Improving health globally by studying health locally

User Manual of the Data Exploration Portal (DEP)

DEP release version 7.0, November 2024

Technical and Programming Team

Brandon R. Grossardt[*], Cynthia M. Bock-Goodner[†], Barbara A. Abbott[‡], Andrew W. Ringlien[†], Michael M. Truong[†], Tracy M. Safran[†]

Principal Investigators

Jennifer St. Sauver[‡§] and Walter A. Rocca[‡¶]

REP Collaborators

MAYO CLINIC

OLMSTED MEDICAL CENTER

MAYO CLINIC HEALTH SYSTEM

COUNTY OF Olmsted

ZUMBRO VALLEY HEALTH CENTER

[*]Division of Clinical Trials and Biostatistics, Mayo Clinic, Rochester, MN
[†]Division of Information Management and Analytical Systems, Mayo Clinic, Rochester, MN
[‡]Division of Epidemiology, Mayo Clinic, Rochester, MN
[§]Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN
[¶]Department of Neurology, Mayo Clinic, Rochester, MN

Table of Contents

Abstract

The Rochester Epidemiology Project is a collaboration of clinics, hospitals, and other medical facilities in Minnesota and Wisconsin and includes data from community members who have agreed to share their medical records for research. Using medical record information, medical scientists can discover the causes of diseases, how people respond to medical and surgical therapies, and what may happen to people long term. Research studies conducted in the local community may improve the health of people both locally and globally.

As a way to share data with researchers, clinicians, and the general public, we have developed an online portal for exploring relationships between various characteristics. The portal allows users to choose two characteristics, to summarize the prevalence of each separately by age and sex, and to explore the co-occurrence of the two characteristics. In addition, users are presented with shaded maps to help visualize the prevalence and co-occurrence of the two selected characteristics across the REP region. Finally, users can see whether the 1-year period prevalence of characteristics changed over recent years (trends).

1. Inclusion

1.1 Population

The Rochester Epidemiology Project (**REP**) is a medical records-linkage system which links information across multiple medical providers to individual persons living in the community.[1,2] The data available include medical diagnostic codes, surgical procedure codes, drug prescriptions, demographic characteristics, and death information including causes of death. Using the extensive REP infrastructure, the Data Exploration Portal (**DEP**) includes summary-level information for all persons who lived in the 27-county region of southern Minnesota and western Wisconsin (see **Figure 1**).[3] These persons were identified using the personal residency timelines constructed by the REP infrastructure.[4,5]

DEP version 7.0 (released November 2024) includes persons who were residents of the 27-county region on 1 Jan 2022. The included population comprises more than 735,000 persons (see **Table 1**). Only persons who gave permission for medical record information to be used for research purposes are summarized.[6] More than 93% of persons gave permission for some or all of their medical record information to be used for research (see **Table 1**, footnote). About 62% of the complete population across the 27-county region is included (see **Table 2**).

Figure 1. *The 27-county region of the Rochester Epidemiology Project. It includes 19 counties in southern Minnesota, and 8 counties in western Wisconsin*

1.2 Characteristics

DEP users can choose from among 1,253 characteristics determined by diagnosis code condition groupings, procedure and service code groupings, medication class groupings, and personal characteristics available from medical records (e.g., race and ethnicity). The complete code listings defining each of the 1,253 characteristics can be found as part of the documentation for the Data Exploration Portal at https://rochesterproject.org/portal/.

Table 1. REP Census Population on 1 Jan 2022 included in the Data Exploration Portal.*

	Men, N						Women, N						Both sexes, N					
	Age group, yrs						Age group, yrs						Age group, yrs					
County	0-20	21-39	40-64	65-79	≥80	All Ages	0-20	21-39	40-64	65-79	≥80	All Ages	0-20	21-39	40-64	65-79	≥80	All Ages
Minnesota counties																		
Olmsted	22,694	21,364	23,412	9,091	2,748	79,309	22,128	24,215	24,758	10,606	4,026	85,733	44,822	45,579	48,170	19,697	6,774	165,042
Dodge	2,932	2,082	3,084	1,136	324	9,558	2,804	2,329	3,125	1,180	457	9,895	5,736	4,411	6,209	2,316	781	19,453
Mower	5,766	4,156	5,216	2,402	760	18,300	5,724	4,398	5,399	2,500	1,200	19,221	11,490	8,554	10,615	4,902	1,960	37,521
Goodhue	5,356	4,403	6,218	3,166	950	20,093	5,082	4,634	6,557	3,389	1,314	20,976	10,438	9,037	12,775	6,555	2,264	41,069
Fillmore	1,971	1,476	2,482	1,362	432	7,723	1,944	1,750	2,588	1,449	585	8,316	3,915	3,226	5,070	2,811	1,017	16,039
Wabasha	2,433	1,875	3,203	1,721	527	9,759	2,439	2,068	3,301	1,831	577	10,216	4,872	3,943	6,504	3,552	1,104	19,975
Winona	1,646	1,463	2,143	1,312	391	6,955	1,638	1,674	2,336	1,343	451	7,442	3,284	3,137	4,479	2,655	842	14,397
Houston	0,909	781	1,241	761	176	3,868	916	736	1,260	741	294	3,947	1,825	1,517	2,501	1,502	470	7,815
Freeborn	3,780	2,616	3,818	1,988	639	12,841	3,441	2,783	3,961	2,125	1,040	13,350	7,221	5,399	7,779	4,113	1,679	26,191
Steele	4,642	3,374	4,828	2,119	593	15,556	4,539	3,561	4,935	2,236	895	16,166	9,181	6,935	9,763	4,355	1,488	31,722
Rice	2,099	1,704	2,603	1,342	484	8,232	2,213	2,206	2,937	1,454	547	9,357	4,312	3,910	5,540	2,796	1,031	17,589
Blue Earth	5,243	4,894	5,104	2,537	807	18,585	5,217	5,750	5,368	2,675	1,120	20,130	10,460	10,644	10,472	5,212	1,927	38,715
Waseca	2,071	1,462	2,279	1,102	291	7,205	2,064	1,865	2,436	1,100	424	7,889	4,135	3,327	4,715	2,202	715	15,094
Faribault	0,894	724	1,185	740	225	3,768	867	844	1,229	712	354	4,006	1,761	1,568	2,414	1,452	579	7,774
Martin	2,068	1,555	2,121	1,201	384	7,329	1,984	1,640	2,242	1,231	645	7,742	4,052	3,195	4,363	2,432	1,029	15,071
Watonwan	1,019	694	907	517	194	3,331	1,065	849	974	507	238	3,633	2,084	1,543	1,881	1,024	432	6,964
Brown	0,383	303	589	451	168	1,894	347	396	658	431	183	2,015	730	699	1,247	882	351	3,909
Nicollet	2,283	1,746	2,547	1,328	444	8,348	2,286	2,318	2,677	1,372	616	9,269	4,569	4,064	5,224	2,700	1,060	17,617
Le Sueur	1,918	1,291	2,139	1,156	334	6,838	1,882	1,535	2,224	1,162	446	7,249	3,800	2,826	4,363	2,318	780	14,087
Wisconsin counties																		
Eau Claire	7,692	8,062	8,889	4,375	1,222	30,240	7,659	8,457	8,916	5,006	1,821	31,859	15,351	16,519	17,805	9,381	3,043	62,099
Trempealeau	2,307	1,786	2,803	1,360	300	8,556	2,271	1,911	2,691	1,367	477	8,717	4,578	3,697	5,494	2,727	777	17,273
La Crosse	5,564	6,270	7,255	3,152	734	22,975	5,751	6,077	7,240	3,534	1,108	23,710	11,315	12,347	14,495	6,686	1,842	46,685
Buffalo	0,737	648	1,133	702	178	3,398	803	672	1,102	612	232	3,421	1,540	1,320	2,235	1,314	410	6,819
Pepin	0,426	323	647	412	130	1,938	439	392	611	386	149	1,977	865	715	1,258	798	279	3,915
Dunn	4,519	3,697	5,036	2,522	626	16,400	4,315	3,942	4,929	2,668	912	16,766	8,834	7,639	9,965	5,190	1,538	33,166
Barron	2,422	2,112	3,300	1,826	510	10,170	2,484	2,278	3,317	1,926	716	10,721	4,906	4,390	6,617	3,752	1,226	20,891
Chippewa	3,596	3,104	4,776	2,373	618	14,467	3,563	3,391	4,655	2,465	888	14,962	7,159	6,495	9,431	4,838	1,506	29,429
All 27 counties	97,370	83,965	108,958	52,154	15,189	357,636	95,865	92,671	112,426	56,008	21,715	378,685	193,235	176,636	221,384	108,162	36,904	736,321

* Persons who have given permission for all or part of their medical record information to be used for research purposes. The complete population enumerated by the REP Census on 1 Jan 2022 comprised 784,837 persons (379,836 men and 405,001 women). Corresponding inclusion rates are 93.8% overall, 94.2% for men, and 93.5% for women.

Table 2. REP Census coverage as compared to the 2022 US Intercensal Population Estimates.*

	Men, % coverage					Women, % coverage					Both sexes, % coverage				
	Age group, yrs					Age group, yrs					Age group, yrs				
County	0-39	40-64	65-79	≥80	All Ages	0-39	40-64	65-79	≥80	All Ages	0-39	40-64	65-79	≥80	All Ages
Minnesota counties															
Olmsted	101.0	95.7	96.0	99.7	98.7	106.0	100.5	97.0	90.7	102.4	103.5	98.1	96.5	94.2	100.6
Dodge	90.2	89.1	89.8	94.7	89.9	96.1	94.6	91.7	104.6	95.5	93.1	91.8	90.8	100.3	92.7
Mower	93.1	85.3	89.7	98.2	90.5	100.0	93.7	93.6	94.3	96.9	96.5	89.4	91.6	95.7	93.7
Goodhue	84.5	78.6	83.9	109.4	83.4	89.2	84.5	85.4	99.5	87.6	86.8	81.5	84.7	103.4	85.5
Fillmore	66.9	72.0	77.7	88.7	71.2	75.1	81.0	83.6	76.4	78.4	70.9	76.3	80.6	81.2	74.8
Wabasha	88.6	90.1	89.5	96.2	89.6	95.0	97.4	94.6	82.9	94.9	91.8	93.7	92.1	88.7	92.3
Winona	23.4	30.5	36.4	47.5	28.1	25.2	34.9	37.0	35.2	30.1	24.3	32.6	36.7	40.0	29.1
Houston	39.6	40.1	44.2	43.2	40.7	42.2	41.0	44.9	49.7	42.7	40.8	40.5	44.6	47.0	41.7
Freeborn	88.3	78.4	77.1	86.1	83.2	91.2	85.9	79.4	91.7	87.6	89.7	82.0	78.2	89.5	85.4
Steele	84.1	81.6	82.7	79.7	82.9	88.7	86.4	82.4	79.6	86.6	86.4	84.0	82.5	79.6	84.8
Rice	20.2	24.2	32.1	46.1	23.6	25.6	30.6	32.8	38.0	28.6	22.8	27.2	32.4	41.4	26.0
Blue Earth	47.5	57.1	66.9	80.3	53.0	54.1	62.5	65.4	71.0	58.3	50.7	59.8	66.1	74.6	55.6
Waseca	77.3	78.6	77.5	82.0	77.9	82.3	80.9	79.8	85.3	81.6	79.8	79.7	78.6	83.9	79.8
Faribault	49.9	53.3	57.9	72.8	53.4	56.4	58.3	58.7	68.1	58.3	53.0	55.7	58.3	69.8	55.8
Martin	79.9	70.0	67.9	75.6	74.5	82.7	76.8	69.7	86.6	78.9	81.3	73.3	68.8	82.1	76.7
Watonwan	62.1	54.0	61.3	74.3	60.1	69.9	60.2	62.6	59.6	65.3	66.0	57.0	61.9	65.5	62.7
Brown	10.6	15.1	22.4	27.2	14.6	12.6	17.1	20.4	19.4	15.8	11.6	16.1	21.3	22.5	15.2
Nicollet	43.9	47.5	59.4	75.0	48.1	51.8	56.0	59.0	62.3	54.6	47.8	51.5	59.2	67.0	51.3
Le Sueur	44.9	41.7	56.0	65.1	46.0	50.6	47.5	55.8	59.9	50.8	47.6	44.5	55.9	62.1	48.4
Wisconsin counties															
Eau Claire	52.8	59.9	64.5	79.1	57.1	55.3	61.1	66.2	72.5	59.2	54.0	60.5	65.4	75.0	58.2
Trempealeau	51.6	55.3	59.6	55.2	54.1	55.8	58.9	62.5	58.2	57.8	53.6	57.0	61.0	57.0	55.9
La Crosse	37.2	42.0	39.3	38.0	38.9	36.5	42.2	40.5	37.9	38.7	36.8	42.1	39.9	37.9	38.8
Buffalo	48.3	49.2	52.4	58.9	49.9	52.2	49.9	54.1	56.2	52.0	50.2	49.6	53.2	57.3	50.9
Pepin	45.8	52.1	57.2	68.1	51.2	53.9	51.8	58.1	72.7	55.1	49.7	51.9	57.7	70.5	53.1
Dunn	64.8	74.1	85.1	96.0	71.0	67.8	78.1	87.8	92.9	74.4	66.2	76.0	86.5	94.1	72.7
Barron	43.1	42.7	42.3	54.1	43.3	47.1	44.7	45.4	49.4	46.2	45.1	43.7	43.8	51.3	44.7
Chippewa	40.2	40.5	46.0	56.0	41.7	46.9	44.9	48.4	51.2	46.7	43.3	42.6	47.2	53.1	44.1
All 27 counties	58.9	60.3	63.2	72.6	60.4	63.4	65.0	65.1	67.9	64.4	61.1	62.6	64.2	69.8	62.4

* The percent coverage is calculated as the number of persons included in the REP population on 1 Jan 2022 (numerator; see **Table 1**) divided by the number of persons estimated by the US Intercensal Population Estimates on 1 Jul 2022 (denominator; Vintage 2023).[7]

1.2.1 Diagnosis-based groupings

First, we defined 20 chronic conditions plus anxiety disorders using the International Classification of Diseases, Tenth Revision (**ICD-10**) diagnosis code listings as recommended by the US Department of Health and Human Services (**DHHS**; 21 total conditions).[8] More details about the REP experience using these 21 conditions are reported elsewhere.[9–13] The ICD-10 code listing for the DHHS conditions was updated with the assistance of a coding expert and using the originally published **ICD-9** codes.[8]

Second, we used the Clinical Classifications Software-Refined (**CCSR**) to define diagnosis code groupings comprising 543 categories within 22 chapters. In addition, we constructed 24 additional categories by rolling-up more granular categories (e.g., the 5 categories of BLD001 through BLD005 are all types of anemia, and can be rolled-up into an inclusive category of "*Anemia, All Types*"). In total, we include 567 characteristics based on the CCSR diagnosis code groupings. These CCSR groupings are made available by the Agency for Healthcare Research and Quality (**AHRQ**) as part of the Hospital Cost and Utilization Project (**HCUP**).[14,15] The REP has reported extensively on the previous generation of these code-sets called the Clinical Classifications Software (**CCS**).[3,16]

Lastly, in keeping with previous versions of the DEP, we identified a handful of neurologic conditions that are well characterized by a very small number of ICD-10 codes. For these conditions, we created a REP-defined sub-level grouping within the respective CCSR categories. This list includes 6 conditions: Amyotrophic Lateral Sclerosis, Huntingtons Chorea, Mild Cognitive Impairment, and Restless Legs Syndrome are nested under [*NVS006*]; Alzheimer Disease and Dementia with Lewy Bodies are nested under [*NVS011*].

1.2.2 Procedure and service groupings

We used the CCS to define procedure and service code groupings comprising 244 distinct categories within 17 chapters. These groupings are made available by AHRQ as part of the HCUP (the same as the diagnosis code groupings). The groupings are based on ICD-10 procedure codes,[17] Current Procedural Terminology (**CPT®**) codes, and Healthcare Common Procedure Coding System (**HCPCS**) codes.[18] More details about the 244 CCS-defined procedure and service groups are available elsewhere.[17,18] As of November 2024, AHRQ/HCUP have not released a *refined* version (CCSR) of procedure and service groupings that includes both ICD and CPT/HCPCS codes. We expect that future versions of the DEP may transition to a refined version of CCS code groupings for procedures and services.

1.2.3 Medication groupings

We used the Veterans Administration Drug Classification System (**VA Class**) mappings to define medication groupings. The VA Class mappings consist of unique categories of medications in a nested and hierarchical structure and each category is of the form of a two-character alpha code combined with a three-digit numeric value (e.g., [CV700] = Diuretics). The groupings were created as part of the US Department of Veterans Affairs, Veterans Health Administration as a method of organizing medications into meaningful groups. Specifically, the REP uses various mapping files to convert medications to RxNorm ingredient level information, and then further to VA Class.[19–21]

1.2.4 Personal characteristics

We included 9 personal characteristics gathered from electronic medical record (**EMR**) based clinical systems. These characteristics include self-reported race and ethnicity information. In particular, we include separate race characteristic choices for White, non-White, Black, and Asian races (4 characteristics). The counts in other race groups are too small to include separately. Similarly, one ethnicity characteristic is included for Hispanic ethnicity.

In addition, self-reported smoking habit characteristics are included for ever smokers and current smokers (2 characteristics). Finally, body mass index (**BMI**) is calculated from clinical visit height and weight measurements and used to define two characteristics for BMI ≥ 25 kg/m^2 (overweight and obese) and for BMI ≥ 30 kg/m^2 (obese). BMI cut-offs for persons age 2 through 19 were determined from the 85th (overweight) and 95th (obese) percentiles of the age and sex specific tables from the year 2000 Centers for Disease Control (**CDC**) Growth Charts.[22,23]

2. Data Summaries

2.1 Prevalence

By default, we define a characteristic as *prevalent* if a person received one or more codes from within a characteristic-defining code group in the 5-year period before the prevalence date of 1 Jan 2022 (any code between 1 Jan 2017 and 31 Dec 2021). In addition, the user has the option to change the prevalence time-period from a 5-year window to a 1-year window before prevalence date (1 Jan 2021 through 31 Dec 2021; see **Figure 2**).

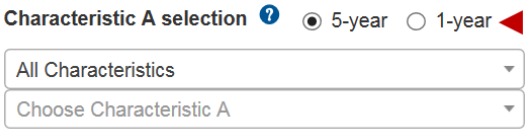


Figure 2. *Choosing the prevalence period for a characteristic. The default period is the 5-year window before 1 Jan 2022. The user can also select a shorter 1-year period (red arrow).*

For each characteristic, prevalence is calculated by dividing the number of persons with a specified characteristic (numerator) by the total number of persons in the population (denominator). Prevalence calculations separately within an age and sex stratum are calculated by including only numerator and denominator persons from within that stratum. The prevalence of the co-occurrence of the two selected characteristics is calculated in the same way.

Users should take note that prevalence scales vary across characteristics. Common characteristics may be reported as *prevalence per 100 persons*, whereas less common characteristics may be reported as *prevalence per 10,000 persons*. It is important that users pay attention to the scale (y-axis label) of displayed prevalence values.

2.2 Observed-to-expected ratios

As a measure of the excess co-occurrence (i.e., clustering) of two characteristics, a ratio is calculated by dividing the number of observed persons with both characteristics by the expected number of persons with both characteristics under the assumption of conditional independence (observed-to-expected ratio; **OER**).[24] We report the OERs separately in sex and age strata and overall.

As a simple example, consider two characteristics in a population of 1,000 persons that occur at frequencies of 10% (characteristic A; *n=100* persons) and 20% (characteristic B; *n=200* persons). Under the assumption of conditional independence, one would expect that the two characteristics would co-occur in 2% ($0.10 \times 0.20 = 0.02$) of the persons (*n=20*). Suppose instead we observe 65 persons with both characteristics A and B. The OER is calculated as the observed number of persons divided by the expected number of persons with both characteristics ($65 \div 20 = 3.25$).

An OER < 1.0 indicates that fewer persons were observed with co-occurring characteristics A and B than would be expected under the assumption of conditional independence. An OER > 1.0 indicates that more persons with co-occurring characteristics A and B were observed than would be expected under the assumption of conditional independence.

For a specific age stratum (e.g., ages 0-20 yrs), the OER is calculated as the sum of the observed persons with both characteristics A and B (the C_i values in **Table 3A**) divided by the sum of the expected persons with both A and B (see **Tables 3B – 3D**). For example, the OER for the age stratum 0 to 20 years is calcu-

lated for i = 0 to 20. Similarly, the overall (all ages) OER is calculated for i = 0 to 99.

Presented here is a general example of how the OER calculations are performed for characteristics A and B. Calculations separately within age, sex, or county strata are performed in an analogous manner. Mathematically, the OER is defined as:

$$OER = \frac{observed}{expected} = \frac{\sum_i C_i}{\sum_i (E_i \cdot N_i)}$$

We determined if the calculated OER differed significantly from 1.0 by calculating the 95% confidence interval. Because the co-occurrence of some combinations of characteristics is quite rare, we calculated the confidence interval for the OER directly from the Poisson distribution. In particular, we calculate the 95% confidence interval around the observed number of persons with both characteristics A and B; the 95% confidence interval of the OER itself is then calculated by dividing each of the confidence interval points by the expected number of persons. The SAS macro for calculating these 95% confidence intervals is given in **Appendix A**.[25]

Table 3A. Observed counts of characteristics.

Age, yrs*	Total	Char A	Char B	Char A & B
0-1	N_0	A_0	B_0	C_0
1-2	N_1	A_1	B_1	C_1
2-3	N_2	A_2	B_2	C_2
...	...	...	...	...
98-99	N_{98}	A_{98}	B_{98}	C_{98}
99+	N_{99}	A_{99}	B_{99}	C_{99}

* Age intervals include the lower value and exclude the upper value. For example, the interval 0-1 includes all persons of ages birth through the day before the first birthday.

Table 3B. Prevalence of characteristics.

Age, yrs	Prev A	Prev B	Prev A & B
0-1	A_0/N_0	B_0/N_0	C_0/N_0
1-2	A_1/N_1	B_1/N_1	C_1/N_1
2-3	A_2/N_2	B_2/N_2	C_2/N_2
...	...	...	...
98-99	A_{98}/N_{98}	B_{98}/N_{98}	C_{98}/N_{98}
99+	A_{99}/N_{99}	B_{99}/N_{99}	C_{99}/N_{99}

Table 3C. Expected prevalence of A & B.

Age, yrs	Prev A	Prev B	E(Prev A & B)*
0-1	A_0/N_0	B_0/N_0	$E_0=(A_0 \cdot B_0)/{N_0}^2$
1-2	A_1/N_1	B_1/N_1	$E_1=(A_1 \cdot B_1)/{N_1}^2$
2-3	A_2/N_2	B_2/N_2	$E_2=(A_2 \cdot B_2)/{N_2}^2$
...	...	...	...
98-99	A_{98}/N_{98}	B_{98}/N_{98}	$E_{98}=(A_{98} \cdot B_{98})/{N_{98}}^2$
99+	A_{99}/N_{99}	B_{99}/N_{99}	$E_{99}=(A_{99} \cdot B_{99})/{N_{99}}^2$

* The expected prevalence of co-occurrence of characteristics A and B under the assumption of conditional independence is calculated as the product of the independent prevalences.

Table 3D. Expected number of persons with A & B.

Age, yrs	Total	E(Prev A & B)	E(No. A & B)
0-1	N_0	$E_0=(A_0 \cdot B_0)/{N_0}^2$	$E_0 \cdot N_0$
1-2	N_1	$E_1=(A_1 \cdot B_1)/{N_1}^2$	$E_1 \cdot N_1$
2-3	N_2	$E_2=(A_2 \cdot B_2)/{N_2}^2$	$E_2 \cdot N_2$
...	...	...	...
98-99	N_{98}	$E_{98}=(A_{98} \cdot B_{98})/{N_{98}}^2$	$E_{98} \cdot N_{98}$
99+	N_{99}	$E_{99}=(A_{99} \cdot B_{99})/{N_{99}}^2$	$E_{99} \cdot N_{99}$

2.3 Standardization

For the county-specific calculations displayed on the map, prevalence and OER values are directly standardized by age (and sex when appropriate) to the US total population from the 2020 US Decennial Census (see **Appendix B**).[7] This standardization process makes prevalence and OER values comparable across counties while accounting for the differences in age and sex make-up. See section **3.4** for more information on prevalence and OER reporting by county.

2.4 Privacy

We take the privacy of persons included in the REP seriously. Accordingly, we make efforts to maintain anonymity and to protect the confidentiality of persons. We have taken several steps to ensure this continued trust.

Firstly, we only include information for persons who have given permission for their medical record information to be used for research

purposes.[6] Secondly, we present all information in aggregate summary form. No specific numerators are reported for prevalence and no observed counts are given for OER calculations. Relationships between the two user-selected characteristics are aggregated and summarized in graphical and table form. Lastly, we only report prevalence and OER values when more than 10 persons are in a stratum defined by age, sex, and county.

3. User Interface

3.1 Choosing Characteristics

Users may select characteristics by performing a free text search of the descriptions from a subset of the characteristics (e.g., only medications), or from the full characteristic list. To search by subsetting to a type of a characteristic first, click on the box labeled *All Characteristics* and choose the type of characteristic of interest from the drop down list (see **Figure 3**). After a subset of characteristics is chosen (or from the complete list of all characteristics), click on the box labeled *Choose Characteristic A*, and start typing the text of the characteristic of interest. The selection list will narrow to include only items matching the search criteria text (see **Figure 4**). Characteristic B is selected in the same way.

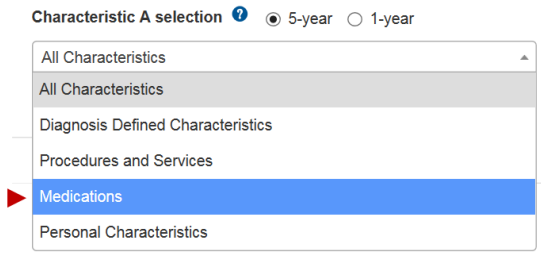


Figure 3. *Users can subset to the type of characteristic of interest by selecting the subset category from the top drop-down list (e.g., Medications; see red arrow).*

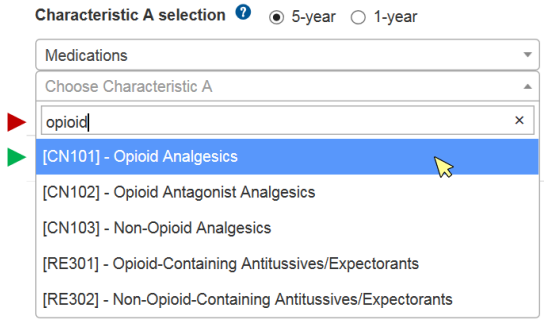


Figure 4. *After beginning to type the characteristic name (red arrow), the selection list updates with matching characteristics, and the user selects the item of interest (green arrow).*

Once a user has selected a characteristic of interest, the full list of codes used to define that characteristic is available to the user by following the *List of codes* link below the selection box (see **Figure 5**).

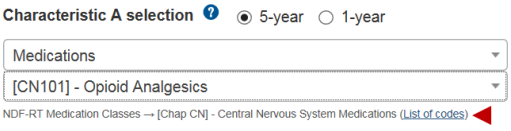


Figure 5. *Clicking on the List of codes link (red arrow) will open a separate window with the full listing of codes used to define the chosen characteristic.*

Some characteristics may not be named using lay-person terms or may be described differently than is standard practice within a given specialty. When users are unable to locate a characteristic of interest, we encourage a search on several varying names or descriptions of the characteristic. For example, a search for "Gehrig" will not return any results; however, a search for "amyotrophic" will return the condition Amyotrophic Lateral Sclerosis (ALS; also known as "Lou Gehrig Disease").

3.2 Data Summary Views

Users may choose from three tabs which provide different views of summary information. The first tab (called *Prevalence*; see section **3.3**) summarizes characteristics and their co-occurrence for the combined 27-county population overall. The second tab (called *Geography*; see section **3.4**) summarizes characteristics and

their co-occurrence directly standardized (by age and sex) to the 2020 US Census population and displayed separately for each of the 27 counties via a color shaded map. The third tab (called *Trends*; see section **3.5**) summarizes the prevalence of characteristics in the 27-county populations on January 1st for years 2017 through 2022. **Figure 6** shows the tabs where users select the data summary view of interest within the DEP.

Figure 6. *The tabs where a user selects the view of interest.*

3.3 The Prevalence Tab

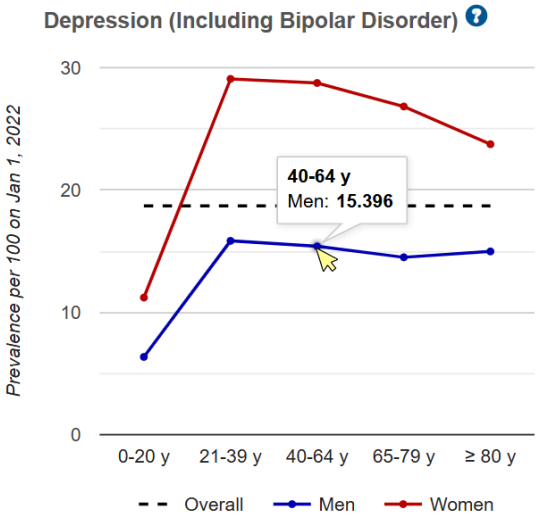


Figure 7. *The prevalence per 100 persons for depression as defined by the DHHS diagnosis code grouping.*

The prevalence of each chosen characteristic is summarized as a line graph for men (blue line) and women (red line) and across 5 age groups (0-20 y, 21-39 y, 40-64 y, 65-79 y, and $\geq$ 80 y). Hovering over points on the graph will display the prevalence value. The overall pooled prevalence (men and women combined for all ages) is also displayed on the graph as a horizontal dashed black line. As an example, **Figure 7** shows the prevalence graph for depression. Note that the scale of the *y*-axis may vary because some characteristics are less common than others. The prevalence of the co-occurrence of both characteristics is also displayed as a line graph.

A table summarizing OER values separately in sex and age strata is also reported (see **Figure 8**). OERs are not calculated if fewer than 11 persons with both characteristics are observed within a specific stratum (shown as "$n < 11$" in a table cell; see Age Group 0-20 y for Women in **Figure 8**). Similarly, for characteristics applicable to only men (e.g., *cancer of prostate*) or women (e.g., *oophorectomy*), "N/A" is reported in the OER table for the unaffected sex and Total columns (see columns for Men and Total in **Figure 8**).

The OER table is constructed to draw attention to OER values reaching conventional statistical significance ($p < 0.05$). OER values significantly different from 1.0 are denoted with an asterisk (*) and are shaded in orange (OER > 1.0; more co-occurrence of the two characteristics than expected) or purple (OER < 1.0; less co-occurrence than expected) using a color palette from ColorBrewer 2.0®.[26] Note that if an OER is not statistically different from 1.0, the value is still displayed in the table but is not color shaded or denoted with an asterisk.

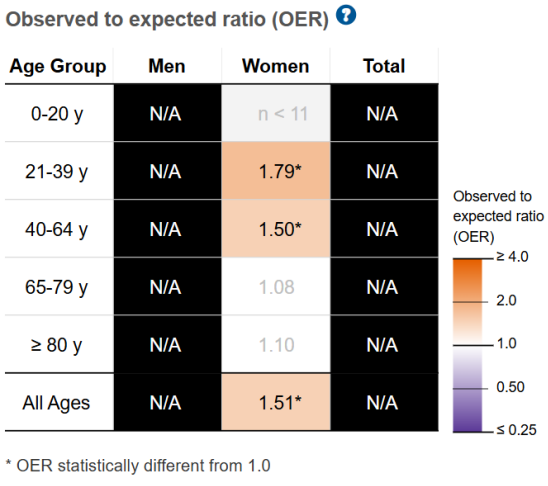


Figure 8. *An example of the OER table for the co-occurrence of the two characteristics* depression *and* oophorectomy.

3.4 The Geography Tab

Data are presented by geography in a visual format via shading of a map of the 27-county region of the REP. The user chooses whether to shade the counties by prevalence of individual characteristics, prevalence of both characteristics, or by OERs for the co-occurrence of two selected characteristics. The user can also choose to show data overall or separately for one sex.

Prevalence and OER values are directly standardized by age (and sex when appropriate) to the total 2020 US Census population for each county; this allows for comparison between counties while accounting for differences in age and sex distributions.[7] **Figure 9** shows an example of a prevalence shaded map. Similarly, **Figure 10** shows an example of an OER shaded map. When hovering over a specific county with the mouse, the numerical value of the prevalence or OER is displayed (see Barron Co., WI in **Figures 9 and 10**).

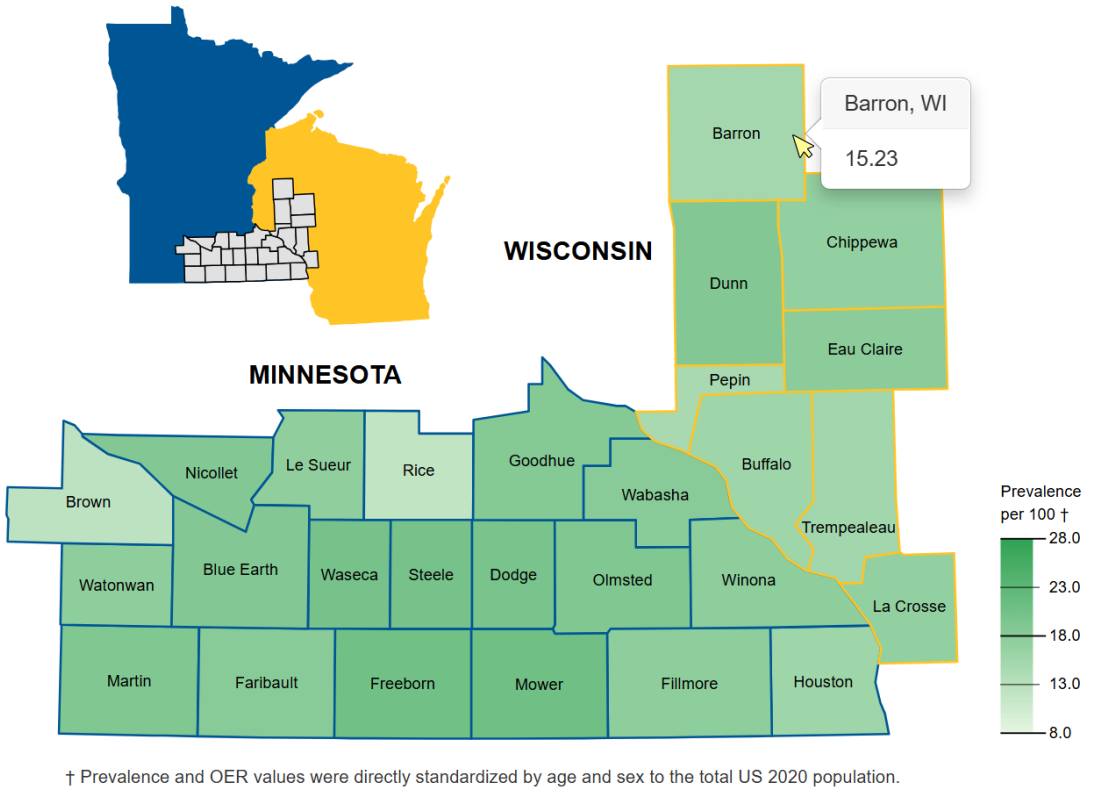


Figure 9. *The 5-year period prevalence shaded map of the 27-county region for* depression *in men and women combined. Note that the prevalence value for Barron County, Wisconsin is displayed when hovering over the county with the mouse*

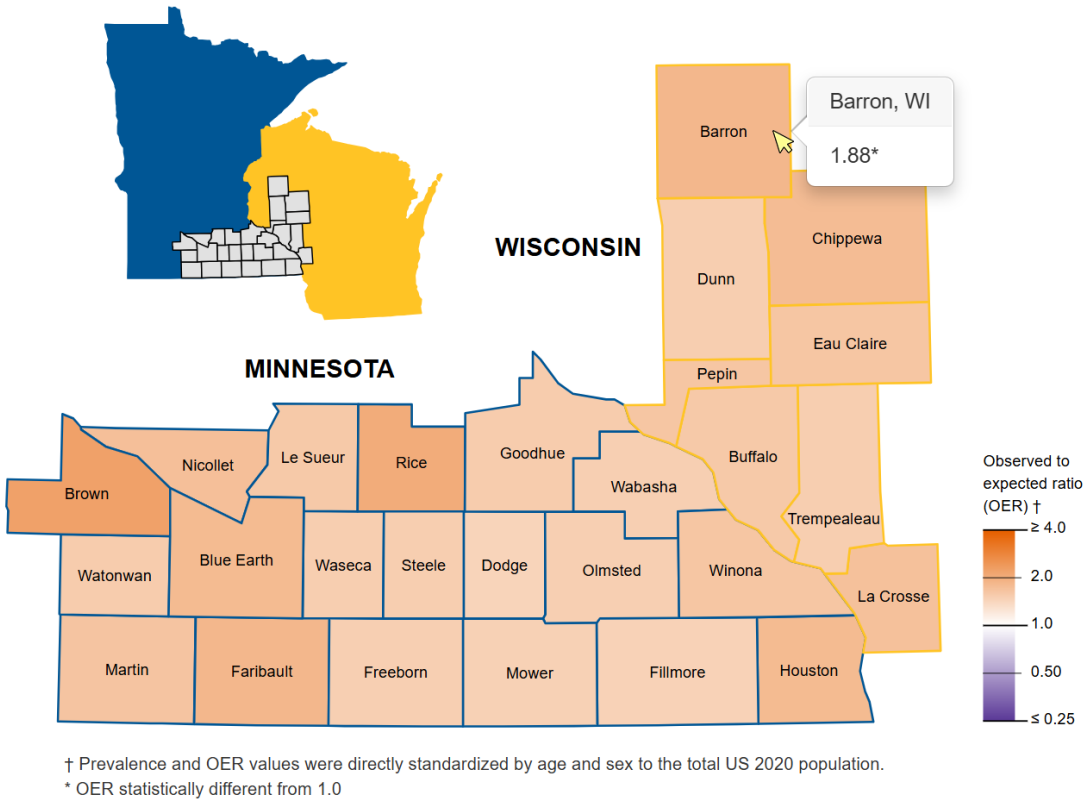


Figure 10. *The OER shaded map of the 27-county region for the co-occurrence of* depression *(5-year period prevalence) and* diabetes *(5-year period prevalence) in men and women combined. Note that the OER value for Barron County, Wisconsin is displayed when hovering over the county with the mouse.*

3.5 The Trends Tab

The *Prevalence* and *Geography* tabs only include summary data for the 27-county population identified in the most recent prevalence sample (e.g., 1 Jan 2022 in DEP version 7.0). However, it may be beneficial for users to explore whether the prevalence of a characteristic is changing over time (calendar year). For this reason, we developed the *Trends* view.

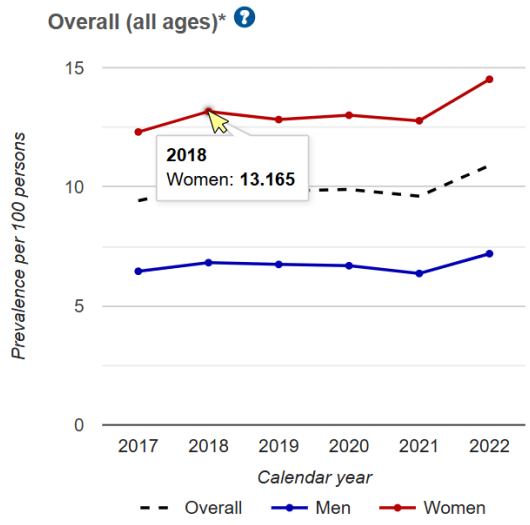
This tab displays prevalence for the 27-county population over the 6 most recent calendar years separately for characteristics A and B (see **Table 4** for the population size over these 6 years). Prevalence trends are displayed in 6 separate panels: (1) One panel plots the prevalence of a characteristic for all ages combined and directly standardized by age and sex to the 2020 US Census population (see **Figure 11**), and (2) five panels display prevalence separately for each of the broad age groups (non-standardized, raw prevalence). Separate lines are shown for men (blue) and women (red), and for both sexes combined (dashed black). **Figure 11** shows an example of a trend plot for *depression* across years 2017 to 2022 for all ages.

Of particular note, users should be aware that only prevalence in the 1-year period before January 1st of each year are shown. This is because many persons in consecutive year prevalence populations overlap; thus, 5-year period prevalence would not be independent. Even when a user selects the 5-year period radial button for a characteristic, *only the 1-year*

prevalence is shown in the Trends tab. Just as before, the prevalence scale (i.e., the *y*-axis label) may vary across the panels. Similarly, no trend lines are shown fwhen fewer than 11 persons have a characteristic in an age- and sex-specific stratum. If a characteristic is too uncommon for both men and women, then the respective panel is shaded gray with no trend lines.

Table 4. Persons included in trends populations.

	Age groups, population on January 1st					
Sex/year	0-20 y	21-39 y	40-64 y	65-79 y	≥ 80 y	All ages
Men						
2017	96,903	78,857	107,851	42,997	14,282	340,890
2018	96,840	78,668	106,773	44,593	14,267	341,141
2019	96,146	78,387	105,416	45,911	14,306	340,166
2020	98,494	84,070	108,331	47,911	14,452	353,258
2021	100,653	88,638	111,940	50,431	14,839	366,501
2022	97,370	83,965	108,958	52,154	15,189	357,636
Women						
2017	94,473	89,301	112,690	47,428	21,836	365,728
2018	94,292	88,906	111,514	48,896	21,757	365,365
2019	93,111	87,669	109,030	49,968	21,338	361,116
2020	96,093	91,778	111,035	51,673	21,298	371,877
2021	98,663	95,239	114,701	54,085	21,471	384,159
2022	95,865	92,671	112,426	56,008	21,715	378,685
All persons						
2017	191,376	168,158	220,541	90,425	36,118	706,618
2018	191,132	167,574	218,287	93,489	36,024	706,506
2019	189,257	166,056	214,446	95,879	35,644	701,282
2020	194,587	175,848	219,366	99,584	35,750	725,135
2021	199,316	183,877	226,641	104,516	36,310	750,660
2022	193,235	176,636	221,384	108,162	36,904	736,321



* Prevalence is standardized by age and sex to the total US 2020 population.

Figure 11. *The 1-year period prevalence per 100 persons of* depression *across the 27-county populations on January 1st of 2017 through 2022, separately by sex.*

Interpretation of trend data *should be undertaken with caution*. There are substantial changes that have taken place during the years 2017 to 2022 that are unrelated to medical characteristics. For example, the electronic medical record systems used at REP participating providers for Mayo Clinic and Olmsted Medical Center transitioned from Cerner or other legacy EMR systems to EPIC-based EMRs during the years from 2017 through 2018. Such fundamental changes in the source data may introduce spurious "trends" in summary-level data.

We also note that prevalence decreases for many diagnosis and procedure code-based categories for the plotted year point of 2021 (all diagnoses and procedure codes from calendar year 2020). This is likely due to the world-wide COVID-19 pandemic and the delay of some procedures and the decrease in general medical visits during lock-down.

4. Registration and Login

4.1 Why Register?

Users of the Data Exploration Portal have the option to create a login account (registration). The primary benefits of creating a personal login account are:

1. The ability to **download data from tables and figures** as a comma-separated values (**CSV**) file.
2. The ability to **save searches** for user-selected combinations of characteristics.
3. The ability to receive **email updates** when new features are added to the DEP (users must opt in to receive these emails).

4.2 Creating an Account

Creating a login account is easy. From the DEP home page, users can click on *My Account* on the top navigation bar, and then click on *Login* (see **Figure 12**). The registration form requires

a first name, last name, an email address (your login username), and a password. At the time of registration, users must also agree to the Terms and Conditions for using the DEP (see **Appendix C**) and can decide whether they wish to receive email updates related to the DEP.

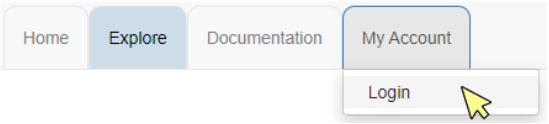


Figure 12. *The* My Account *navigation bar used to create a new account or to log in to an existing account.*

4.3 Downloading Data

Once a login account has been created, users have the ability to download a CSV file of the underlying data for two selected characteristics. These data are the tabular form of the data displayed graphically in line charts (prevalence separate by age, sex, and calendar year), color-shaded tables (OERs), and maps (both prevalence and OER values, separately by county and sex). The link *Download data in CSV format* is located at the lower left corner of the web page after two characteristics have been selected.

4.4 Saving Searches

Users are able to save a search for a specific combination of characteristics to their personal DEP account (see *Save Search* link in **Figure 13**). This may be useful so that once a combination of two characteristics of interest is found, the user does not have to recreate the search. The capability to save searches may be of particular convenience for users who have found potential hypothesis generating associations between two characteristics and wish to save these for further investigation. Once a search is saved, it can be found in a user's profile page under the *My Account* tab at the top of the webpage (see **Figure 12**). Note that specific user-selected searches can also be sent via email using the *Email Search* link (see **Figure 13**). Emailing search links does not require a registered DEP-user account. Unfortunately, upon release of the DEP v. 6.0 in November 2023, we were forced to clear all user-saved searches because of the transition to CCSR groupings. There is no direct mapping from the previous diagnosis based CCS groupings (DEP versions 1.0 through 5.0) to the newer refined CCSR mappings.

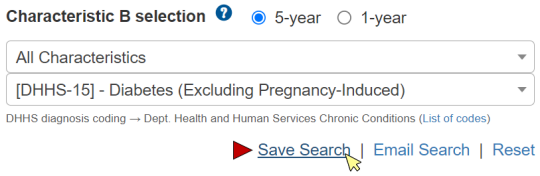


Figure 13. *The* Save Search *and* Email Search *options are located under the Characteristic B selection box (red arrow).*

4.5 DEP Update Emails

At the time of registration, users may decide whether they wish to receive emails announcing future developments and improvements to the Data Exploration Portal. User-provided email addresses are not used for messages unrelated to the DEP and the email list is not sold or shared with non-REP entities. DEP users can expect to receive 1 or 2 email update messages per year, and email preferences can be changed at any time in the *My Account* settings.

Appendix A. SAS Macro for exact Poisson confidence limits.

```
######################################################################
#
#  This SAS macro calculates the exact confidence limits for
#  a Poisson count.
#
#  Algorithm:  Daly, Leslie.  Simple SAS macros for the
#                 calculation of exact binomial and Poisson
#                 confidence limits.
#                 Comput Biol Med.  1992;22(5):351-61.
#
#  Functions called:  LOG, GAMINV
#
#  Usage:  %CIPOISS(cl, x, ll, lu)
#            Input parameters:
#                   cl - confidence level proportion (e.g.  0.95)
#                   x - observed number of events
#            Output parameters:
#                   ll - Lower confidence limit
#                   lu - Upper confidence limit
#
#  Usage notes:  Missing values for the output parameters
#                   are generated if:
#                   (1) X is less than zero, or if
#                   (2) CL is not between 0.0 and 1.0.
######################################################################

%macro cipoiss(cl, x, ll, lu);

if (&x) lt 0 or not(0 lt (&cl) lt 1) then do;
   &ll = .;
   &lu = .;
   end;
else do;
   if (&x) ne 0 then do;
      &ll = gaminv((1 - (&cl))/2, (&x));
      &lu = gaminv((1 + (&cl))/2, (&x) + 1);
      end;
   else if (&x) eq 0 then do;
      &lu = -log(1 - (&cl));
      &ll = 0;
      end;
   end;

%mend;
```

Appendix B. Total 2020 US Census reference population.[7]

	Population, N				Population, N		
Age, yrs	Men	Women	Total	Age, yrs	Men	Women	Total
0 - 1	1,907,982	1,825,191	3,733,173	51 - 52	2,073,714	2,072,770	4,146,484
1 - 2	1,928,926	1,846,660	3,775,586	52 - 53	2,025,205	2,031,284	4,056,489
2 - 3	1,980,392	1,891,347	3,871,739	53 - 54	2,043,490	2,065,387	4,108,877
3 - 4	2,028,781	1,935,784	3,964,565	54 - 55	2,075,610	2,104,337	4,179,947
4 - 5	2,068,682	1,978,806	4,047,488	55 - 56	2,155,586	2,206,290	4,361,876
5 - 6	2,081,588	1,994,479	4,076,067	56 - 57	2,189,190	2,241,576	4,430,766
6 - 7	2,072,810	1,973,925	4,046,735	57 - 58	2,179,978	2,246,798	4,426,776
7 - 8	2,069,511	1,982,754	4,052,265	58 - 59	2,195,068	2,265,391	4,460,459
8 - 9	2,086,029	1,985,329	4,071,358	59 - 60	2,172,140	2,255,058	4,427,198
9 - 10	2,109,096	2,003,206	4,112,302	60 - 61	2,133,820	2,218,734	4,352,554
10 - 11	2,156,295	2,053,910	4,210,205	61 - 62	2,089,447	2,194,509	4,283,956
11 - 12	2,211,556	2,107,809	4,319,365	62 - 63	2,071,968	2,185,703	4,257,671
12 - 13	2,279,954	2,170,544	4,450,498	63 - 64	1,993,341	2,120,834	4,114,175
13 - 14	2,267,973	2,160,853	4,428,826	64 - 65	1,920,791	2,068,477	3,989,268
14 - 15	2,214,702	2,114,888	4,329,590	65 - 66	1,862,194	2,019,270	3,881,464
15 - 16	2,195,896	2,097,472	4,293,368	66 - 67	1,776,409	1,946,760	3,723,169
16 - 17	2,212,634	2,117,723	4,330,357	67 - 68	1,705,389	1,885,107	3,590,496
17 - 18	2,183,358	2,088,367	4,271,725	68 - 69	1,643,084	1,826,890	3,469,974
18 - 19	2,192,682	2,099,362	4,292,044	69 - 70	1,519,408	1,714,706	3,234,114
19 - 20	2,226,508	2,132,951	4,359,459	70 - 71	1,473,667	1,668,891	3,142,558
20 - 21	2,204,976	2,111,958	4,316,934	71 - 72	1,428,482	1,625,457	3,053,939
21 - 22	2,185,108	2,098,588	4,283,696	72 - 73	1,428,101	1,624,945	3,053,046
22 - 23	2,171,094	2,088,208	4,259,302	73 - 74	1,427,302	1,605,107	3,032,409
23 - 24	2,178,603	2,100,104	4,278,707	74 - 75	994,054	1,171,117	2,165,171
24 - 25	2,207,336	2,122,545	4,329,881	75 - 76	978,108	1,157,187	2,135,295
25 - 26	2,238,236	2,159,712	4,397,948	76 - 77	954,152	1,143,458	2,097,610
26 - 27	2,285,714	2,206,197	4,491,911	77 - 78	928,222	1,125,382	2,053,604
27 - 28	2,328,006	2,252,294	4,580,300	78 - 79	785,137	971,983	1,757,120
28 - 29	2,365,648	2,308,699	4,674,347	79 - 80	695,899	877,237	1,573,136
29 - 30	2,403,221	2,345,659	4,748,880	80 - 81	625,445	808,302	1,433,747
30 - 31	2,381,929	2,330,762	4,712,691	81 - 82	582,704	763,655	1,346,359
31 - 32	2,326,475	2,274,884	4,601,359	82 - 83	527,239	707,254	1,234,493
32 - 33	2,287,877	2,235,990	4,523,867	83 - 84	461,784	636,588	1,098,372
33 - 34	2,270,075	2,211,100	4,481,175	84 - 85	420,219	595,022	1,015,241
34 - 35	2,283,203	2,221,117	4,504,320	85 - 86	353,370	531,827	885,197
35 - 36	2,257,491	2,205,937	4,463,428	86 - 87	314,703	487,916	802,619
36 - 37	2,233,065	2,185,501	4,418,566	87 - 88	260,977	419,520	680,497
37 - 38	2,261,913	2,218,258	4,480,171	88 - 89	233,027	390,809	623,836
38 - 39	2,237,770	2,198,148	4,435,918	89 - 90	203,429	350,919	554,348
39 - 40	2,223,666	2,189,021	4,412,687	90 - 91	175,423	318,266	493,689
40 - 41	2,182,540	2,153,190	4,335,730	91 - 92	147,744	279,258	427,002
41 - 42	2,095,733	2,078,051	4,173,784	92 - 93	117,520	233,133	350,653
42 - 43	2,062,666	2,052,806	4,115,472	93 - 94	96,380	200,853	297,233
43 - 44	2,005,945	2,003,387	4,009,332	94 - 95	75,546	164,269	239,815
44 - 45	1,966,990	1,966,469	3,933,459	95 - 96	56,845	132,299	189,144
45 - 46	1,977,202	1,976,212	3,953,414	96 - 97	43,169	105,435	148,604
46 - 47	1,941,478	1,945,479	3,886,957	97 - 98	31,452	79,773	111,225
47 - 48	1,984,563	1,991,172	3,975,735	98 - 99	21,712	57,655	79,367
48 - 49	2,095,666	2,093,677	4,189,343	99 +	40,537	121,545	162,082
49 - 50	2,220,803	2,209,702	4,430,505				
50 - 51	2,143,996	2,135,626	4,279,622	**All ages**	**164,192,524**	**167,256,756**	**331,449,280**

* Age intervals include the lower value and exclude the upper value. For example, the interval 0-1 includes all persons of ages birth through the day before the first birthday.

Appendix C. Terms and Conditions of use for the Rochester Epidemiology Project Data Exploration Portal.

Terms and conditions of use

Welcome
This is the website of the Rochester Epidemiology Project (REP) which is a collaboration between Mayo Clinic, Mayo Clinic Health System, Olmsted Medical Center, and Olmsted County Public Health Services ("REP", "REP Partners", "We", or "Us"). This website is an online information and communication service provided by the REP, and this policy is intended to cover the uses for the Site, although additional conditions, restrictions, and privacy policies may apply.

Terms of use agreement
Use of the Site is subject to the following terms of use. The REP partners may modify these terms and conditions at any time, and such modifications shall be effective immediately upon posting the modified terms and conditions on the Site. You agree to review the agreement periodically to be aware of such modifications, and your access or use of the service constitutes your acceptance of the agreement as it appears at the time of your access or use. From time to time, the REP may offer website visitors the opportunity to participate in additional features or services related to the site. You may be required to enter into additional agreements or authorizations before you can access such features or services.

Termination of use
The REP Partners, in their sole discretion, may terminate your account or your use of the Site at any time. The REP Partners reserve the right to change, suspend or discontinue all or any aspects of the Site at any time without prior notice.

Indemnification
You agree that you will indemnify the REP Partners against any damages, losses, liabilities, judgments, costs or expenses (including reasonable attorneys' fees and costs) arising out of a claim by a third party relating to materials you have posted or other actions taken by you on the Site.

Intellectual property
The marks "REP", "Rochester Epidemiology Project", and the REP logo are trademarks and/or service marks of the REP Partners. The content and design of the REP Site is protected by U.S. and international copyright laws. You may not copy, reproduce, republish, upload, post, display, transmit or frame any of these materials without prior written consent from the copyright owners, except that you may view, download, display and print a single copy of these materials on a single computer for personal, noncommercial use only, so long as: you do not alter or modify the materials in any way; you include all applicable notices and disclaimers (including copyright notices); and you do not use the materials in a way that suggests an association with REP or an affiliated REP entity. You understand and agree that title to these materials shall not pass to you or any other user.

Disclaimers
YOU UNDERSTAND AND AGREE THAT the REP SITE AND ANY SERVICES, CONTENT OR INFORMATION CONTAINED ON OR PROVIDED BY the REP SITE IS PROVIDED ON AN "AS IS" BASIS. THE REP PARTNERS DO NOT MAKE ANY EXPRESS OR IMPLIED WARRANTIES, REPRESENTATIONS OR ENDORSEMENTS WHATSOEVER (INCLUDING WITHOUT LIMITATION WARRANTIES OF TITLE OR NONINFRINGEMENT, OR THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE) WITH REGARD TO THE SERVICE OR ANY MERCHANDISE, INFORMATION OR SERVICE PROVIDED THROUGH THE SERVICE. IN ADDITION, THE REP PARTNERS DO NOT GUARANTEE THAT USE OF THE REP SITE WILL BE FREE FROM TECHNOLOGICAL DIFFICULTIES INCLUDING, BUT NOT LIMITED TO, UNAVAILABILITY OF INFORMATION, DOWNTIME, SERVICE DISRUPTIONS, VIRUSES OR WORMS, AND YOU UNDERSTAND THAT YOU ARE RESPONSIBLE FOR IMPLEMENTING SUFFICIENT PROCEDURES AND CHECKPOINTS TO SATISFY YOUR PARTICULAR REQUIREMENTS FOR ACCURACY OF DATA INPUT AND OUTPUT.

Errors on the REP Site
Content and availability of services are subject to change without notice. Errors will be corrected where discovered, and the REP Partners reserve the right to revoke any stated offer and to correct any errors, inaccuracies, or omissions.

Limitations of liability
THE USE OF THE SITE IS ENTIRELY AT YOUR OWN RISK, AND IN NO EVENT SHALL THE REP PARTNERS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, CONSEQUENTIAL, SPECIAL, EXEMPLARY, PUNITIVE, OR ANY OTHER MONETARY OR OTHER DAMAGES, FEES, FINES, PENALTIES, OR LIABILITIES ARISING OUT OF OR RELATING IN ANY WAY TO THIS SERVICE, OR SITES ACCESSED THROUGH THIS SERVICE, AND/OR CONTENT OR INFORMATION PROVIDED HEREIN. YOUR SOLE AND EXCLUSIVE REMEDY FOR DISSATISFACTION WITH THE SERVICE IS TO STOP USING THE SERVICE. USER HEREBY ACKNOWLEDGES THAT THIS PARAGRAPH SHALL APPLY TO ALL CONTENT, AND SERVICES AVAILABLE THROUGH THE SITE. SOME STATES DO NOT ALLOW THE EXCLUSION OR LIMITATION OF INCIDENTAL OR CONSEQUENTIAL DAMAGES, SO THE ABOVE LIMITATION OR EXCLUSION MAY NOT APPLY TO YOU.

Links to third party websites
The Site may contain links to other sites operated by third parties, including but not limited to third party sites affiliated with the REP Partners. These links are available for your convenience and are intended only to enable access to these Third Party Sites and for no other purpose. The REP Partners do not warrant or make any representation about the substance, quality, functionality, accuracy, fitness for a particular purpose, merchantability or any other representation about any Third Party Site or its content. A link to a Third Party Site on the REP Site does not constitute sponsorship, endorsement, approval, or responsibility for any Third Party Site. The REP Partners make no representation or warranty as to any products or services offered on any Third Party Site. The conditions of use and privacy policy of any Third Party Site may differ substantially from the conditions of use and legal notices that apply to your use of the Site. Please review the conditions of use for all Third Party Sites for more information about the terms and conditions that apply to your use of Third Party Sites.

Contact information
We encourage you to contact us regarding any questions, comments, suggestions, or other feedback you have about the Web site. Please contact us at info@rochesterproject.org.

Privacy Policy

Web privacy policy
We take your privacy seriously, and we want you to know how we collect, use, share, and protect your information. In addition to this privacy policy, users of the REP Site should consult the REP Site terms of use as well as any product specific terms and conditions that apply.

What information we collect
We respect the right to privacy of all visitors to the REP Site. We collect information necessary to maintain an account (first name, last name, institution, city, state, and e-mail address). The personal information that you submit, may be used for purposes related to our research and educational operations or to address any inappropriate use of our website. We do not save personal information to use for other purposes, nor do we provide it to any other organizations.

Email communications and related services
We may use the e-mail address provided to send you surveys which you may or may not choose to participate in. In addition, e-mail communications that you send to us via the email links on our site may be shared with the REP staff person who is most able to address your inquiry. We make every effort to respond in a timely fashion once communications are received. Once we have responded to your communication, it is discarded or archived, depending on the nature of the inquiry.

The email functionality on our site does not provide a completely secure and confidential means of communication. It is possible that your email communication may be accessed or viewed by another Internet user while in transit to us. If you wish to keep your communication private, do not use our email. You may decide at some point that you no longer wish to receive communications from our site. To stop receiving communications, change the settings within your account profile or send an email message to info@rochesterproject.org.

IP addresses
The Web server automatically collects the Internet Protocol (IP) address of the computers that access our site. An IP address is a number that is assigned to your computer when you access the Internet. It is not truly personally identifiable information because many different individuals can access the Internet via the same computer. We use this information in aggregate form to understand how our site is being used and how we can better serve visitors.

Cookies and other tracking technology
We collect information about visitors to our Site using "cookies" (see definition on Wikipedia) and similar technology such as Web beacons, Web bugs, pixel tags, and so on. We use this technology to recognize a repeat visitor and offer the visitor a set of services or information requested in a previous visit. We use session cookies to track a visitor's path through our site during a visit, to help us understand how people use our site.

How we use the information we collect
We use the information we collect for things like, but not limited to:
- Conducting analyses
- Communicating things like special events and surveys
- Establishing and managing your accounts with us
- Operating, evaluating, and improving our web site

Data retention
We will retain your information for as long as your account is active or as needed to provide you services, comply with our legal obligations, resolve disputes, and enforce our agreements. Except for authorized law enforcement investigations or other valid legal processes, we will not share any personally identifiable information we receive from you with any parties outside of the REP Partners.

We may share some information with third parties
We may share your personally identifiable information with third parties who we have engaged to help us provide the services. In each case, we will ensure that these third parties have agreed not to use or disclose your personal information except to help us provide the services.

We may provide third parties with aggregate statistics about our visitors, traffic patterns, and related site information. These data reflect site-usage patterns gathered during visits to our website each month, but they do not contain behavioral or identifying information about any individual member unless that member has given us permission to share that information.

To help us determine the effectiveness of the REP Site, we work with Web analytics tools hosted by third parties who receive non-identifiable information from your browser, including but not limited to the site or the advertisement you came from, your IP address, your general geographic location, your browser and platform information, and the pages you view within our Site.

Privacy policy updates
We may need to update our privacy policy as technology changes and the REP evolves. If we make significant changes to the privacy policy, we'll post a prominent message on our website.

Contact information:
Rochester Epidemiology Project
Attn: Administrative Manager
PO Box 115
Rochester, MN 55903
Toll free: (877) 276-6654
Email: info@rochesterproject.org

This policy was last updated November 1, 2016.

References

1. Rocca WA, Yawn BP, St. Sauver JL, Grossardt BR, Melton LJ 3rd. History of the Rochester Epidemiology Project: half a century of medical records linkage in a US population. *Mayo Clin Proc.* 2012;87(12):1202-13.

2. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ 3rd, Pankratz JJ, Brue SM, Rocca WA. Data Resource Profile: The Rochester Epidemiology Project (REP) medical records-linkage system. *Int J Epidemiol.* 2012;41(6):1614-24.

3. St Sauver JL, Grossardt BR, Finney Rutten LJ, Roger VL, Majerus M, Jensen DW, Brue SM, Bock-Goodner CM, Rocca WA. Rochester Epidemiology Project Data Exploration Portal. *Prev Chronic Dis.* 2018;15:E42.

4. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ 3rd, 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(9):1059-68.

5. Rocca WA, Grossardt BR, Brue SM, Bock-Goodner CM, Chamberlain AM, Wilson PM, Finney Rutten LJ, St Sauver JL. Data Resource Profile: Expansion of the Rochester Epidemiology Project medical records-linkage system (E-REP) *Int J Epidemiol.* 2018;47(2);368-368j.

6. Citizens' Council on Health Care What Consent? - MN's Medical & Genetic Research Law. Accessed October 2016.

7. U.S. Census Bureau, Population Division. Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2023 (CC-EST2023-ALLDATA; Vintage 2023, released June 2024). Accessed August 2024.

8. Goodman RA, Posner SF, Huang ES, Parekh AK, Koh HK. Defining and Measuring Chronic Conditions: Imperatives for Research, Policy, Program, and Practice. *Prev Chronic Dis.* 2013;10:120239.

9. Rocca WA, Boyd CM, Grossardt BR, Bobo WV, Finney Rutten LJ, Roger VL, Ebbert JO, Therneau TM, Yawn BP, St Sauver JL. Prevalence of multimorbidity in a geographically defined American population: patterns by age, sex, and race/ethnicity. *Mayo Clin Proc.* 2014;89(10):1336-49.

10. St Sauver JL, Boyd CM, Grossardt BR, Bobo WV, Finney Rutten LJ, Roger VL, Ebbert JO, Therneau TM, Yawn BP, Rocca WA. Risk of developing multimorbidity across all ages in an historical cohort study: differences by sex and ethnicity. *BMJ Open.* 2015;5(2):e006413.

11. Bobo WV, Yawn BP, St Sauver JL, Grossardt BR, Boyd CM, Rocca WA. Prevalence of Combined Somatic and Mental Health Multimorbidity: Patterns by Age, Sex, and Race/Ethnicity. *J Gerontol A Biol Sci Med Sci.* 2016;71(11):1483-1491.

12. St Sauver JL, Chamberlain AM, Bobo WV, Boyd CM, Finney Rutten LJ, Jacobson DJ, McGree ME, Grossardt BR, Rocca WA. Implementing the US Department of Health and Human Services definition of multimorbidity: a comparison between billing codes and medical record review in a population-based sample of persons 40-84 years old. *BMJ Open.* 2021;11(4):e042870.

13. Bobo WV, Grossardt BR, Virani S, St Sauver JL, Boyd CM, Rocca WA. Association of Depression and Anxiety With the Accumulation of Chronic Conditions. *JAMA Netw Open.* 2022;5(5):e229817.

14. Healthcare Cost and Utilization Project (HCUP) supported by the Agency for Healthcare Research and Quality (AHRQ). Clinical Classifications Software Refined (CCSR) for ICD-10-CM, Version 2023-1 (FY2022). Accessed March 2023.

15. Healthcare Cost and Utilization Project (HCUP) supported by the Agency for Healthcare Research and Quality (AHRQ). Clinical Classifications Software Refined (CCSR): User Guide. Accessed March 2023.

16. St Sauver JL, Warner DO, Yawn BP, Jacobson DJ, McGree ME, Pankratz JJ, Melton LJ 3rd, Roger VL, Ebbert JO, Rocca WA. Why patients visit their doctors: assessing the most prevalent conditions in a defined American population. *Mayo Clin Proc.* 2013;88(1):56-67.

17. Healthcare Cost and Utilization Project (HCUP) supported by the Agency for Healthcare Research and Quality (AHRQ). Beta Clinical Classifications Software (CCS) for ICD-10-CM/PCS, Version 2020-2 (FY2019). Accessed June 2023.

18. Healthcare Cost and Utilization Project (HCUP) supported by the Agency for Healthcare Research and Quality (AHRQ). Clinical Classifications Software (CCS) for Services and Procedures, Version 2022-1 (FY 2021). Accessed June 2023

19. US National Library of Medicine. National Drug File - Reference Terminology Source Information. Accessed August 2017

20. US National Library of Medicine. RxNav: Navigating RxNorm Drugs. Accessed September 2017

21. Veterans Administrations National Formulary. Pharmacy Benefits Management Services. Accessed September 2021

22. Centers for Disease Control and Prevention. Defining Childhood Obesity: BMI for Children and Teens. Accessed August 2017

23. Centers for Disease Control and Prevention. Clinical Growth Charts: Boys and Girls BMI-for-age. Accessed August 2017

24. Gordon C. 3319.0.55.001 - Multiple Cause of Death Analysis, 1997-2001: Appendix V. Method of calculating observed to expected ratio as a measure of strength of association. Australian Bureau of Statistics. Accessed April 2017.

25. Daly L. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. *Comput Biol Med.* 1992;22(5):351-61.

26. Brewer C, Harrower M, and The Pennsylvania State University. ColorBrewer 2.0: Color Advice for Maps Accessed July 2016.

Funding

This study was made possible using the resources of the Rochester Epidemiology Project (REP) medical records-linkage system, which is supported by the National Institute on Aging (NIA; AG 058738), by the Mayo Clinic Research Committee, and by fees paid annually by REP users. The content of this manual is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health (NIH) or the Mayo Clinic.

Documentation updated on November 13, 2024.