
DATA RESOURCE PROFILE

Data Resource Profile: The Rochester Epidemiology Project (REP) medical records-linkage system

Jennifer L St Sauver,¹ Brandon R Grossardt,² Barbara P Yawn,³ L Joseph Melton III,¹
Joshua J Pankratz,⁴ Scott M Brue⁴ and Walter A Rocca^{1,5*}

¹Division of Epidemiology, Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, MN, USA,

²Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, MN, USA, ³Department of Research, Olmsted Medical Center, Rochester, MN, USA, ⁴Division of Research and Education Support Systems, Department of Information Technology, College of Medicine, Mayo Clinic, Rochester, MN, USA and ⁵Department of Neurology, College of Medicine, Mayo Clinic, Rochester, MN, USA

*Corresponding author. Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. E-mail: rocca@mayo.edu

Accepted 9 October 2012

The Rochester Epidemiology Project (REP) medical records-linkage system was established in 1966 to capture health care information for the entire population of Olmsted County, MN, USA. The REP includes a dynamic cohort of 502 820 unique individuals who resided in Olmsted County at some point between 1966 and 2010, and received health care for any reason at a health care provider within the system. The data available electronically (electronic REP indexes) include demographic characteristics, medical diagnostic codes, surgical procedure codes and death information (including causes of death). In addition, for each resident, the system keeps a complete list of all paper records, electronic records and scanned documents that are available in full text for in-depth review and abstraction. The REP serves as the research infrastructure for studies of virtually all diseases that come to medical attention, and has supported over 2000 peer-reviewed publications since 1966. The system covers residents of all ages and both sexes, regardless of socio-economic status, ethnicity or insurance status. For further information regarding the use of the REP for a specific study, please visit our website at www.rochesterproject.org or contact us at info@rochesterproject.org. Our website also provides access to an introductory video in English and Spanish.

Data resource basics

The Rochester Epidemiology Project (REP) medical records-linkage system was established in 1966 to provide longitudinal medical data for a complete population residing in a well-defined geographic region. The REP captures virtually all individuals who have resided in Olmsted County, MN, USA at some time from 1966 to the present, regardless of

age, sex, ethnicity, disease status, socio-economic status or insurance status. The REP has been continuously funded by the National Institutes of Health for 47 years, and is currently funded by the National Institute on Aging (grant AG034676). Further details about the technical, organizational and methodological developments and about the major events and protagonists of the history in the REP are

available elsewhere.¹ The REP records-linkage system provides the infrastructure to study specific diseases and health outcomes that come to medical attention in the Olmsted County population across all age groups and in both men and women. Common uses of the REP are incidence and prevalence studies, case-control studies, cohort studies, cost or cost-effectiveness studies and natural history or outcomes studies.

Data resource area and population coverage

The REP currently includes a dynamic cohort of 502 820 unique individuals who contributed a total of 6 239 353 person-years of follow-up. These individuals were residents of Olmsted County at some time between 1 January 1966 and 31 December 2010 and received health care from a participating care provider during the same period. More than 50 health care providers have participated in the REP since 1966 including local private practitioners, state hospitals and a tuberculosis sanitarium. When these practices closed over the years, they donated their medical records to the REP. The REP has scanned and indexed these records, and they are available for research studies. Medical diagnoses and surgical procedures have been coded and stored electronically to facilitate the identification of participants for studies. The current care providers participating in the REP are the Mayo Clinic and its two affiliated hospitals (St Marys and Rochester Methodist), the Olmsted Medical Center, its branch offices and its affiliated hospital (Olmsted Medical Center Hospital), and the Rochester Family Medicine Clinic (a private medical care practice in Olmsted County). Dental clinics are now being incorporated into the system.¹ Data collection is ongoing, and new participants and medical records are added to the REP infrastructure either quarterly or twice a year.

We have previously shown that the linkage of information from these health care providers captures virtually the entire Olmsted County population. Indeed, REP estimates of the Olmsted County population are 2–4% higher than those reported by the US Decennial Census.² Table 1 shows the age and sex distribution of the Olmsted County population based on the REP estimates for 1966 and at 10-year intervals from 1970 to 2010. Other characteristics of this population (e.g. ethnic group and education) have been reported elsewhere.³

In 1996, the State of Minnesota introduced a law to protect the confidentiality of medical record information (Minnesota state privacy law—Statute 144.335; amended in 1997). This law requires all Minnesota health care providers to make two attempts (at least 60 days apart) to obtain written permission from each patient seen after 1 January 1997 before medical

records can be used for research. If a patient does not respond to either contact, authorization is implied, and the record may be used for research. Additionally, authorization is implied for patients who only received care before 1 January 1997. The authorization does not expire, but can be revoked upon patient request. Parents or guardians are asked to authorize use of medical records for children <18 years of age. Once children turn 18 years old, they must sign their own authorization. All health care providers who participate in the REP have established procedures to comply with this law.^{1,2}

Table 2 shows the age and sex distribution of the Olmsted County residents who have agreed to allow their medical record information to be used for research. Because research authorization is specific to a health care provider, participants can agree to participate at some providers, but refuse at others. Of the 1 January 2000 residents, 85% agreed to participate at all providers and an additional 13% agreed to participate for at least one provider. Only 2% refused to participate at all providers (in total, 98%). Rates were similar for the 2010 population, with overall participation for at least one provider of 98%. Thus, the REP captures health care information on virtually the entire Olmsted County population.

Frequency of follow-up

Follow-up of patients is done at the discretion of patients and their health care providers as part of routine care. Each time an Olmsted County resident visits a REP health care provider, the information from that clinical visit is automatically integrated into the REP research infrastructure. To describe follow-up patterns by age and sex, we defined a cohort of 127 203 participants who resided in Olmsted County on 1 January 2000. The baseline for each participant was the visit closest to 1 January 2000. We then followed the cohort to determine the percentage of participants who had returned for a health care visit within 1, 2 and 3 years after baseline. Overall, 80% of Olmsted County residents were seen at least once within 1 year and 93% within 3 years (Figure 1). More than 90% of infants (0–2 years of age) and >90% of older adults (≥ 70 years) returned for a visit within 1 year (Figure 1). Women returned sooner (more frequently) than men, and >85% of women at all ages returned within 3 years. Men in the 19–25-year age range were the least likely to return for a health care visit at a participating provider; only 80% were seen at least once within 3 years. In summary, the vast majority of the population had at least one follow-up visit within 3 years.

Attrition from the REP occurs when individuals either die or move out of the county and no longer receive their health care at one of the participating REP providers. Attrition through migration out of Olmsted County is tracked via address data obtained

Table 1 Distribution of the population of Olmsted County, MN, USA, as enumerated by the Rochester Epidemiology Project by age, sex and selected calendar year from 1966 to 2010

Age group ^a	Population on 1 January					
	1966 <i>n</i> (%)	1970 <i>n</i> (%)	1980 <i>n</i> (%)	1990 <i>n</i> (%)	2000 <i>n</i> (%)	2010 <i>n</i> (%)
Women						
0–4	4841 (11.3)	4258 (8.9)	3567 (7.1)	4789 (8.4)	4591 (7.0)	5726 (7.3)
5–9	4066 (9.5)	4466 (9.3)	3168 (6.3)	4266 (7.5)	4425 (6.7)	5237 (6.7)
10–14	3066 (7.2)	3833 (8.0)	3284 (6.5)	3447 (6.1)	4795 (7.3)	4556 (5.8)
15–19	5030 (11.7)	5237 (10.9)	5344 (10.6)	3882 (6.8)	4763 (7.2)	4929 (6.3)
20–24	5349 (12.5)	6380 (13.3)	6920 (13.7)	4921 (8.6)	5090 (7.7)	6296 (8.0)
25–29	3936 (9.2)	4830 (10.1)	5468 (10.8)	6026 (10.6)	4935 (7.5)	6931 (8.8)
30–34	2683 (6.3)	3363 (7.0)	4143 (8.2)	5823 (10.2)	4837 (7.3)	5792 (7.4)
35–39	2222 (5.2)	2404 (5.0)	3262 (6.5)	4505 (7.9)	5670 (8.6)	4830 (6.2)
40–44	2012 (4.7)	2204 (4.6)	2561 (5.1)	3720 (6.5)	5543 (8.4)	4659 (5.9)
45–49	1795 (4.2)	2019 (4.2)	2049 (4.1)	3013 (5.3)	4509 (6.8)	5546 (7.1)
50–54	1681 (3.9)	1750 (3.6)	1921 (3.8)	2458 (4.3)	3698 (5.6)	5707 (7.3)
55–59	1462 (3.4)	1587 (3.3)	1794 (3.6)	1875 (3.3)	2926 (4.4)	4555 (5.8)
60–64	1249 (2.9)	1444 (3.0)	1528 (3.0)	1744 (3.1)	2310 (3.5)	3575 (4.6)
65–69	1149 (2.7)	1242 (2.6)	1391 (2.8)	1628 (2.9)	1839 (2.8)	2829 (3.6)
70–74	946 (2.2)	1112 (2.3)	1287 (2.5)	1413 (2.5)	1716 (2.6)	2223 (2.8)
75–79	648 (1.5)	934 (1.9)	1124 (2.2)	1233 (2.2)	1564 (2.4)	1767 (2.3)
80–84	422 (1.0)	562 (1.2)	836 (1.7)	1044 (1.8)	1208 (1.8)	1481 (1.9)
85–89	222 (0.5)	267 (0.6)	534 (1.1)	713 (1.3)	869 (1.3)	1099 (1.4)
≥90	95 (0.2)	124 (0.3)	298 (0.6)	451 (0.8)	596 (0.9)	710 (0.9)
All ages	42 874 (100.0)	48 016 (100.0)	50 479 (100.0)	56 951 (100.0)	65 884 (100.0)	78 448 (100.0)
Men						
0–4	5263 (15.2)	4846 (12.0)	3946 (9.0)	4953 (9.6)	4843 (7.9)	6086 (8.8)
5–9	4392 (12.7)	5015 (12.4)	3405 (7.8)	4582 (8.9)	4866 (7.9)	5320 (7.7)
10–14	3270 (9.4)	4195 (10.4)	3681 (8.4)	3773 (7.3)	4944 (8.1)	4676 (6.7)
15–19	3048 (8.8)	3847 (9.5)	4719 (10.8)	3668 (7.1)	4886 (8.0)	4869 (7.0)
20–24	2677 (7.7)	3290 (8.1)	4785 (10.9)	4094 (8.0)	4481 (7.3)	4675 (6.7)
25–29	2956 (8.5)	3551 (8.8)	4530 (10.3)	5155 (10.0)	4454 (7.3)	5334 (7.7)
30–34	2457 (7.1)	3023 (7.5)	3580 (8.2)	5297 (10.3)	4702 (7.7)	5135 (7.4)
35–39	1890 (5.5)	2290 (5.7)	2868 (6.5)	4214 (8.2)	5313 (8.7)	4400 (6.3)
40–44	1742 (5.0)	1965 (4.9)	2416 (5.5)	3292 (6.4)	5172 (8.4)	4208 (6.1)
45–49	1475 (4.3)	1832 (4.5)	2015 (4.6)	2766 (5.4)	4274 (7.0)	4900 (7.1)
50–54	1334 (3.9)	1557 (3.8)	1804 (4.1)	2307 (4.5)	3308 (5.4)	4892 (7.1)
55–59	1136 (3.3)	1366 (3.4)	1699 (3.9)	1848 (3.6)	2672 (4.4)	4097 (5.9)
60–64	887 (2.6)	1114 (2.7)	1343 (3.1)	1578 (3.1)	2115 (3.4)	3048 (4.4)
65–69	754 (2.2)	907 (2.2)	1041 (2.4)	1363 (2.6)	1682 (2.7)	2435 (3.5)
70–74	567 (1.6)	698 (1.7)	818 (1.9)	1020 (2.0)	1392 (2.3)	1914 (2.8)
75–79	409 (1.2)	485 (1.2)	569 (1.3)	739 (1.4)	1085 (1.8)	1444 (2.1)
80–84	220 (0.6)	329 (0.8)	367 (0.8)	467 (0.9)	660 (1.1)	1076 (1.6)
85–89	102 (0.3)	141 (0.3)	199 (0.5)	259 (0.5)	328 (0.5)	559 (0.8)
≥90	36 (0.1)	62 (0.2)	71 (0.2)	92 (0.2)	142 (0.2)	259 (0.4)
All ages	34 615 (100.0)	40 513 (100.0)	43 856 (100.0)	51 467 (100.0)	61 319 (100.0)	69 327 (100.0)
Total	77 489	88 529	94 335	108 418	127 203	147 775

^aAge was stratified in 5-year age groups, except for the group ≥90.

Table 2 Distribution of the population of Olmsted County, MN, USA, by response to Minnesota research authorization in two separate calendar years, by age and sex

Group ^a	Authorization in 1 January 2000 population			Authorization in 1 January 2010 population		
	Yes, all ^b n (%)	Yes, some ^b n (%)	No, all ^b n (%)	Yes, all ^b n (%)	Yes, some ^b n (%)	No, all ^b n (%)
Women						
0–9	8182 (90.7)	588 (6.5)	246 (2.7)	9549 (87.1)	877 (8.0)	537 (4.9)
10–19	7961 (83.3)	1463 (15.3)	134 (1.4)	8429 (88.9)	774 (8.2)	282 (3.0)
20–29	8208 (81.9)	1593 (15.9)	224 (2.2)	11 212 (84.8)	1716 (13.0)	299 (2.3)
30–39	8927 (85.0)	1358 (12.9)	222 (2.1)	8823 (83.1)	1533 (14.4)	266 (2.5)
40–49	8581 (85.4)	1271 (12.6)	200 (2.0)	8644 (84.7)	1350 (13.2)	211 (2.1)
50–59	5685 (85.8)	845 (12.8)	94 (1.4)	8781 (85.6)	1288 (12.6)	193 (1.9)
60–69	3685 (88.8)	405 (9.8)	59 (1.4)	5458 (85.2)	835 (13.0)	111 (1.7)
70–79	2953 (90.0)	290 (8.8)	37 (1.1)	3547 (88.9)	390 (9.8)	53 (1.3)
80–89	1892 (91.1)	157 (7.6)	28 (1.3)	2321 (90.0)	225 (8.7)	34 (1.3)
≥90	542 (90.9)	41 (6.9)	13 (2.2)	643 (90.6)	53 (7.5)	14 (2.0)
All ages	56 616 (85.9)	8011 (12.2)	1257 (1.9)	67 407 (85.9)	9041 (11.5)	2000 (2.5)
Men						
0–9	8755 (90.2)	658 (6.8)	296 (3.0)	10 034 (88.0)	851 (7.5)	521 (4.6)
10–19	8042 (81.8)	1663 (16.9)	125 (1.3)	8412 (88.1)	824 (8.6)	309 (3.2)
20–29	6721 (75.2)	2056 (23.0)	158 (1.8)	8333 (83.3)	1508 (15.1)	168 (1.7)
30–39	8160 (81.5)	1713 (17.1)	142 (1.4)	7657 (80.3)	1689 (17.7)	189 (2.0)
40–49	7937 (84.0)	1415 (15.0)	94 (1.0)	7500 (82.3)	1480 (16.2)	128 (1.4)
50–59	5105 (85.4)	824 (13.8)	51 (0.9)	7627 (84.8)	1276 (14.2)	86 (1.0)
60–69	3407 (89.7)	370 (9.7)	20 (0.5)	4664 (85.1)	770 (14.0)	49 (0.9)
70–79	2290 (92.5)	176 (7.1)	11 (0.4)	2996 (89.2)	339 (10.1)	23 (0.7)
80–89	925 (93.6)	60 (6.1)	3 (0.3)	1502 (91.9)	122 (7.5)	11 (0.7)
≥90	137 (96.5)	5 (3.5)	0 (0.0)	238 (91.9)	19 (7.3)	2 (0.8)
All ages	51 479 (84.0)	8940 (14.6)	900 (1.5)	58 963 (85.1)	8878 (12.8)	1486 (2.1)
Total	108 095 (85.0)	16 951 (13.3)	2157 (1.7)	126 370 (85.5)	17 919 (12.1)	3486 (2.4)

^aAge was stratified in 10-year age groups, except for the group ≥90.

^bResidents of Olmsted County can give research authorization to all health care providers, some health care providers, or none.

at the time of contact with a health care provider. We used the cohort of participants who resided in Olmsted County on 1 January 2000 to investigate the attrition rates (lost to follow-up) through 31 December 2010. Figure 2 shows the age-specific percentage of participants who were followed completely through death or through 31 December 2010 (panel A), of participants who were lost to follow-up after at least one return visit (panel B) and of participants who were never seen after the baseline visit (panel C).

Attrition rates were highest for residents 15–29 years of age at baseline and lowest in persons ≥65 years of age. Overall, 63% of the participants who lived in Olmsted County on 1 January 2000 were alive and still resided in Olmsted County 11 years later, 8% had moved outside of Olmsted County but were still alive and routinely receiving medical care at

one of the participating REP providers, and 6% were followed completely through the time of death (complete 11-year follow-up for 77% of participants). Over the 11 years of follow-up, 18% returned for at least one visit after baseline but were eventually lost to follow-up, and 0.2% did not return for a visit but were known to be deceased via state or national sources of information. Only 4% of participants were never seen again after the baseline visit.

Some participants who reside in Olmsted County may eventually move away but continue to receive medical care at one of the REP participating providers; this information remains part of the REP. Inclusion in a specific REP study is often based on disease or exposure status and on Olmsted County residency on a particular index date. Medical information from REP health care providers is often available for many years both before and after that index date,

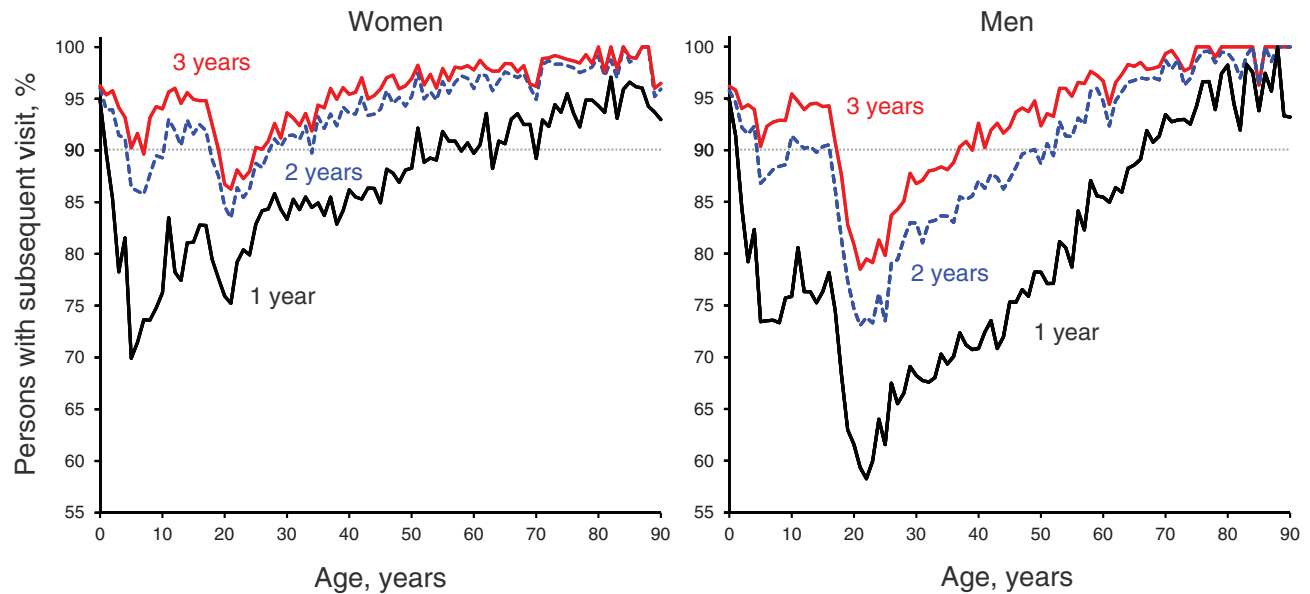


Figure 1 Age-specific percentage of persons returning within 1, 2 and 3 years after the baseline visit (1 January 2000) in women (left panel) and men (right panel). A horizontal reference line at 90% serves as a visual aid

and this information is accessible for research regardless of residency.² A total of 502 820 persons have lived in Olmsted County at some time between 1 January 1966 and 31 December 2010. These persons have accumulated 4 923 024 person-years of medical information while they were living in Olmsted County and 1 316 329 person-years of information at participating health care providers while they were living elsewhere (6 239 353 total person-years). Unfortunately, participants who move out of the region and do not return for care are lost to follow-up. If these participants are systematically different from those who are followed, some follow-up bias may be introduced.⁴

Measures

Electronic data

All health care providers participating in the REP contribute electronic demographic information (name, sex, date of birth, address), a care provider-specific identification number (e.g. a Mayo Clinic patient number) and diagnostic codes for medical conditions and surgical procedures. Data are obtained either quarterly or twice a year from all providers. After the diagnostic and surgical codes have been properly linked to the corresponding participants, they are stored in electronic REP indexes. We emphasize that the linkage occurs both within and across institutions.^{1,2}

Medical diagnosis data have been coded using three different coding systems, depending on the site from which the data were received and depending on the year of the diagnosis, including the 'Berkson Coding

System' (developed by Joseph Berkson at the Mayo Clinic in 1935),¹ the Hospital Adaptation of the International Classification of Diseases, Eighth Revision (H-ICDA) coding system⁵ and the International Classification of Diseases, Ninth Revision (ICD-9) coding system.⁶ Surgical procedures were coded using the Berkson coding system from 1935 to 1987, and the ICD-9 coding system from 1987 to the present. More recently, some outpatient surgical procedures are coded in the Current Procedural Terminology (CPT) coding system.

Investigators use combinations of Berkson, H-ICDA, ICD-9 or CPT codes to identify lists of potential participants with a disease or procedure of interest.¹ Owing to the complex nature of data retrieval and the many coding systems used in the indexes over time, the REP employs medical index retrieval specialists specifically trained to identify the proper codes and to obtain the lists of relevant participants. It is also possible to use the REP indexes to identify referent participants for cohort studies or control participants for case-control studies.

The REP also captures electronic death information through multiple sources. Dates of death are routinely tracked and documented through each health care provider, and are obtained at the same time as the diagnosis and procedure updates. We also receive electronic Minnesota State Death Certificates and match these certificates to all individuals in the REP database on a quarterly basis. Therefore, cause of death information is available if the individual died within the State of Minnesota. Additionally, we supplement these data with information obtained biennially from the National Death Index for Olmsted County residents who migrate out of the county and

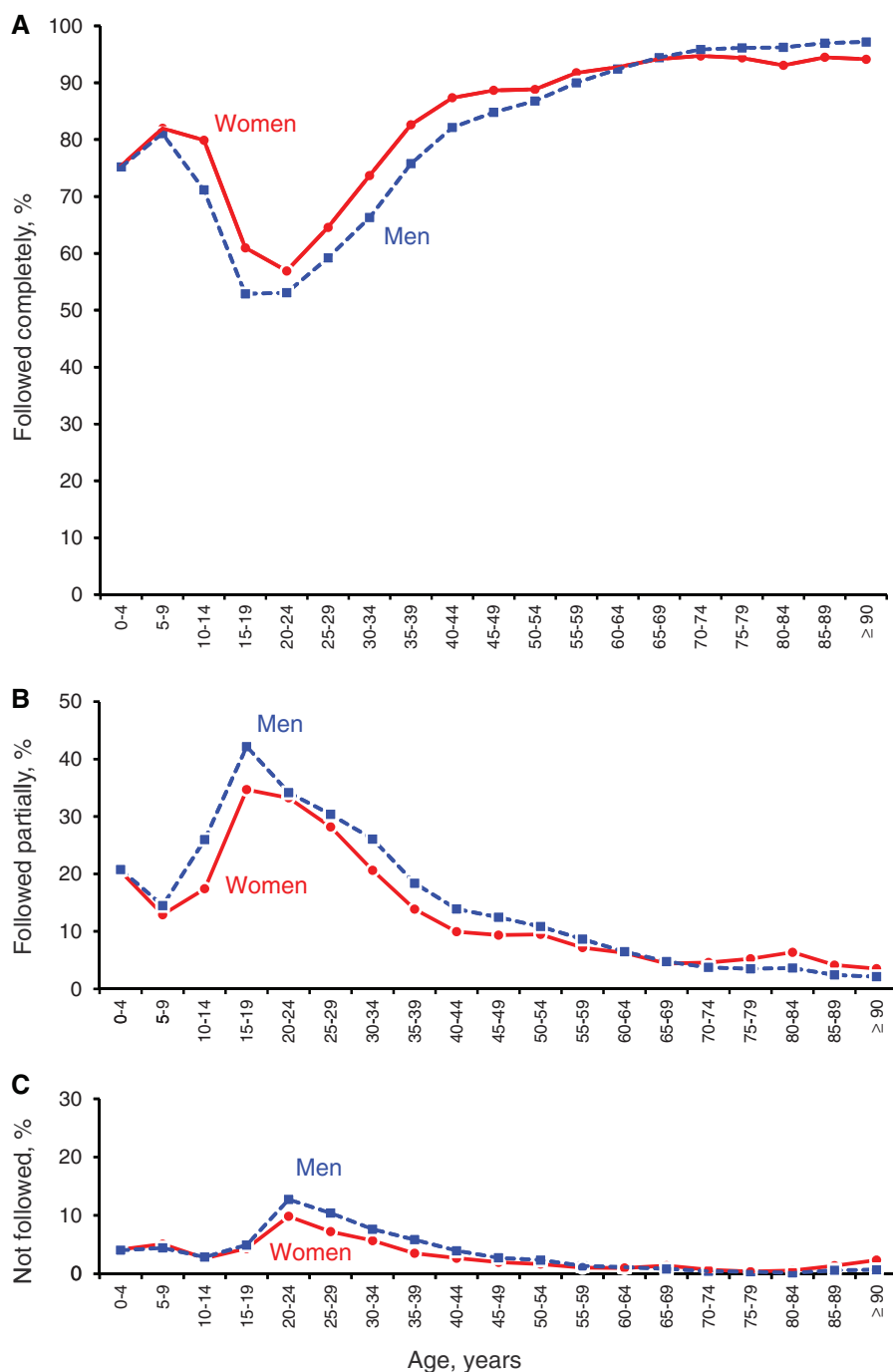


Figure 2 Follow-up status of Olmsted County residents from 1 January 2000 through 31 December 2010 (11 years of follow-up). Percentage of participants who were followed completely through death or through 31 December 2010 (censored alive at end of follow-up); (A) percentage of participants who were lost to follow-up after at least one return visit (partially followed); (B) percentage of participants who never returned to one of the health care providers participating in the REP after the baseline visit; (C) no follow-up

die outside Minnesota. As we previously reported, death rates in the Olmsted County population are similar to death rates in Minnesota and the rest of the United States.³

We also capture electronic data for residents who undergo an autopsy in Olmsted County. Autopsy

rates in Olmsted County have typically been higher than in the rest of Minnesota.⁷ Figure 3A shows average age- and sex-specific autopsy rates for the years 1966–2010 combined. Autopsy rates were highest in the younger population, particularly young men, but declined with increasing age. Figure 3B shows

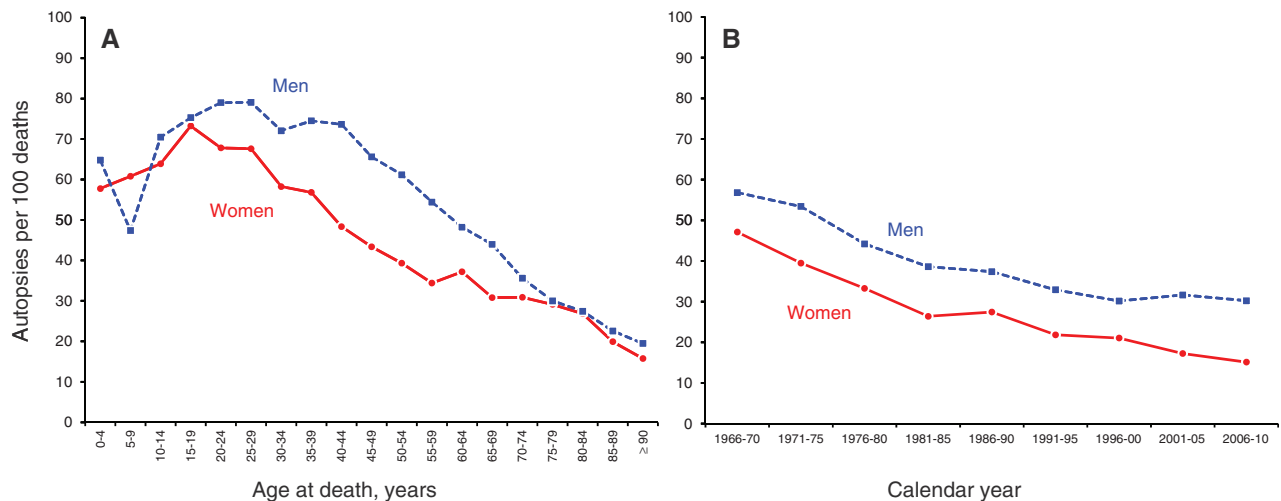


Figure 3 Age- and sex-specific autopsy rates in Olmsted County, MN, for the period 1966–2010 combined (A) and age-adjusted autopsy rates over almost half a century (B) rates adjusted to the age distribution of deaths in the 2006–2010 period

age-adjusted autopsy rates over almost half a century for men and women separately. In all time periods, the autopsy rates were higher in men, and declined equally in men and women over time.

Medical record data

Following approval from the Institutional Review Boards of the Mayo Clinic and the Olmsted Medical Center, the REP also provides access to the full text of medical records for the participants who have been identified (if they have provided research authorization). Because the residents of Olmsted County frequently obtain their care from multiple health care providers, the REP matches these multiple medical records to individual residents.² A listing of all medical records matched to an individual patient can be accessed through a web-based application called the 'REP Browser'. Figure 4 shows an example of a search for 'La Tester' (artificial patient name). Following entry of this name into the REP Browser (circled portion of Figure 4), information was returned on all records belonging to Lars Tester. In this example, Lars Tester has three medical records available as part of the REP from multiple health care sites (rectangular box of Figure 4). Two of these records are available under the name 'Lars Tester', whereas one is available under the name 'L B Tester'. The extensive matching processes conducted by the REP staff,² combined with the easy retrieval of all information available for an individual through the REP Browser, make it possible for investigators to determine which medical records exist for an individual participant. These records can then be retrieved and reviewed to obtain patient information that is not available electronically from the indexes (e.g. detailed symptoms or functional outcomes). Some participants only have paper medical records, some only have electronic medical

records and many have a combination of paper and electronic records, all of which are available for review.

Data resource use: key findings and publications

The REP has supported >2000 publications across a wide range of diseases. A complete listing of REP publications is available on the REP website: <http://www.rochesterproject.org>. The 10 most cited publications resulting from studies supported by the REP are listed in Table 3.^{8–18} As expected, the number of citations also reflects the time since publication.

Studies of prevalence and of secular trends in incidence are hallmarks of the REP, and these types of studies are among the most cited REP publications. For example, the REP has made it possible to study changes in the incidence of conditions as diverse as deep vein thrombosis and pulmonary embolism,⁹ epilepsy¹¹ and esophageal adenocarcinoma¹⁴ (Table 3). Additionally, the Olmsted County population is stable, and it is possible to follow patients with a variety of conditions over decades to characterize long-term outcomes. For example, Owan *et al.*¹³ described long-term outcomes of heart failure among patients with preserved ejection fraction, whereas Senni *et al.*¹⁵ described survival rates in congestive heart failure patients.

The REP also serves as an ideal population-based sampling frame to study conditions that may not come to medical attention, or to obtain data that may not be routinely collected as part of clinical care. For example, not all patients with gastroesophageal reflux will seek clinical care, particularly if

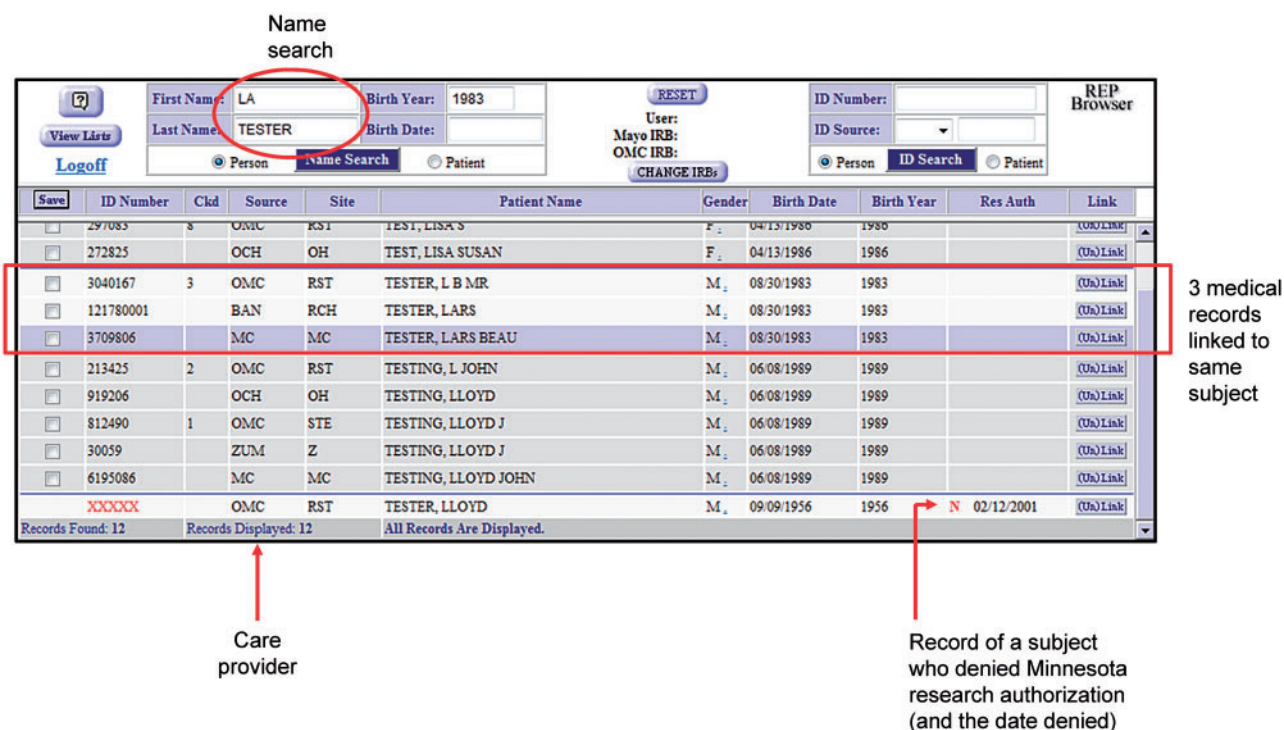


Figure 4 A screen shot of the REP Browser is provided as an example of the layout of information for a given subject in the system. The data shown are artificial data and do not refer to real persons. Only the three records shown within the rectangular box were recognized as linked to the test subject. OMC, Olmsted Medical Center; BAN, Rochester Family Medicine Clinic (Dr Banfield); MC, Mayo Clinic

the symptoms are mild. Studies that attempt to describe this condition may be biased if they only focus on patients who come to medical attention, because less severe cases will not be identified. To address this problem, Locke *et al.*⁸ used the REP as a sampling frame to contact a random sample of Olmsted County residents. Study participants completed standard questionnaires, and the results were used to describe the prevalence and to characterize the symptoms of gastroesophageal reflux in the community. Similarly, before the initiation of widespread screening for prostate cancer, Oesterling *et al.*¹² collected blood specimens and measured prostate-specific antigen levels in a random sample of healthy men residing in Olmsted County. The age-specific reference ranges obtained from this study are still used for screening men for prostate cancer at many health care institutions.

Finally, the REP is an ideal resource for identifying population-based controls or unexposed cohorts, making it possible to conduct unbiased studies of risk factors. An example of this type of study was performed by Heit *et al.*,¹⁶ who used the REP infrastructure to describe risk factors for venous thromboembolism. This study was the first to show that hospital, nursing home or other long-term care confinement was an important risk factor for venous thromboembolism.

Strengths and weaknesses

The primary strength of the REP is the ability to capture information on the health care of all residents of Olmsted County regardless of age, sex, ethnicity, socio-economic status, insurance status or setting of care delivery. The REP allows investigators to conduct population-based research on a wide range of diseases and conditions, to follow patients from primary to tertiary care, without regard to insurance, and to access the full text of medical records. Therefore, patients can be followed across the full spectrum of disease, from symptoms through final diagnosis, without relying only on administrative data. Finally, the population is relatively stable, so the duration of medical record information available to investigators is substantial (Figure 2).²

There are some limitations. The size of the Olmsted County population limits studies of rare conditions (e.g. pancreatic or ovarian cancer). Additionally, it is difficult to study diseases or exposures that do not come to medical attention or are not routinely documented in the medical record (e.g. mild cognitive impairment, gastroesophageal reflux or other preclinical stages of disease). However, medical record information for a specific study may be supplemented by collecting further data through mail, telephone or in-person interviews.^{8,19,20} Olmsted County residents

Table 3 The ten most-cited publications that relied on the Rochester Epidemiology Project records-linkage system^a

Author	Year	Title	Journal	Issue and pages	No. of citations
Locke <i>et al.</i>	1997 ⁸	Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota	<i>Gastroenterology</i>	112(5):1448–56	1262
Silverstein <i>et al.</i>	1998 ⁹	Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study	<i>Arch Intern Med</i>	158(6):585–93	1029
Redfield <i>et al.</i>	2003 ¹⁰	Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic	<i>JAMA</i>	289(2):194–202	1028
Hauser <i>et al.</i>	1993 ¹¹	Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935–1984	<i>Epilepsia</i>	34(3):453–68	792
Oesterling <i>et al.</i>	1993 ¹²	Serum prostate-specific antigen in a community-based population of healthy men. Establishment of age-specific reference ranges	<i>JAMA</i>	270(7):860–64	791
Owan <i>et al.</i>	2006 ¹³	Trends in prevalence and outcome of heart failure with preserved ejection fraction	<i>N Engl J Med</i>	355(3):251–59	741
Pera <i>et al.</i>	1993 ¹⁴	Increasing incidence of adenocarcinoma of the esophagus and esophagogastric junction	<i>Gastroenterology</i>	104(2):510–13	691
Senni <i>et al.</i>	1998 ¹⁵	Congestive heart failure in the community: a study of all incident cases in Olmsted County, Minnesota, in 1991	<i>Circulation</i>	98(21):2282–89	667
Heit <i>et al.</i>	2000 ¹⁶	Risk factors for deep vein thrombosis and pulmonary embolism: a population-based case-control study	<i>Arch Intern Med</i>	160(6):809–15	650
Cooper <i>et al.</i>	1992 ¹⁷	Incidence of clinically diagnosed vertebral fractures: a population-based study in Rochester, Minnesota, 1985–1989	<i>J Bone Miner Res</i>	7(2):221–27	642

^aThe papers are presented in descending order from the highest number of citations to the lowest. The number of citations is also influenced by the time since publication. Another paper on the history of the Rochester Epidemiology Project, published in 1996, has been cited 917 times.¹⁸

may also be invited to a physical examination, to contribute biospecimens or to undergo imaging or laboratory tests for specific research studies.^{12,21,22}

Finally, the ethnic and socio-economic characteristics of the Olmsted County population are similar to other populations in the upper Midwest region of the United States but are different from the characteristics of other populations.³ In particular, some racial and ethnic groups are under-represented. For this reason, results from studies in this population must be considered on a case-by-case basis when attempting to generalize to other populations.³

Data resource access

Details regarding access to REP data for research are available on our website at: www.rochesterproject.org. Inquiries regarding use of the REP for specific

research studies are welcomed. For further information, please contact us at info@rochesterproject.org. Our website also provides access to a video in English or in Spanish that can serve as a brief introduction to REP resources.

Funding

The REP is currently supported by the National Institute on Aging of the National Institutes of Health under Award Number R01 AG034676. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additionally, this publication was supported by CTSA Grant Number UL1 TR000135 from the National Center for Advancing Translational Science (NCATS).

Acknowledgements

We thank Lori Klein for assistance with manuscript preparation.

Conflict of interest: None declared.

KEY MESSAGES

- Studies of prevalence and of secular trends in incidence are hallmarks of the Rochester Epidemiology Project (REP). The REP has made it possible to study changes in the incidence of conditions ranging from deep vein thrombosis and pulmonary embolism to epilepsy and esophageal adenocarcinoma.
- The REP may serve as a population-based sampling frame to study conditions that may not come to medical attention, or to obtain data that may not be routinely collected as part of clinical care. One of these studies established age-specific reference ranges for serum prostate-specific antigen to be used in prostate cancer screening.
- The REP is an ideal resource for identifying population-based controls or unexposed cohorts, making it possible to conduct unbiased studies of risk factors. One of these studies was the first to show that hospital, nursing home or other long-term care confinement is an important risk factor for venous thromboembolism.

References

- ¹ Rocca WA, Yawn BP, St Sauver JL, Grossardt BR, Melton LJ III. History of the Rochester Epidemiology Project: half a century of medical records linkage in a United States population. *Mayo Clin Proc* 2012; (in press).
- ² St Sauver JL, Grossardt BR, Yawn BP, Melton LJ III, Rocca WA. Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project. *Am J Epidemiol* 2011; **173**:1059–68.
- ³ St Sauver JL, Grossardt BR, Leibson CL, Yawn BP, Melton LJ III, Rocca WA. Generalizability of epidemiological findings and public health decisions: an illustration from the Rochester Epidemiology Project. *Mayo Clin Proc* 2012; **87**:151–60.
- ⁴ Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; **32**:51–63.
- ⁵ Activities CoPaH. *H-ICDA, Hospital Adaptation of ICDA*. 2nd edn. Ann Arbor, MI: National Center for Health Statistics, 1973.
- ⁶ World Health Organization. *Manual of the International Classification of Diseases, Injuries, and Causes of Death*, based on the recommendations of the ninth revision conference 1975, and adopted by the 29th World Health Assembly. Geneva, 1977.
- ⁷ Targonski P, Jacobsen SJ, Weston SA *et al*. Referral to autopsy: effect of antemortem cardiovascular disease: a population-based study in Olmsted County, Minnesota. *Ann Epidemiol* 2001; **11**:264–70.
- ⁸ Locke GR III, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ III. Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota. *Gastroenterology* 1997; **112**: 1448–56.
- ⁹ Silverstein MD, Heit JA, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ III. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch Int Med* 1998; **158**: 585–93.
- ¹⁰ Redfield MM, Jacobsen SJ, Burnett JC Jr, Mahoney DW, Bailey KR, Rodeheffer RJ. Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. *JAMA* 2003; **289**: 194–202.
- ¹¹ Hauser WA, Annegers JF, Kurland LT. Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935–1984. *Epilepsia* 1993; **34**:453–68.
- ¹² Oesterling JE, Jacobsen SJ, Chute CG *et al*. Serum prostate-specific antigen in a community-based population of healthy men. Establishment of age-specific reference ranges. *JAMA* 1993; **270**:860–64.
- ¹³ Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM. Trends in prevalence and outcome of heart failure with preserved ejection fraction. *N Engl J Med* 2006; **355**:251–59.
- ¹⁴ Pera M, Cameron AJ, Trastek VF, Carpenter HA, Zinsmeister AR. Increasing incidence of adenocarcinoma of the esophagus and esophagogastric junction. *Gastroenterology* 1993; **104**:510–13.
- ¹⁵ Senni M, Tribouilloy CM, Rodeheffer RJ *et al*. Congestive heart failure in the community: a study of all incident cases in Olmsted County, Minnesota, in 1991. *Circulation* 1998; **98**:2282–89.
- ¹⁶ Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ III. Risk factors for deep vein thrombosis and pulmonary embolism: a population-based case-control study. *Arch Int Med* 2000; **160**: 809–15.
- ¹⁷ Cooper C, Atkinson EJ, O'Fallon WM, Melton LJ III. Incidence of clinically diagnosed vertebral fractures: a population-based study in Rochester, Minnesota, 1985–1989. *J Bone Mineral Res* 1992; **7**:221–27.
- ¹⁸ Melton LJ III. History of the Rochester Epidemiology Project. *Mayo Clin Proc* 1996; **71**:266–74.
- ¹⁹ Rocca WA, Peterson BJ, McDonnell SK *et al*. The Mayo Clinic family study of Parkinson's disease: study design, instruments, and sample characteristics. *Neuroepidemiology* 2005; **24**:151–67.
- ²⁰ Rocca WA, Bower JH, Maraganore DM *et al*. Increased risk of cognitive impairment or dementia in women

who underwent oophorectomy before menopause. *Neurology* 2007;**69**:1074–83.

- ²¹ Roberts RO, Geda YE, Knopman DS *et al.* The incidence of MCI differs by subtype and is higher in men: the Mayo Clinic Study of Aging. *Neurology* 2012;**78**: 342–51.

- ²² Whitwell JL, Wiste HJ, Weigand SD *et al.* Comparison of imaging biomarkers in the Alzheimer disease neuroimaging initiative and the Mayo Clinic Study of Aging. *Arch Neurol* 2012;**69**:614–22.