacking (or lacking in quality) in most CPRs.

The effect of history, or events that occurred during the observation time frame of the study, may also introduce "confounding" to the study. The type of documentation may have an impact. Many institutions chart by exception, so that normal factors or factors influenced by the history of the institution are not documented. If these events are not in the record, the investigator will not only be unable to analyze the effect of history but may not even know there is one and thus generate wrong conclusions.

Comparability of Comparison Groups

The lack of concurrent controls is a concern in many observational studies. When a control is not available within the CPR, researchers may be able to use databases from other organizations as a control. This strategy demands the comparability of essential data elements across databases; to date, this has been difficult, as each organization measures its data differently.

Another strategy used by researchers is to examine clinical outcomes before and after any policy change (a nonconcurrent control group) to determine comparability of comparison groups. However, it is difficult to interpret conclusions from this design because of other changes that may have been occurring at the time of the policy change (the influence of history).

Horns of a Dilemma: Ethical Issues

All researchers who use the CPR as a data source are obliged to address the ethical issues inherent in this data collection approach. Both scientific and human subject review committees are required. Frequently, one of the first issues centers around determining whether a CPR study is the most appropriate source of data. This is not a trivial issue. If threats to the validity of the study cannot be controlled, formal clinical trials may be required. Data analysis, particularly when statistical tools are part of the CPR, presents another potential set of ethical challenges.

Another ethical challenge centers on achieving consensus about an appropriate and complete data set. Investigators must consider whether it is responsible to conduct studies, disseminate results, and perhaps change practice or policy when the data set on which a study is based is incomplete or very general. Generally, it is better to acknowledge uncertainties and confounding factors than to present a false patina of objectivity.

Access to the CPR presents many issues. Processes that establish electronic safeguards around authorization and authentication of persons obtaining access to the electronic record have been well described. Several studies suggest the CPR actually presents more security than paper-based records. The issue of informed consent has not received full discussion, particularly around process improvement projects later presented as research.

Critical Links: Healthcare Classifications and Terminologies

If researchers are to be able to use patient data from the CPR, then the implications for the accurate use of healthcare classifications and terminologies become critical. Current data entry is concerned with documenting care to support billing (using ICD and CPT codes supported by billing rules), regulatory requirements (meeting Joint Commission criteria), and risk management (documentation of high risk events). But documentation practice also needs to focus on patient-centered care documentation (What does the patient look like? What is happening to the patient? What are the patient's responses?).

This requirement, which supports outcomes management, decision support, and research, involves the inclusion of terminologies and classifications. HIM professionals should develop understanding and knowledge of the types and purposes of these languages (see below).

Common Classifications and Terminologies

Nursing-focused

- North American Nursing Diagnosis Association's (NANDA) Taxonomy
- Nursing Intervention Classification (NIC)
- Nursing Outcomes Classification (NOC)
- Home Health Care Classification

Medicine-focused

- Systematized Nomenclature of Human and Veterinary Medicine Reference Terminology (SNOMED-RT)
- LOINC (laboratory observations)
- ICD-9-CM
- ICD-10-CM, ICD-10-PCS

- Omaha System
- Patient Care Data Set
- AORN Perioperative Data Set
- CPT-4
- International Classification of Impairments, Handicaps, and Disabilities (ICIDH)

A variable in a research study has an allowable range of values, controlled by the researcher, that facilitate the aggregation and analysis of data. A data element in a database, in this case the CPR, has an allowable range of coded values that facilitate retrieval of information. Classifications and terminologies provide the allowable, coded data values for data elements such as patient problem lists, laboratory observations, patient signs and symptoms, therapeutic interventions, and outcomes.

Implications for use of these languages include:

- are the data collectors (clinicians and coders) accurate?
- is the use of the language consistent between clinicians?
- is the right data being captured and documented?

Furthermore, traditional classifications may not be granular enough to capture clinical data. For instance, a DRG or even an ICD code will not be specific enough to allow a researcher or clinician to describe and understand a patient's response. Terminologies such as SNOMED-RT may be required to represent patient data in the CPR. The terminology then can be mapped to the classification to meet the needs of aggregated data to support billing, resource consumption, and other regulatory reporting.

How HIM Professionals Can Help Researchers

As the patient record changes from a paper-based document that supports clinician communication, billing, and risk management to an electronic database that supports the above plus outcomes management, quality improvement, clinical decision support, and research, HIM professionals need to develop new skills. AHIMA has begun delineating these new roles and skills in the Vision 2006 project (see below).

Traditional and Emerging HIM Practices

Traditional Practice

Department-based

Physical records

Aggregation and display of data

Forms and records design

Confidentiality and release of information

Emerging HIM Practice

Information-based

Data item definition (including knowledge of classifications and terminologies); data modeling; data administration; data auditing
Electronic searches; shared knowledge sources; statistical and modeling techniques

Logical data views; data flow and re-engineering; application development; application support

Security, audit, and control programs; risk assessment and analysis; prevention and control measures

Source: *A Blueprint for the 21st Century*. Chicago, IL: AHIMA, 1998.

HIM professionals need to know the sources, accuracy, and limitations of patient data within the CPR. Researchers will consult with HIM professionals to determine data retrieval methods and to evaluate the accuracy, validity, and usefulness of patient data to support specific research questions.

This presents a new set of issues for the HIM professional: data design, identification of data elements, identification of data values, who can enter data, where data is entered, and data accuracy, integrity, validity, and business rules. Therefore, the HIM professional needs to be part of the CPR development team from its initial beginnings, involved in the training of clinicians and others who enter data, and involved in the training of people using the data -- in this instance, researchers.

HIM professionals can support research to improve patient care by becoming familiar with the terminologies and classifications of all healthcare providers. They can also take an active role in these systems' implementation in the patient record (whether paper or computer-based) and learn how to work with very large databases. And they can become an

invaluable member of the clinical research team. To do so, they need to become experts in the characteristics and limitations of the data in the patient record and take an active role in the design of databases and provider views -- specifically with an eye to data elements common across providers and their unique patient care perspectives.

This is an exciting time and great opportunity for HIM professionals to integrate the practice of health information management into clinical research projects. Knowing more about the research process, research data requirements, and standardized clinical languages will help HIM professionals achieve AHIMA's dynamic vision of improved healthcare through quality information management.

Notes

1. Computer-based Patient Record Institute. *Mission and Goals*. Bethesda, MD: 1997, p. 3.

References

Aaronson, L.S., and M.E. Burman. "Use of Health Records in Research: Reliability and Validity Issues." *Research in Nursing & Health* 17 (1994): 67-73.

AHIMA. *A Blueprint for the 21st Century*. Chicago, IL: 1998, p. 4.

Uddin, D.E., and P.R. Martin. "Core Data Set: Importance to Health Services Research, Outcomes Research, and Policy Research." *Computers in Nursing* 15, no. 2, Suppl. (1997): S38-S42.

Judith J. Warren is clinical nurse researcher and associate professor at the University of Nebraska Medical Center and College of Nursing, Omaha, NE. She can be reached at jwarren@unmc.edu. **Marcelline Harris** is a NLM post-doctoral Fellow at the University of Minnesota, Minneapolis, MN. **Edward O. Warren** is president of Warren Associates, LLC, Plattsmouth, NE.

Article citation:

Warren, Judith J.; Harris, Marcelline R.; Warren, Edward O.. "Mining the CPR (and Striking Research Gold)" *Journal of AHIMA* 70, no.7 (July 1999): 50-54.

Driving the Power of Knowledge

Copyright 2022 by The American Health Information Management Association. All Rights Reserved.