REVIEW

MAYO CLINIC

History of the Rochester Epidemiology Project: Half a Century of Medical Records Linkage in a US Population

Walter A. Rocca, MD, MPH; Barbara P. Yawn, MD, MSc; Jennifer L. St. Sauver, PhD, MPH; Brandon R. Grossardt, MS; and L. Joseph Melton III, MD, MPH

Abstract

The Rochester Epidemiology Project (REP) has maintained a comprehensive medical records linkage system for nearly half a century for almost all persons residing in Olmsted County, Minnesota. Herein, we provide a brief history of the REP before and after 1966, the year in which the REP was officially established. The key protagonists before 1966 were Henry Plummer, Mabel Root, and Joseph Berkson, who developed a medical records linkage system at Mayo Clinic. In 1966, Leonard Kurland established collaborative agreements with other local health care providers (hospitals, physician groups, and clinics [primarily Olmsted Medical Center]) to develop a medical records linkage system that covered the entire population of Olmsted County, and he obtained funding from the National Institutes of Health to support the new system. In 1997, L. Joseph Melton III addressed emerging concerns about the confidentiality of medical record information by introducing a broad patient research authorization as per Minnesota state law. We describe how the key protagonists of the REP have responded to challenges posed by evolving medical knowledge, information technology, and public expectation and policy. In addition, we provide a general description of the system; discuss issues of data quality, reliability, and validity; describe the research team structure; provide information about funding; and compare the REP with other medical information systems. The REP can serve as a model for the development of similar research infrastructures in the United States and worldwide.

© 2012 Mayo Foundation for Medical Education and Research ■ Mayo Clin Proc. 2012;87(12):1202-1213

From the Division of Epidemiology, Department of Health Sciences Research (W.A.R., B.P.Y., J.L.S., L.J.M.), Division of Biomedical Statistics and Informatics, Department of Health Sciences Research (B.R.G.), and Department of Neurology (W.A.R.), Mayo Clinic, Rochester, MN; and Department of Research, Olmsted Medical Center, Rochester, MN (B.P.Y.).

The ability to link patient-specific health information across diverse health care providers (hospitals, physician groups, and clinics) is increasingly recognized as key to improving the quality and efficiency of medical care. Such a medical records linkage system can be used to improve continuity of care (1) by making remote medical events and treatments available to the care physician at the time of a medical visit without relying on the patient's recollection and (2) by avoiding the need to repeat diagnostic tests. The linkage of medical records also provides an ability to access patient outcomes, including the effectiveness of treatment. When the system is applied to a sufficiently large group of patients over an extended time, medical records linkage systems also support more general research into disease trends in the community. Such systems become particularly informative when they cover all the residents in a well-defined population. Population-based research is a major source of evidence to support medical and public health practices.[1-5]

For almost half a century, the Rochester Epidemiology Project (REP) has maintained a comprehensive medical records linkage system for persons residing in Olmsted County, Minnesota, to support clinical and epidemiologic research (http://www.RochesterProject.org). This unique research infrastructure resulted from a series of particular circumstances and from the contributions of several protagonists from within and external to Mayo Clinic, the founding and hosting institution. Details about the methods currently used in the REP to link medical records from multiple health care providers to specific individuals, to track residency status over time, and to compile a population census are provided elsewhere.[6,7] Details about the demographic, ethnic, and socioeconomic characteristics of the Olmsted County population are also reported elsewhere.[1]

In this article, we provide a brief history of the REP divided into 2 broad periods, before and after 1966, the year in which the REP was established and first funded by the National Institutes of Health (NIH). The "History Before 1966" section describes the circumstances that led to the creation of a medical records linkage system at Mayo Clinic. The "History After 1966" section then describes the circumstances that led to the creation of a medical records linkage system across all health care providers serving the population of Olmsted County. Table 1 provides an outline of critical dates, key protagonists, major events, and technical advances over more than a century. Table 2 provides a summary of the guiding principles that have made the REP possible.

TABLE 1. Timeline of the Rochester Epidemiology Project (REP): Key Protagonists, Major Events, and Technical Advances[a]

Critical dates	Key protagonists	Major events and technical advances
		Before 1966
1885	W. W. Mayo, W. J. Mayo, C. H. Mayo	Initiated a partnership of father and 2 sons that developed into the present-day Mayo Clinic
1885-1907	Initially, Mayo partnership, later a group practice	Used leather-bound ledgers. Data collected in chronological order and separately by each physician
1905	C. H. Mayo	Published the first clinical series of cases from Mayo Clinic[8]
1907	H. S. Plummer	Introduced the unit medical record to assemble all pages pertaining to the same patient (dossier). Introduced a patient registration number (later called the Mayo Clinic number)
1910-1930	M. Root	Introduced index cards to find patients with a specific diagnosis or surgery (2 index systems)
1935	J. Berkson	Introduced 2 new indexes: 1 for diagnoses and 1 for surgical procedures. Introduced Berkson classification codes. Introduced Hollerith punch cards for mechanical data processing
1949	—	Opening of Olmsted Medical Center as a multispecialty clinic
1950	A. R. MacLean et al	Published population-based incidence rates of multiple sclerosis in Rochester, MN[9]
		After 1966
1966	L. T. Kurland	Created the medical records linkage system. Obtained the first supporting grant from the National Institutes of Health. Established Olmsted County as the epidemiologic population. Initiated consortium collaboration of all health care providers in Olmsted County with Mayo Clinic
1981	L. T. Kurland and C. A. Molgaard	Published a *Scientific American* article on the REP[10]
1991	B. P. Yawn	Opened a Department of Research at Olmsted Medical Center focusing on primary care research using the REP
1996	L. J. Melton III	Published the first article on the history of the REP[6]
1997	L. J. Melton III	Minnesota state privacy law required each patient to sign an authorization to review medical records for research (statute 144.335). Organized massive mailings and contacts to obtain the authorizations
2002	S. J. Jacobsen	Introduced an electronic portal to the REP (the REP Browser). Introduced the first intramural REP website
2006-2012	W. A. Rocca and B. P. Yawn	Initiated joint leadership of the REP by Mayo Clinic and Olmsted Medical Center co-principal investigators. Introduced the first extramural REP website[b] Developed the REP Census enumeration and personal timelines.[7] Added a drug prescription index and other indexes to the system. Initiated community engagement activities. Addressed the generalizability of findings[1]
2011-2012	J. St. Sauver et al	Published 2 articles describing methodological aspects of the REP[1,7]
Immediate future	—	Expand the REP to the remaining health care providers in Olmsted County. Expand the REP to an 8-county region of southeastern Minnesota. Add new indexes and computerized databases

[a] REP = Rochester Epidemiology Project.

[b] Extramural REP website: http://www.RochesterProject.org.

In a third section, "The REP Today," we provide a general description of the system; discuss issues of data quality, reliability, and validity; describe the research team structure; provide information about funding; and compare the REP with other medical information systems. The REP can serve as a model for the development of similar research infrastructures in the United States and worldwide.

HISTORY BEFORE 1966

Early Ledgers of the Mayo Clinic Practice

Establishment of the REP in 1966 was made possible by a chain of events and technical developments that took place at Mayo Clinic, the founding institution, during the first half of the 1900s. Details about the early history of Mayo Clinic and

TABLE 2. Guiding Principles That Make the REP Possible

1. Clinical and epidemiologic research based on medical records is indispensible for the progress of surgical and medical practice
2. Medical records can be used to provide continuity of care to the individual patient but also for research and education. Medical records should be archived and preserved
3. The medical data collected by individual physicians (or health care professionals) during routine medical care should be shared for the good of practice, research, and education (sharing within institutions)
4. Sharing of data collected by different institutions is indispensible to provide a complete picture of health and disease in a geographically defined population (sharing across institutions; population-based research)
5. Clinical and epidemiologic research based on routinely collected and linked medical data is essential to improve the health of the community being studied and to guide medical and public health decisions at the national level
6. There is a complex risk-benefit balance between the desire to maintain confidentiality of medical data and the need to conduct clinical and epidemiologic research to improve health

about the development of a records linkage system at Mayo Clinic have been reported more extensively elsewhere.[10,11] Herein, we provide a brief synopsis of the key events and protagonists narrated from today's perspective.

Mayo Clinic was founded in the late 1800s as a family partnership by William W. Mayo and his 2 sons, William J. and Charles H. (Figure 1, A). In 1889, they helped the Sisters of St. Francis of Assisi open Saint Marys Hospital, which was then the only hospital in the region. Around 1903, the practice was transformed into a group practice.[10,11] While the group practice remained small and surgical procedure oriented, an infor-

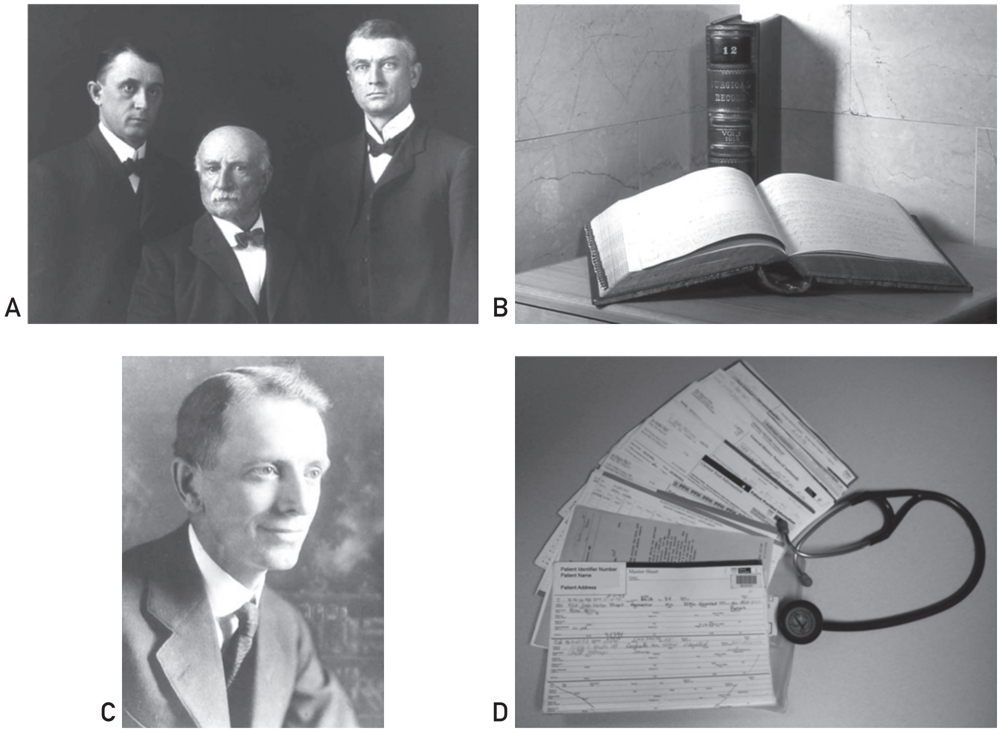

FIGURE 1. A, The father, W. W. Mayo (1819-1911, center), and the 2 Mayo brothers, W. J. Mayo (1861-1939, right) and C. H. Mayo (1865-1939, left). B, Leather-bound ledgers used before 1907. C, H. S. Plummer (1874-1936). D, The new medical record introduced by Dr Plummer in 1907 to assemble forms for each patient into a dossier (paper file). Photographs courtesy of the Mayo Historical Unit, Mayo Clinic, Rochester, Minnesota.

mal system of record keeping was adequate to ensure continuity of care for returning patients. Thus, from 1885 until 1907, patient records were kept by individual physicians in leather-bound ledgers (Figure 1, B). Case histories were generally brief (4 or 5 to a page) and were usually entered in chronological order, similar to a diary. From the very beginning, it was clear to Mayo Clinic physicians that progress in surgical or medical practice could be obtained only through research on the outcomes of their medical and surgical interventions (Table 2).[11] Thus, the ledger system became the basis for the continuity of care of individual patients and for clinical research. Although Mayo Clinic physicians used the data archived in the ledgers to describe series of surgical cases and to report the results of new surgical techniques,[10,12] it was cumbersome to trace the history of patients through multiple ledgers archived in different locations and owned by individual physicians.

Introduction of the Unit Medical Record

Henry Plummer, a young clinical associate of the group practice (Figure 1, C), addressed both major problems of the ledgers (the fragmentation of information and the individual ownership).[12] By 1907, Plummer introduced a system whereby medical information was written on unbound paper forms (loose pages) kept in a single file, or dossier, for each patient (Figure 1, D). Each patient was assigned a unique registration number that was repeated on each page of the various forms included in the dossier (the Mayo Clinic number). This numbering system is still used today and encompasses millions of unique patients. The dossier included notes made by each physician who examined the patient. All the notes from each medical specialty were kept together, and new forms were introduced as new specialties were developed at Mayo Clinic. The results of laboratory tests were transcribed in chronological order on separate forms that were designed to allow a rapid visual scan of current and historical results. Correspondence with the patient was also kept in the file, as were birth and death records for local residents (Figure 1, D). Plummer's system ensured that all the medical information pertaining to an individual patient could be found conveniently in a single dossier of documents archived in a central location. In addition, Plummer persuaded the other members of the group practice to establish, as policy, that the dossiers should serve as an institutional resource so that all the records would be available for teaching and research to all the members, regardless of which physician had treated a particular patient (Table 2).[10]

However, it became evident that the use of the dossiers for education and research required an additional set of tools (Table 2).[10] To identify groups of patients with the same disease or the same surgical procedure, Plummer created 2 simple indexes: 1 organized by diagnosis and 1 by surgical procedure. The identification number (the Mayo Clinic number), age, sex, and date of medical contact for each patient with a given disease or surgical procedure were written by Plummer's secretary, Mabel Root, on 5×8-inch index cards (Figure 2, A and B). In both indexes, the major headings were organs and organ systems (eg, thyroid); listed alphabetically under these headings were specific diseases or surgical procedures (eg, hypothyroidism) (Figure 2, B). The indexes made it possible to locate records of patients who had similar diseases or who had undergone similar surgical procedures and to assemble clinical or surgical series for education or research.[12]

The major drawback of the Plummer-Root system was that diseases were listed under organs. Diseases that can affect many organs (such as cancer) were dispersed throughout the index. In addition, diseases were not grouped into meaningful functional categories (eg, cardiovascular diseases).[10] At the same time, diagnostic terminology was expanding rapidly, as shown by the publication of a *Standard Nomenclature of Diseases and Operations* in 1933 (2nd edition in 1935) by the American Medical Association.[14] Ultimately, the increased diagnostic sophistication and the growth in patient volume overburdened the Plummer-Root system as they had overburdened the ledger system 25 years earlier.

Advances in Medical Record Indexing

In the 1930s, Joseph Berkson of the Mayo Clinic Department of Physiology was asked to undertake a second reorganization of the medical record system (Figure 2, C).[15] In 1935, Berkson developed 2 new indexes, 1 for surgical procedures and 1 for diagnoses, based on disease codes. However, Berkson decided to use neither the standard nomenclature of the American Medical Association[14] nor the *International Classification of Causes of Death* (later called the *International Classification of Diseases [ICD]* and currently available in its 10th revision; http://www.who.int/whosis/icd10/).[5] In place of these published nomenclatures, Berkson devised his own diagnostic codes, although they did share several features with the published classification systems.[16] Similar to the standard nomenclature of the American Medical Association,[14] Berkson's codes had headings for organs and diseases, reflecting increased interest in disease processes. In each disease category, he created a classification labeled "except as above," much like the "other" and "unspecified" categories of the *ICD*. Berkson's codes covered more than 20,000 diseases and sites in the body, easily accommodating the level of diagnostic specificity at

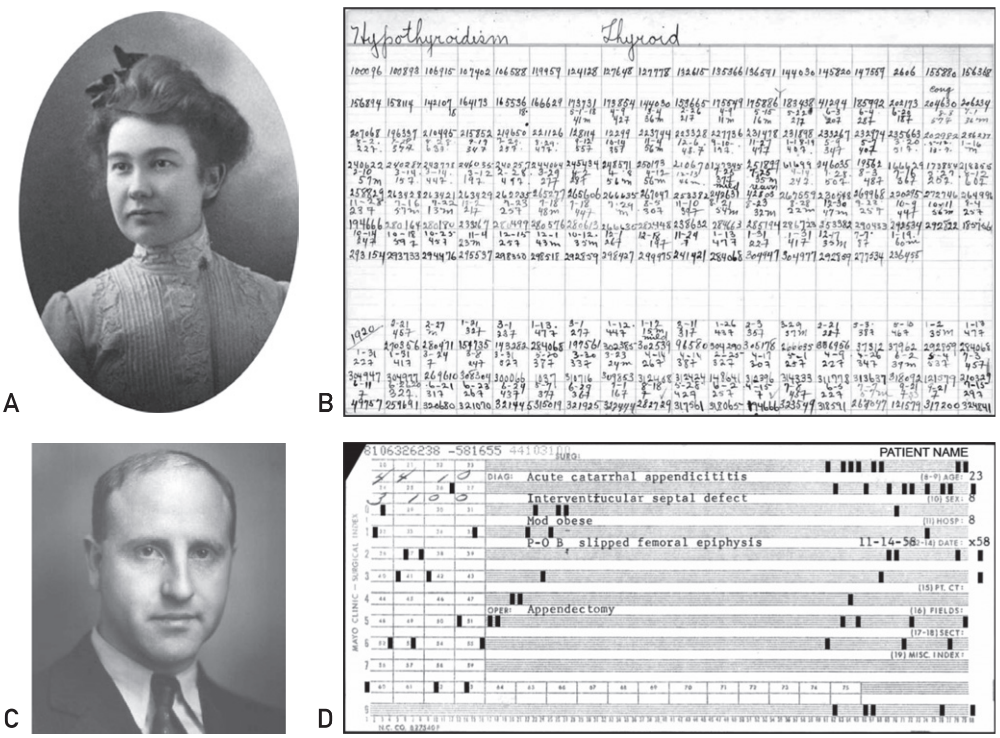

FIGURE 2. A, Mabel Root (1878-1981), secretary to Dr Plummer. B, Index card listing patients seen at Mayo Clinic for hypothyroidism (organ = thyroid; disease = hypothyroidism). C, J. Berkson (1899-1982). D, A Hollerith punch card used for mechanical data processing. Photographs courtesy of the Mayo Historical Unit, Mayo Clinic, Rochester, Minnesota.

that time. The numerical codes in Berkson's indexes became known locally as the "Berkson codes" and were entered on Hollerith punch cards, which were then the most efficient medium for mechanical data processing (Figure 2, D).

It is difficult in retrospect to determine whether the decision of Berkson to develop an independent coding system was a strategic decision or an error. On the one hand, creating a local coding system allowed more flexibility and reflected more directly local needs and practices. On the other hand, the use of a customized classification system made the clinical studies published by investigators from Mayo Clinic more difficult to compare with studies conducted elsewhere that used consensus-based classification systems.

A second original feature of Berkson's diagnostic index was the inclusion on the punch cards of an item indicating residency in Rochester, the central city in Olmsted County. Because almost all care for major diseases in the area was being provided by Mayo Clinic at that time, it became possible to identify quickly and by mechanical means all local patients who had a given disorder. This geographic code became important when the need to focus epidemiologic research on a well-defined population became clear. After World War II, epidemiology developed rapidly as a major research method in cancer, cardiovascular disease, and other chronic diseases,[2,17,18] and investigators recognized the need for population-based research to validate the scientific evidence previously based only on series of patients seen in hospitals or specialized centers. The risk of selection bias when describing the clinical spectrum of severity and the outcomes of diseases was finally recognized and addressed.[19-21] In 1946, Joseph Berkson described the selection bias that may occur in case-control studies based on hospital admissions (known as Berkson bias).[22,23]

HISTORY AFTER 1966

Creation of a Medical Records Linkage System for Olmsted County

Stimulated by the publication of a study on the incidence of multiple sclerosis in Olmsted County,[9] Leonard T. Kurland came to Mayo Clinic for a fel-

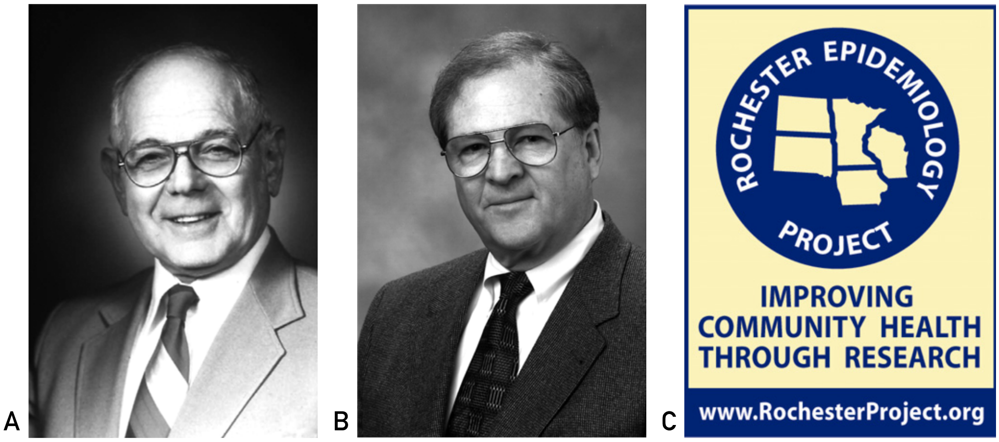
A
B
C
ROCHESTER EPIDEMIOLOGY PROJECT
IMPROVING COMMUNITY HEALTH THROUGH RESEARCH
www.RochesterProject.org

FIGURE 3. A, L. T. Kurland (1921-2001), founder of the Rochester Epidemiology Project. B, L. J. Melton III (1944-). C, Current logo of the Rochester Epidemiology Project. Photographs courtesy of the Mayo Historical Unit, Mayo Clinic, Rochester, Minnesota.

lowship in neurology and returned a decade later to become head of what is now called the Department of Health Sciences Research (Figure 3, A).[24] Kurland was trained in neurology and epidemiology, and he was the first to more fully exploit the unique potential of the Mayo Clinic records archive for generating accurate frequency and natural history data from a geographically defined population. In 1966, Kurland obtained funding from the NIH to create indexes of diagnostic codes for the other (non–Mayo Clinic) health care facilities providing care to residents of Rochester and Olmsted County and to link their medical records with those already present at Mayo Clinic. This was the official beginning of the REP medical records linkage system.[3-5] The original 1966 federal grant also helped provide appropriate statistical support for the system and allowed for the hiring of a population geneticist. The major new partner was Olmsted Medical Group (now known as Olmsted Medical Center), which was established in 1949 as a primary care, multispecialty clinic.[25] The result was linkage of medical data from almost all the sources of medical care used by the local population.

The coding system introduced by Berkson is still in use today to retrieve diagnoses or surgical procedures dating from 1935 to 1975 for research projects.[10,16] However, by the 1970s, the catch-all category “except as above” became obsolete because of the proliferation of new diagnoses, and the searches for new diseases became time-consuming and, therefore, expensive. To solve this problem, Kurland introduced in 1975 a new disease coding system that was based on the 8th revision of the *ICD*, hospital adaptation.[26] This decision not only resolved a practical issue by reducing the time and cost of identifying patients with a given disease but also realigned the research performed at Mayo Clinic with research performed elsewhere. By adopting an international classification system, the investigators using the REP were able to better communicate their methods and to compare their results with those from other institutions. The REP has continued in this tradition with the more recent implementation of the *ICD-9* and the current preparations for the clinical introduction of the *ICD-10*.

Response to Confidentiality Concerns

Continuing the work of Kurland, L. Joseph Melton III was the principal investigator of the REP from 1991 to 2000 (Figure 3, B), and in 1996 he wrote the first article on the history of the REP.[6] During this time, federal and state law allowed the use of medical record data for epidemiologic studies without written consent or authorization. However, in January 1996, the state of Minnesota passed a new law requiring a general written authorization from each patient before their medical records could be reviewed for research. The law was amended in 1997 to clarify its implementation (Minnesota state privacy law, Statute 144.335).[27-29] The authorization applied to all residents seen after that date and did not expire but could be revoked at any time.

Mayo Clinic, Olmsted Medical Center, Rochester Family Medicine Clinic, and other health care facilities affiliated with the REP established procedures to comply with the new law. Two contacts to obtain research authorization from each participant are attempted in writing (mailing) or in person with at least 60 days between attempts and the contacts include language indicating that lack of response will be taken as implied approval for medical record review (as per Minnesota law). If the patient gives explicit authorization or does not respond after these 2 attempts, then the record is considered accessible for research purposes. Authorization for research use of existing medical record data is also implied for patients never seen after January 1, 1997.

In 2 studies of the impact of research authorization on medical record research, 97% of 2463 unique individuals seen between 1994 and 1996 provided authorization at Mayo Clinic (stratified random sample from 309,930 patients referred from any region) and 96% of all 15,997 patients seen in January or February 1997 provided research authorization at Olmsted Medical Center.[28,29] In the Mayo Clinic study, refusal of research authorization was somewhat higher for women, younger patients, patients living in the local community, and patients with previous diagnoses considered more sensitive (eg, mental disorders).[29] When restricting the sample to residents of Olmsted County between 1998 and 2007, 90.7% of patients gave authorization to all the health care providers included in the REP, an additional 7.2% gave authorization to at least 1 health care provider, and only 2.1% denied authorization to all REP health care providers.[7]

A new level of regulation was introduced in 2002 by the Health Insurance Portability and Accountability Act (HIPAA).[30] As allowed by the law, REP studies that involve only review of existing medical record data (passive medical record review) can be conducted without obtaining study-specific written informed consent from each participant if the investigators obtain a HIPAA waiver from the Mayo Clinic Institutional Review Board (IRB) and the Olmsted Medical Center IRB. The waiver is provided because obtaining written consent for each specific study would be almost impossible to implement (the study may span decades and include thousands of patients, many of whom may have died) and would pose a major burden on some patients (repeated mail contacts by different investigators). This practice is consistent with the recommendation by the Council for International Organizations of Medical Sciences that formal informed consent requirements should be waived for medical record studies based on large historical cohorts, such as the REP.[5,31-33] However, if any contact with the individual is made as part of the study (via letter, telephone, or face-to-face), written informed consent and HIPAA authorization are required.

By using the general Minnesota research authorization and the HIPAA waiver, investigators using the REP have been allowed to conduct studies with high participation rates.[7,33] The experience with participation in passive research in the REP may serve as a model for future attempts to create medical records linkage systems in other US populations that are at the same time efficient while respecting patients' privacy concerns.[33,34]

Introduction of Electronic Medical Records and Recent Developments

Steven Jacobsen directed the REP from 2000 to 2006. During these years, the traditional paper medical records contained in the dossiers were progressively replaced by electronic medical records. The transformation was complete by 2004 at Olmsted Medical Center and by 2005 at Mayo Clinic. Jacobsen's major contribution was to exploit the new electronic capabilities to create an electronic portal to search medical records for a given individual. This portal, called the REP Browser, allows electronic access to an index of all medical records for each individual and of their archival location (paper record or electronic record) for IRB-approved studies. Starting in 2002, Jacobsen also introduced a more formal and permanent linkage of medical records to single individuals. The linkage had historically been created on a study-by-study basis using probability scores, and the investigators had to clarify matches with low scores. This study-by-study linkage was expensive, time consuming, and susceptible to errors.[7] The permanent linkage of records initiated by Jacobsen was finalized in more recent years and was supplemented with manual verification of uncertain matches. Details about the linkage methods used currently in the REP are reported elsewhere.[7] Dr Jacobsen also introduced the first internal website to assist users in the design and conduct of REP studies.

In 2006, Walter Rocca from Mayo Clinic and Barbara Yawn from Olmsted Medical Center became joint directors of the REP. This joint leadership recognizes that the REP is a consortium of multiple institutions. Rocca and Yawn further developed the REP infrastructure in 7 new directions. First, they developed a plan to include in the REP the few health care providers in Olmsted County currently not participating in the REP (eg, dental practices, optometrists, and chiropractors). Several of these health care providers have now been included in the consortium. Second, they developed a plan to expand the coverage of the REP to include the 8-county region of southeastern Minnesota (Dodge, Goodhue, Wabasha, Winona, Houston, Fillmore, Mower, and Olmsted counties). The expansion is under way and will increase the REP population from approximately 140,000 persons to approximately 350,000 persons (counting only current residents). Third, they are introducing new electronic indexes for drug prescriptions, medical services, costs of services, and immunizations. In particular, they recently completed the integration of drug prescription data from Olmsted Medical Center and Mayo Clinic. Drug prescriptions were linked and coded using a common nomenclature (RxNorm of the National Library of Medicine; http://www.nlm.nih.gov/research/umls/rxnorm/) and were grouped using the National Drug File–Reference Terminology (http://www.usgovxml.com/dataservice.aspx?ds=NDFRT). This new component of the medical records linkage system will facilitate pharmacoepidemiologic studies (Table 1).

Fourth, in 2011 Rocca and Yawn started a program of community engagement to increase awareness in the Olmsted County community of the REP and to create a partnership between the REP leadership and the community. As a result of these activities, a REP Community Advisory Board was established in September of 2012. Fifth, in 2009, they introduced the first extramural website to facilitate use of the REP research infrastructure by external investigators (http://www.RochesterProject.org). Sixth, they developed a REP Census enumeration and defined a personal timeline for each individual who has resided in Olmsted County since 1966.[7] Finally, they formally addressed the demographic and socioeconomic similarities and differences of the Olmsted County community compared with the state of Minnesota, the Upper Midwest, and the entire United States. These comparisons are important to establish the generalizability of studies conducted using the REP to other populations or to the entire country.[1]

THE REP TODAY

General Description

The medical records linkage system of the REP now encompasses 6,239,353 person-years of follow-up for a total of 502,820 unique individuals attended at least once between 1966 and 2010 (counting both current and previous residents). The REP does not collect and store data following a specified format as typically used in a cohort study (eg, the Framingham Heart Study) or in cross-sectional surveys (eg, the National Health and Nutrition Examination Survey).[35,36] Medical information is recorded as part of routine medical practice using the format dictated by each specific health care provider. Only demographic data, diagnostic codes, surgical procedure codes, and drug prescriptions are currently organized in electronic indexes that can be searched by computer.

Figure 4 summarizes the steps involved in creating the records linkage system and in using it to test clinical or epidemiologic hypotheses. Multiple medical records for the same individual are linked within and across institutions to create a compre-

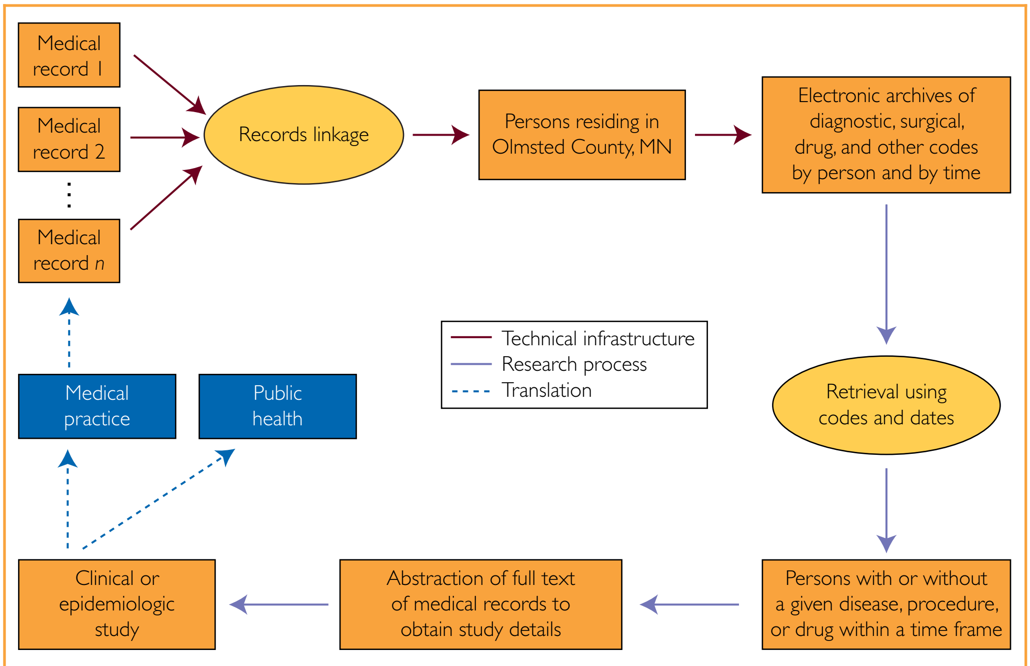
Medical record 1
Medical record 2
⋮
Medical record *n*
Records linkage
Persons residing in Olmsted County, MN
Electronic archives of diagnostic, surgical, drug, and other codes by person and by time
Technical infrastructure
Research process
Translation
Retrieval using codes and dates
Medical practice
Public health
Clinical or epidemiologic study
Abstraction of full text of medical records to obtain study details
Persons with or without a given disease, procedure, or drug within a time frame

FIGURE 4. Scheme of the steps involved in linking medical records, indexing diagnoses and other data, and retrieving medical records for a research study.

hensive medical dossier. Diagnostic codes, surgical codes, and other coded information are abstracted and stored electronically in indexes that can be searched using a computer. Each piece of information also includes a time frame (eg, the date of a surgical intervention or the date of diagnosis). Investigators who use the REP retrieve lists of patients who received a specific diagnosis or procedure within a specified time frame from the computerized indexes (eg, all women who underwent hysterectomy between 1975 and 1995 or all patients who received ≥1 diagnosis of myocardial infarction between 2000 and 2005). Next, a nurse abstractor, physician, or other trained investigator reviews all the records for these patients to verify the diagnosis and apply specific diagnostic criteria. Detailed record abstraction is also used to collect data on exposures (eg, occupation, smoking, and marital status) and outcomes of interest (eg, nursing home admission and disability). Finally, the information obtained is used to design incidence or prevalence studies, case-control studies, cohort studies, cost or cost-effectiveness studies, and natural history or outcome studies.

The REP has been used in studies spanning almost every medical specialty and has yielded more than 2000 publications to date. A guiding principle at the time of the REP inception, as well as today, is that population-based research can lead to improved community health (Table 2 and Figure 4). This idea is represented in the current logo (Figure 3, C). Other guiding principles that have evolved over half a century of experience with medical records linkage are provided in Table 2.

Data Quality, Reliability, and Validity

Historically, the REP has included Mayo Clinic and its affiliated hospitals, Olmsted Medical Center and its affiliated hospital, the University of Minnesota hospitals, and the Veterans Affairs Medical Center, located in Minneapolis, Minnesota, as well as other medical institutions in the region. Seven private general practitioners have had offices in Rochester in the past 46 years, and data from all of their practices have also been incorporated into the REP. Currently, only 1 general practitioner is active in the community (the Rochester Family Medicine Clinic), and he participates in the REP. Two small charity clinics, several dental practices, optometrists, chiropractors, and some vaccination facilities are not included in the REP, although efforts are under way to engage these remaining health care providers in the REP consortium. As a result, information about dental and ophthalmologic diseases currently remains incomplete.

Over the years, a variety of studies on the validity of the linkage methods and the census enumeration have been completed, as reported in detail elsewhere.[7] For example, in a random sample of 400 patients, only 10 had at least 1 record incorrectly included, and 5 were missing at least 1 record. The rate of overinclusion in the matching was 2.5%, and the rate of underinclusion was 1.3%.[7] In addition, the REP Census was found to be valid compared with a list of residents obtained from random-digit dialing, a list of residents of nursing homes and senior citizen complexes, a commercial list of residents, and a manual review of records. Finally, the REP Census counts were comparable with those of 4 decennial US censuses.[7]

Reliability and validity studies of specific variables considered in studies using the REP have been reported by many authors over many years. For example, in a study of Parkinson disease, the information about cigarette smoking, occupation, and years of education obtained from record abstraction was compared with information obtained at interview.[37,38] Use of the REP by multiple investigators over almost half a century has generated a rich documentation of the quality of specific data about exposures or diagnoses. Finally, we reported elsewhere data suggesting that findings of studies using the REP are comparable with findings of studies using other comparable methods of case identification (eg, similar incidence or prevalence or similar time trends).[1]

Research Team Structure

The REP is currently run by a Scientific Steering Committee composed of the 2 co-principal investigators, an anthropologist, a bioethicist, a biostatistician, an expert in information technology, and a pharmacoepidemiologist. A full-time epidemiologist serves as scientific manager and coordinates the day-to-day activities pertaining to research projects, interactions with users of the REP, publications, and other scientific issues. Financial, personnel, and operational activities are coordinated by 2 project managers, 1 at Mayo Clinic and 1 at Olmsted Medical Center. A third part-time project manager assists with the development of collaborations with new care facilities not yet participating in the REP. The data management team includes 3 full-time information technology experts, and the statistical team includes 2 part-time statisticians and 1 part-time data analyst. The day-to-day activities of records handling (storing and retrieving), diagnostic coding, and data verification and correction are conducted by a team of 4 full-time study assistants. Other part-time team members assist with personnel supervision and with the management of data on cost of medical services and procedures. Two part-time administrative assistants provide secretarial and meeting organization support. The operational team meets once per week, and the full team meets once per month. Additional meetings involve specific work groups.

Funding

The REP has been funded continuously by the NIH for 47 years (since its inception in 1966). Several institutes have contributed to the funding over the years. In the most recent 10 years, the REP was funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, and it is currently supported by the National Institute on Aging. However, the federal funding has always been supplemented by funding from Mayo Clinic. To provide some idea of the cost of maintaining the current infrastructure, the REP budget for 2012 was approximately $770,000 from the National Institute on Aging (total direct costs) and $600,000 from Mayo Clinic, for an annual total budget of $1,370,000.

Comparison With Other Medical Information Systems

There is a long tradition of using medical records linkage techniques to create extensive infrastructures for epidemiologic research.[3-5] An ideal medical records linkage system should have 4 characteristics: (1) it should cover a well-defined geographic region, such as a city, county, state, or other geographic entity; (2) it should have existed for 10 years or more to provide historical depth (many important public health questions can be answered only using data with a long interval between exposures and outcomes); (3) it should include a large number of persons so that rare exposures or rare diseases and medical practices can be studied; and (4) it should include as many variables as possible that can be searched electronically (demographic data, diagnostic codes, drug prescriptions or sales, surgical procedures, diagnostic procedures, screening procedures, laboratory results, etc).

Among English-speaking countries, successful examples of medical records linkage systems have been implemented in the United Kingdom,[39-43] Australia,[44] and Canada.[45,46] However, similar systems have been more limited in the United States because of the lack of a national health system.[47] Some health maintenance organizations and other health plans have been able to develop a medical records linkage system for patients affiliated with the plan; however, often the patients covered by the plan do not represent the entirety of a geographically defined population. Two examples are the Kaiser Permanente plans in California and Oregon (http://www.kaiserpermanente.org) and Group Health of Washington State and North Idaho (http://www.ghc.org).

Only in recent years have attempts been made at the federal level in this country to create publicly accessible databases for research.[34] However, these efforts are struggling with equally strong trends toward strict confidentiality of medical record information.[48] Even if these national databases become available to investigators in the US, they may lack historical depth or may be limited to specific age groups (eg, Medicare generally covers only patients aged ≥65 years). In contrast, the REP system covers a complete population of approximately 500,000 persons of all ages residing in a well-defined geographic region (Olmsted County), has existed for almost half a century, and includes electronic indices for diagnostic codes, surgical procedures, and drug prescriptions. The addition of new indexes and computerized databases is ongoing. With the efforts initiated in 2010 toward a national health system, the need to link medical records within and across institutions to improve continuity of care and provide scientific evidence about the effectiveness and cost of medical interventions will increase.[34] The REP is an important model for guiding these developments.

CONCLUSION

The REP medical records linkage system has existed for almost half a century and was made possible by approximately another half century of local developments at Mayo Clinic, the founding institution. Because of this long and complex history, its coverage of an entire population, its geographic location, and its scientific productivity, the REP is unique in the United States. Studies supported by the REP have contributed to transforming medical practices in Olmsted County and worldwide and to improving public health at the community, national, and international levels.

ACKNOWLEDGMENTS

We thank Lori Klein for typing and formatting the manuscript and Mary G. Roberts for other support. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Abbreviations and Acronyms: **HIPAA** = Health Insurance Portability and Accountability Act; ***ICD*** = *International Classification of Diseases*; **IRB** = institutional review board; **NIH** = National Institutes of Health; **REP** = Rochester Epidemiology Project

Grant Support: This study was made possible by the Rochester Epidemiology Project (grant number R01-AG034676; Principal Investigators: Walter A. Rocca, MD, MPH, and Barbara P. Yawn, MD, MSc).

Correspondence: Address to Walter A. Rocca, MD, MPH, Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (rocca@mayo.edu).

REFERENCES

1. St Sauver JL, Grossardt BR, Leibson CL, et al. Generalizability of epidemiologic findings and public health decisions: an illustration from the Rochester Epidemiology Project. *Mayo Clin Proc*. 2012;87(2):151-160.
2. Susser M, Stein Z. *Eras in Epidemiology: The Evolution of Ideas*. New York, NY: Oxford University Press; 2009.
3. Newcombe HB. *Handbook of Record Linkage: Methods for Health and Statistical Studies, Administration, and Business*. New York, NY: Oxford University Press; 1988.
4. Dunn HL. Record linkage. *Am J Public Health Nations Health*. 1946;36(12):1412-1416.
5. Porta MS; International Epidemiological Association. *A Dictionary of Epidemiology*. 5th ed. New York, NY: Oxford University Press; 2008.
6. Melton LJ III. History of the Rochester Epidemiology Project. *Mayo Clin Proc*. 1996;71(3):266-274.
7. St Sauver JL, Grossardt BR, Yawn BP, et al. Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester Epidemiology Project. *Am J Epidemiol*. 2011;173(9):1059-1068.
8. Mayo CH. Mortality, disability and permanency of cure in surgery. *Northwestern Lancet*. 1905;25:179-182.
9. MacLean AR, Berkson J, Woltman HW, et al. Multiple sclerosis in a rural community. *Res Publ Assoc Res Nerv Ment Dis*. 1950;28:25-27.
10. Kurland LT, Molgaard CA. The patient record in epidemiology. *Sci Am*. 1981;245(4):54-63.
11. Clapesattle HB. *The Doctors Mayo*. Garden City, NY: Garden City Publishing; 1943.
12. Mellish MH. *Collected Papers of The Mayo Clinic, Rochester, Minnesota*. Vol XII. Philadelphia, PA: WB Saunders; 1921.
13. Habermann TM, Ziemer RE, Lantz JC. Images and reflections from Mayo Clinic heritage. *Mayo Clin Proc*. 2002;77(11):1182.
14. Logie HB. *A Standard Classified Nomenclature of Disease*. New York, NY: Commonwealth Fund; 1933.
15. O'Fallon JR, Cormack RM, Bithell JF. Obituaries: Joseph Berkson 1899-1982. *Biometrics*. 1983;39(4):1107-1111.
16. Berkson J. A system of codification of medical diagnoses for application to punch cards, with a plan of operation. *Am J Public Health Nations Health*. 1936;26(6):606-612.
17. Morabia A. *A History of Epidemiologic Methods and Concepts*. Boston, MA: Birkhauser Verlag; 2004.
18. Susser M. *Causal Thinking in the Health Sciences: Concepts and Strategies of Epidemiology*. New York, NY: Oxford University Press; 1973.
19. Ellenberg JH, Nelson KB. Sample selection and the natural history of disease: studies of febrile seizures. *JAMA*. 1980; 243(13):1337-1340.
20. Ellenberg JH. Observational data bases in neurological disorders: selection bias and generalization of results. *Neuroepidemiology*. 1994;13(6):268-274.
21. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32(1-2):51-63.
22. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bull*. 1946;2:47-53.
23. Roberts RS, Spitzer WO, Delmore T, et al. An empirical demonstration of Berkson's bias. *J Chronic Dis*. 1978;31(2):119-128.
24. Rocca WA. In Memoriam: Leonard T. Kurland. *Neuroepidemiology*. 2002;21:262-264.
25. Geier GR. *Shadows: A History of the Olmsted Medical Center*. Rochester, MN: Olmsted Medical Center; 2010.
26. Commission on Professional and Hospital Activities, National Center for Health Statistics. *H-ICDA, Hospital Adaptation of ICDA*. 2nd ed. Ann Arbor, MI: National Center for Health Statistics; 1973.
27. Melton LJ III. The threat to medical-records research. *N Engl J Med*. 1997;337(20):1466-1470.
28. Yawn BP, Yawn RA, Geier GR, et al. The impact of requiring patient authorization for use of data in medical records research. *J Fam Pract*. 1998;47(5):361-365.
29. Jacobsen SJ, Xia Z, Campion ME, et al. Potential effect of authorization bias on medical record research. *Mayo Clin Proc*. 1999;74(4):330-338.
30. Office for Civil Rights Health and Human Services. Standards for privacy of individually identifiable health information: final rule. *Fed Regist*. 2002;67:53181-53273.
31. Council for International Organizations of Medical Sciences, World Health Organization. *International Ethical Guidelines for Biomedical Research Involving Human Subjects*. Geneva, Switzerland: World Health Organization; 2002.
32. Council for International Organizations of Medical Sciences, World Health Organization. *International Ethical Guidelines on Epidemiological Studies*. Geneva, Switzerland: World Health Organization; 2009.
33. Hansson MG. Need for a wider view of autonomy in epidemiological research. *BMJ*. 2010;340:c2335.
34. Conway PH, VanLare JM. Improving access to health care data: the Open Government strategy. *JAMA*. 2010;304(9):1007-1008.
35. Higgins MW. The Framingham Heart Study: review of epidemiological design and data, limitations and prospects. *Prog Clin Biol Res*. 1984;147:51-64.
36. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey. http://www.cdc.gov/nchs/nhanes.htm/. Accessed July 31, 2012.
37. Benedetti MD, Bower JH, Maraganore DM, et al. Smoking, alcohol, and coffee consumption preceding Parkinson's disease: a case-control study. *Neurology*. 2000;55(9):1350-1358.
38. Frigerio R, Elbaz A, Sanft KR, et al. Education and occupations preceding Parkinson disease: a population-based case-control study. *Neurology*. 2005;65(10):1575-1583.
39. Acheson ED, Evans JG. The Oxford record linkage study: a review of the method with some preliminary results. *Proc R Soc Med*. 1964;57(4):269-274.
40. Gill L, Goldacre M, Simmons H, et al. Computerised linking of medical records: methodological guidelines. *J Epidemiol Community Health*. 1993;47(4):316-319.
41. Kendrick S, Clarke J. The Scottish record linkage system. *Health Bull (Edinb)*. 1993;51(2):72-79.
42. Kendrick SW, Douglas MM, Gardner D, et al. Best-link matching of Scottish health data sets. *Methods Inf Med*. 1998;37(1): 64-68.
43. Walley T, Mantgani A. The UK General Practice Research Database. *Lancet*. 1997;350(9084):1097-1099.
44. Holman CD, Bass AJ, Rouse IL, et al. Population-based linkage of health records in Western Australia: development of a

health services research linked database. *Aust N Z J Public Health*. 1999;23(5):453-459.

45. Roos LL, Soodeen R-A, Jebamani L. An information-rich environment: linked-record systems and data quality in Canada. In: *Proceedings of Statistics Canada Symposium 2001 - Achieving Data Quality in a Statistical Agency: A Methodological Perspective*. Winnipeg, Canada: Manitoba Centre for Health Policy and Research; 2001:1-6.

46. Roos LL, Menec V, Currie RJ. Policy analysis in an information-rich environment. *Soc Sci Med*. 2004;58(11):2231-2241.

47. Sommer A. *Getting What We Deserve: Health and Medical Care in America*. Baltimore, MD: Johns Hopkins University Press; 2009.

48. Nattinger AB, Pezzin LE, Sparapani RA, et al. Heightened attention to medical privacy: challenges for unbiased sample recruitment and a possible solution. *Am J Epidemiol*. 2010;172: 637-644.