*Information Systems Frontiers 5:4, 429–434, 2003*


# *Data Systems in Support of Research in Managed Care*

***William A. Rush* and Patrick J. O'Connor***
***HealthPartners Research Foundation, PO Box 1524, Minneapolis, MN 55440-1524, USA***
***E-mail: William.A.Rush@HealthPartners.com***
***E-mail: Patrick.J.Oconnor@HealthPartners.com***

***Paul E. Johnson***
***Carlson Information/Decision Science, 3-365 CarlSMgmt, 7041, 321 19th Ave S, Minneapolis, MN 55455, USA***
***E-mail: johns021@umn.edu***

***Abstract.*** *The data systems supporting managed care are not structured to facilitate scientific research. We discuss the different types of research typically done within a managed care setting and the strategies successfully used by an affiliated research foundation to develop a data infrastructure to support a multi-investigator research mission. We describe our path from (1) diverse data systems supporting membership, claims, laboratory and pharmacy functions (2) the development of a corporate data warehouse for reporting and (3) the progression to a research data mart designed specifically to support research. We review the technical and resource issues encountered and how the evolution of the data systems for research paralleled and interacted with the research directions of the foundation.*

***Key Words.*** *managed care, data systems, health research, epidemiology*

## *Introduction*

Twenty years ago health care related studies of large populations required strenuous recruitment efforts, massive data collection, a substantial coding effort and constant validation. Large studies were few, expensive, and rarely replicated. Today in large health care systems with automated health care data, studies that would have been difficult, expensive or even impossible earlier may be routinely conducted. Managed care research efforts are perhaps the best example of this new reality in health care research (Briesacher and Erwin, 1995; Moy et al., 1997; Nelson, Quiter, and LI, 1998).

Managed care organizations (MCO) are not in the business to conduct research. In fact most MCOs in the United States do not get involved in any research (Cutler, 1996). When an MCO does support scientific research it is usually for a number of reasons unrelated to the business of providing health care. For example MCOs that were started as an association with some university will often carry the university's interest in research into their business. Others may start research endeavors because of a socially conscience CEO or board. However, MCO research is born, in order to survive MCO researchers quickly find that substantial support must come from outside sources. This is because when budgets are tight or profits are low, if any significant resources are being dedicated to research, they are quickly withdrawn.

To understand the information system needs of MCO based research one needs to understand that it is not research as it is usually imagined. When one thinks of scientific research in relation to medicine or health care the image that springs to mind for most is that of a laboratory with bubbling beakers and dedicated scientists in white coats looking through microscopes. This is not the reality of scientific research within manage care. Managed care research mainly deals with studies of health outcomes as performed in epidemiology, health services research or clinical trials. The majority of MCO research is devoted to the study of patterns of health risk and disease within populations or to interventions that attempt to improve the efficiency and effectiveness of health care services (Selby, 1997) (Fig. 1).

*To whom correspondence should be addressed.

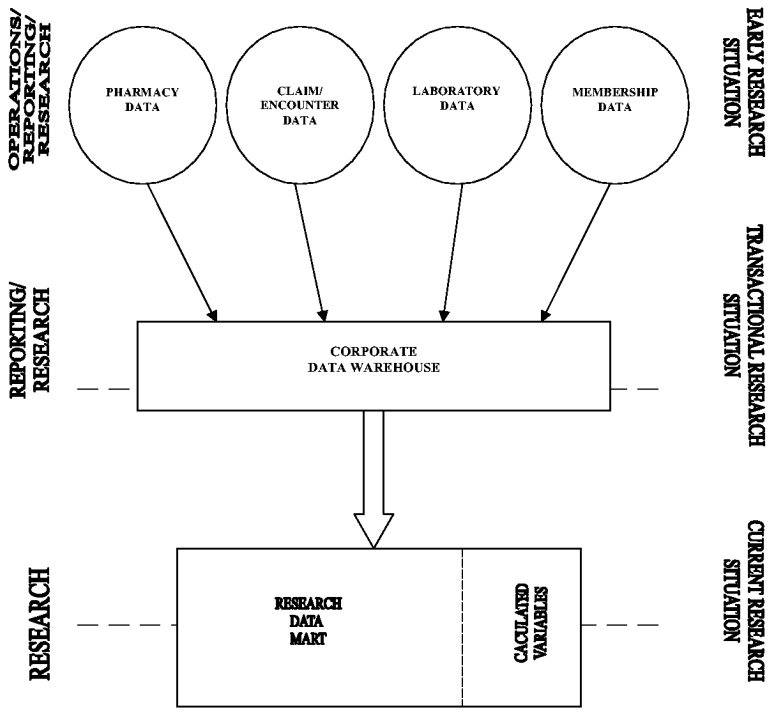


***Fig. 1.*** *History and structure of data systems used to support research.*

## *MCO Research*

Because of their large memberships and extensive data systems Managed Care Organizations (MCO) are ideal locations for many types of health related research (Lapham, Montgomery, and Hoy, 1990; Selby, 1997). Patients are tracked automatically by membership systems, while claims and encounter systems generate a continual stream of experience data. Cohorts can often be readily identified and followed over many years. Studies of relatively rare conditions may be possible because membership is often high enough that sufficient sample size can be obtained to make statistical analysis practical. In managed care organizations with a dental component, it has become possible to link medical and dental conditions at the patient level and investigate the impacts of one upon the other. The broad spectrum of differing physicians, clinics and medical groups under the managed care umbrella allows evaluation of the impact of organizational factors on quality of care, cost of care or clinical outcomes. Differing components of the federal government are also finding managed care populations useful for monitoring for possible adverse events associated with new medications or vaccines (Chen, DeStefano, and Davis, 2000).

A typical research project within an MCO might be to study the people with diabetes and how they are being treated. To accomplish this the research team will have to develop and merge data from a variety of data systems. *Membership* systems identify who is enrolled, for how long and what services are covered by the MCO. *Claims/encounters* systems provide diagnostic and procedure codes to identify diabetes and related conditions. *Laboratory* data systems provide information that can also be used to identify patients with diabetes and to monitor their clinical status. *Pharmacy* systems allow classification of patients by types of treatment and may allow monitoring of medication adherence (Steiner et al., 1988).

## *Strategies for Building Research Support Systems*

While managed care organizations may have six or seven figure memberships and collect terabytes of health care data; their data systems are generally not

designed with any thought towards research. Therefore while the potential is present, along with a great deal of system and data development it is necessary to establish personal and political links within the MCO. The following discussion deals with the political and technical experience at HealthPartners Research Foundation (HPRF), a non-profit research foundation associated with HealthPartners a large Midwestern MCO.

One of the early realizations at HPRF was that the data preparation phase of research shares many of the same aspects as other of the MCOs departments with reporting responsibilities. MCOs need to have data organized so that they can provide summary data reports to national certifying organizations, employer groups, and other departments within the MCO. Departments with a substantial level of reporting responsibility are the natural allies of research's efforts to build a more data analysis orientated data system.

The development of HPRF's present system started about eight years ago when research and reporting data were collected directly from the transaction processing systems used to manage the MCO. Because these transaction systems were not designed to carry the types of loads research processing produced, programs had to run at night when they would not impact normal operations. This had a braking effect on projects because whenever a job was executed there was a long wait before repeating it or executing follow-up programs. It became clear that this situation could not last and plans were initiated for the development of a data warehouse, to support all the reporting/research needs of the health plan.

After extensive study it was decided that a relational data base structure would bet suit the needs. Oracle was chosen as the vendor. The first design of the table structure was not satisfactory. Although the data would reside in a common environment it had exactly the same structure as the various transaction systems. These were systems that worked well for the retrieval and updating of individual records related to a single member but failed when the experience of thousands needed to be analyzed. The design team had focused on learning the nature and structure of the existing data and not on the needs of the users. After additional effort with more user input they produced a more usable system.

Warehouse data was taken from a number of disparate systems that were not maintained within the same or even similar data environments (Teach, 1996; Verma and Harper, 2001). Membership, claims and encounters were both Mumps based systems. The pharmacy system while also built in Mumps was a variation on a VA system using a different style of data storage, and the laboratory system was a proprietary system from an outside vendor. By bringing the data from these systems into a common environment a number of efficiencies were achieved (Center for Health Research, 1997; Verma and Harper, 2001). First there was a reduction in the number of different systems that needed to be accessed in order to acquire data. This meant that analysts no longer needed a detailed understanding of multiple systems. Since there could be a single extract from data sources, standard validity checks could be established and the potential for error decreased. Security could also be more efficiently managed with a single data source. In the past the analysts had to struggle with the problems of merging the data from multiple sources for each project. Member, provider, and facility identifiers might differ across systems. This had meant time consuming and tedious translation for each study. With the warehouse this was accomplished in a single effort when the data was loaded into the warehouse, by creating common identifiers to link/identify data from the various sources. The structure of the tables and the indexes created were optimized to support the types of data manipulations that research/reporting projects require. This could make the difference between an operation that took seconds verses one that took hours.

Research studies frequently require data to be available over significant periods of time, often as long as 7 to 10 years. Unfortunately, HP only needed to maintain active data for two historical years plus the current year. When the MCO no longer needed data for reporting it made good sense to archive it in order to save on storage space and reduce the time needed for reloads. However, HPRF needed to develop a way of maintaining data from the warehouse even after it was no longer present within the warehouse. At this point it became apparent that HPRF needed its own warehouse. But this warehouse only needed to be a subset of the data elements within the HP data warehouse. While the corporate warehouse had gone a long ways towards solving many of the difficulties that arose with dealing with data in the transaction processing systems, it was still serving many departments and was not optimized for any one user. Also it did not permit the addition of other calculated variables that research was increasingly dependent upon. Examples of these would be the Charlson score, a measure of the burden of comorbidity (Charlson et al., 1987; Deyo, Cherkin, and Ciol, 1992; D'Hoore and Tiquin, 1996; Redelmeier

and Lustig, 2001; Walter, Brand, and Counsell, 2001), a measure of medication adherence, i.e. how consistently does a patient take their medication as prescribed (Steiner et al., 1988) and a continuity of care score, i.e. a measure of how often a patient saw the same physician (Eriksson, 1990; Smedby et al., 1984). Each of these measures can be easily calculated from the data available within the warehouse, and while useful for many research studies are of little interest to the MCO.

In the early stages of the development of HPRF's data warehouse, known as the research data mart (RDM), it was realized that only a small subset of the fields in the data warehouse were needed. While each of the fields in the data warehouse was used by someone in the corporation only a small proportion of the fields were used consistently for research. While the RDM could contain fewer over all fields than the corporate data warehouse it needed to maintain them for as long as possible. There was also a need for the RDM to preserve those elements of structure that would allow the RDM to access data from the corporate data warehouse if unanticipated data were needed.

Having a well-designed and functional RDM has helped HPRF deal with the problems of multiple data sources and a system not optimized for research needs. However it did not helped to deal with the issues that arise out of the structure of health care delivery within managed care. Some of these issues as described below have the potential to present MCO research with major problems.

The first relates to the potential gap in communication when data elements change within the MCO's systems. Since the research component of the MCO is not involved in the main functions of the MCO i.e. tracking membership, submitting bills and paying claims, analysts working within the research area are frequently unaware of changes that may have occurred within the structure or content of a data set. The analyst's first indication that changes have occurred is often when something is amiss in the data. If the change is subtle such as an altered method of calculating a variable it may even be missed entirely. Therefore it has been important to develop and maintain channels of communication between HPRF and HP programs and system personnel.

Secondly, even though there are thousands of diagnosis and procedure codes potentially available to health care providers, the reality is that choices are usually checked off on a coding form. Since it would be impossible to list the thousands of ICD diagnoses and CPT procedure codes on both sides of a sheet of paper the most common diagnoses or procedures seen by a particular specialty are chosen. An example of how this is a limitation can be seen in the available codes for non-insulin dependent diabetes (type 2) on the primary care coding forms at HP. The only choice available for any case of type 2 diabetes is 250.00 the code for uncomplicated diabetes. If a patient has renal complications of diabetes the appropriate code would be 250.4 but this is not an option on the form. The consequence is that all type 2 diabetes primary care encounters, complicated or not, are coded with a code indicating it was uncomplicated diabetes.

A third potential set of problems arises out of the consolidation the health care industry is experiencing. The health care industry for the last decade and into the foreseeable future has been undergoing a state of continual change as it tries to adjust to market and competitive forces. One these changes is the merger of health care systems. The merger of health care systems also means the merger of data systems and the merger of data systems means problems.

Data system mergers as a result of corporate mergers can result in exasperating problems for research. Often the merged corporation continues to run the inherited claims processing systems for some extended period of time. HP was created from the merger of Group Health and MedCenters in the early nineties. However, because of prior contracts the two completely different claims processing system from these two prior entities were operated until just recently. There were situations where members could appear in different systems in differing years or encounter records could be duplicated in both systems. This required the research analyst to become familiar with two systems and to develop very different paradigms for creating data sets. Such differing paradigms also hold the potential for introducing unexplained variance. Any research that has as part of its goal the comparison of the two health care components of the merged health care system may find significant differences resulting from data system incompatibilities and not true health care system differences.

Often information that is in the paper medical record, while useful in managing patients' health care and important as a research variable is not present within the data systems. This is because it has no utility for the operational needs of the managed care system. Several examples of this within HP are blood pressure, weight and height. The effect of having data present within the chart but not in the automated systems is that any study that requires this information must do a chart

audit. Chart audits are very expensive relative to data extraction from electronic records. They are expensive because they can take considerable time and charts are usually housed at the member's primary clinic, requiring auditors to travel. When chart audits are necessary project budgets have to be larger. This can result in a tradeoff of fewer charts in order to stay within budget. The consequence being that sample sizes are often smaller and the power to detect significance suffers.

## *Data Systems Interact with Evolution of Foundation*

There is a concept in biology of co-evolution. This is understood to mean that as the environment changes the organisms that inhabit that environment must also change while at the same time contributing to the cycle of change. Such has been the systems process at HPRF.

When HPRF was first formed each project involved a unique approach to data collection. Systems personnel were recruited from the HP systems department. This was a logical approach as in the early days each project was unique and since data was sourced directly from the transaction systems personnel were needed who understood these systems. Statistical support was provided by outside contractors.

However, as time went by it became clear that there were common elements that spanned the majority of research projects. Systems were developed to automate and routinize some of these processes. The result being that project budgets could be reduced and timelines compressed. When new staff were hired emphasis was given to applicants that besides programming experience had data manipulation and simple statistical skills.

With the advent of the HP data warehouse there was another paradigm shift in systems and how they were used in support of research. Data no longer needed to be extracted from multiple independent systems. However, the programmers/analysts learned SQL and how to interface with the data through Microsoft Access or SAS. With more efficient access the data extraction component of projects was reduced in overall proportion of the work. This resulted in the need for greater skill in data manipulation and statistical analysis among programming staff. In fact at this point the programmers became data analysts and began to specialize. Some moved more towards statistical support while others did more data extraction and manipulation (Fig. 2).

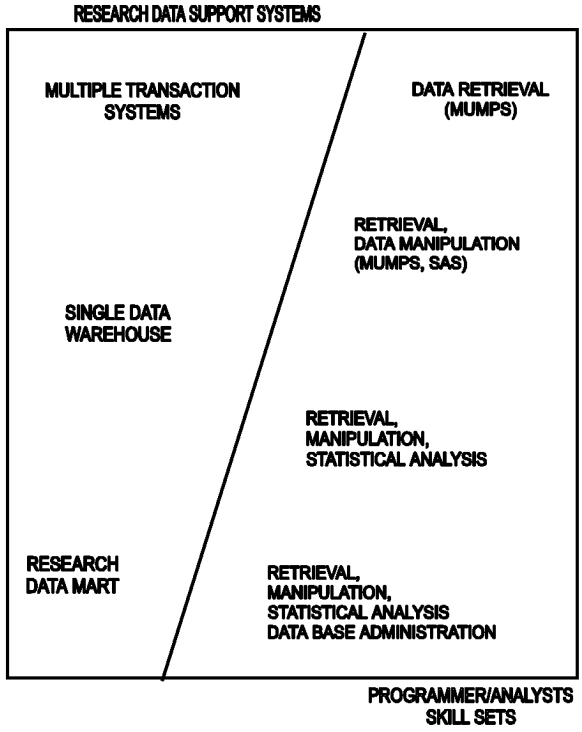


***Fig. 2.*** *The co-evoluation of research systems and programmer/analyst roles.*

The advent of the RDM has resulted in the current situation. While the same skills are needed to extract and analyze data from the RDM as the corporate data warehouse, HPRF has had to develop data base administrators (DBA) to design, deploy and update the RDM. This has resulted in further specialization into this area among the analysts.

If we return to our earlier biology example, we can see that when HPRF was at the beginning of its evolution systems were many and diverse. Personnel were generalists able to cope with the variation in multiple systems but lacking extensive analysis and statistical skills. As the systems environment changed towards a more organized and interrelated systems the programmers were freed to deal more with the data manipulation and analysis components of their work. Finally with the integration of data systems more specifically designed to support research, further diversity was developed through specialization of the analysts. It should also be noted that programmer/analyst staff went from one person at the beginning to a dozen at the current time.

## References


Briesacher B, Erwin W. Managed care organizations: The new research Partners. *Med Interface* 1995;8(12):68–73, 77.

Center for Health Research. *Health Research Databases*. Portland, OR: Kaiser Permanente, 1997.

Charlson M, Pompei P, Ales K, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chron Dis* 1987;40:373–383.

Chen R, DeStefano F, Davis R, et al. The vaccine safety datalink: Immunization research in health maintenance organizations in the USA. *Bull World Health Organ* 2000;78(2):186–194.

Cutler C. Research needs for managed care. *Health Aff (Millwood)* 1996;15(3):93–94.

Deyo R, Cherkin D, Ciol M. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613–619.

D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bses. *J Clin Epidemiol* 1996;49:1429–1433.

Eriksson E. Continuity-of-care measures. *Med Care* 1990;28:180–190.

Lapham S, Montgomery K, Hoy W. HMO databases: Fertile ground for epidemiological research. *Comput Healthc* 1990;11(9):18–20, 22, 24.

Moy E, Mazzaschi A, Levin R, Blake D, Griner P. Relationship between National Institutes of Health Research Awards to US Medical Schools and Managed Care Market Penetration. *JAMA* 1997;278:217–221.

Nelson A, Quiter E, LI S. The state of research within managed care plans: 1997 Survey. *Health Aff (Millwood)* 1998;17(1):28–38.

Redelmeier D, Lustig A. Prognostic indices in clinical practice. *JAMA* 2001;285:3024–3025.

Selby J. Linking automated databases for research in managed care settings. *Ann Intern Med* 1997;127(8 Pt 2): 719–724.

Smedby B, Smedby O, Eriksson E, Mattsson LG, LA. Continuity of care. *Med Care* 1984;22:676–680.

Steiner J, Koepsell T, Fihn S, Inui T. A general method of compliance assessment using centralized pharmacy records. *Med Care* 1988;26:814–823.

Teach E. Tapping your hidden assets. *CFO* 1996;47–56.

Verma R, Harper J. Life cycle of a data warehousing project in healthcare. *J Healthc Inf Manag* 2001;15(2):107–117.

Walter L, Brand R, Counsell S, et al. Development and validation of a prognostic index for a 1-year mortality in older adults after hospitalization. *JAMA* 2001;285:2987–2994.



*William Rush* is a research associate at HealthPartners Research Foundation. For six years he lead the programming and analysis group before moving into an investigator role. His current research is centered within two areas. The first is diabetes health services research. The second focuses on the emerging area of dental health services research.

*Patrick O'Connor* is a Senior Clinical Investigator at HealthPartners Research Foundation. He is a practicing physician in family medicine with a research agenda in chronic disease outcomes research. He has focused his research on developing and assessing interventions to improve morbidity, mortality, function, satisfaction, and costs of care for adults with diabetes, hypertension, lipid disorders, or heart disease.

*Paul Johnson* is Carlson Professor of Decision Sciences and Adjunct Professor of Psychology and Computer Science at the University of Minnesota. He is also a faculty member in the University of Minnesota Center for Cognitive Sciences and the Center for Political Psychology. Professor Johnson's research focuses on the study of variability and expertise in decision-making, the decision failures that arise from the misperception or misrepresentation of information (deception) and the use of knowledge resources and decision technologies in business and professional settings.