NIH Health Care Systems Research Collaboratory
Contact: support@popmednet.org

NIH Collaboratory Distributed Research Network Users' Guide

NIH Collaboratory DRN_Users' Guide_v5.0 for website_16Jun2015.docx
16June2015

Contents

Introduction

The **NIH Collaboratory Distributed Research Network (DRN)** facilitates research partnerships with organizations (Data Partners) that possess electronic health data that have been curated and formatted to support multi-site biomedical research.

The DRN Coordinating Center (DRN CC) administers the network and helps investigators identify and contact Data Partners to propose collaboration on specific research projects.

Kinds of data available and sources

Research ready data sets have been developed as part of the FDA Sentinel program (www.mini-sentinel.org/). Sentinel Data Partners, insurers plus some HMOs, currently participate in the NIH Collaboratory DRN; these data partners represent over 90% of the covered lives in the Mini-Sentinel system. Together, these data partners have over 300 million person-years of observation time and detailed information for billions of medical encounters and outpatient pharmacy dispensings. The age distribution of the population is shown in Figure 1.

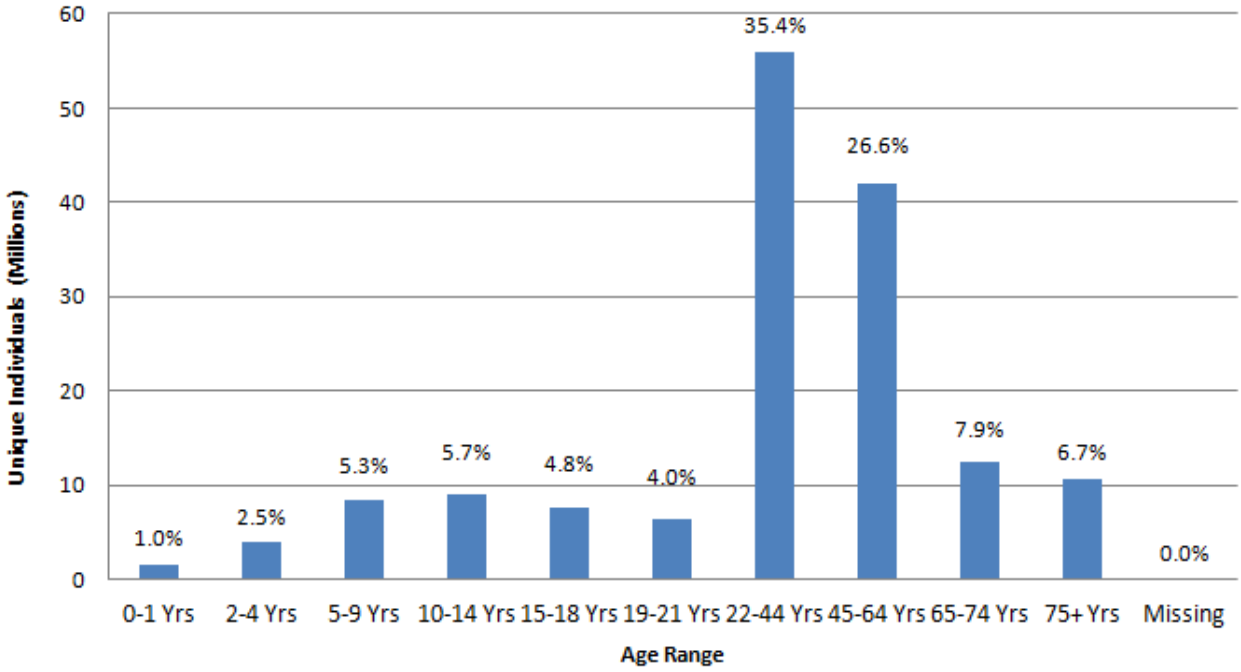
All data partners possess the following demographic, administrative, and medical encounter data, including:

- Enrollment and disenrollment dates (these denote periods during which medically attended care is likely to be observed)
- Age and sex
- Inpatient, emergency department, and ambulatory encounter information
  (1) encounter dates (admission and discharge dates for inpatient care)
  (2) coded diagnoses
  (3) coded procedures
- Outpatient pharmacy dispensing information

The following information is available for a minority of the covered population:

- Vital signs: height, weight, blood pressure, tobacco use
- Laboratory test results for these tests: alkaline phosphatase (ALP), alanine aminotransferase (ALT), absolute neutrophil count (ANC), total bilirubin, creatine kinase total (CK), creatine kinase MB (CK-MB), ratio CK-MB/CK, creatinine, fibrin d-dimer, glucose (fasting and randomized), hemoglobin, glycosylated hemoglobin (HbA1c), influenza virus (A, B, A+B, and not specified), international normalized ratio (INR), lipase, pregnancy test, platelet count, troponin I cardiac, and troponin T cardiac

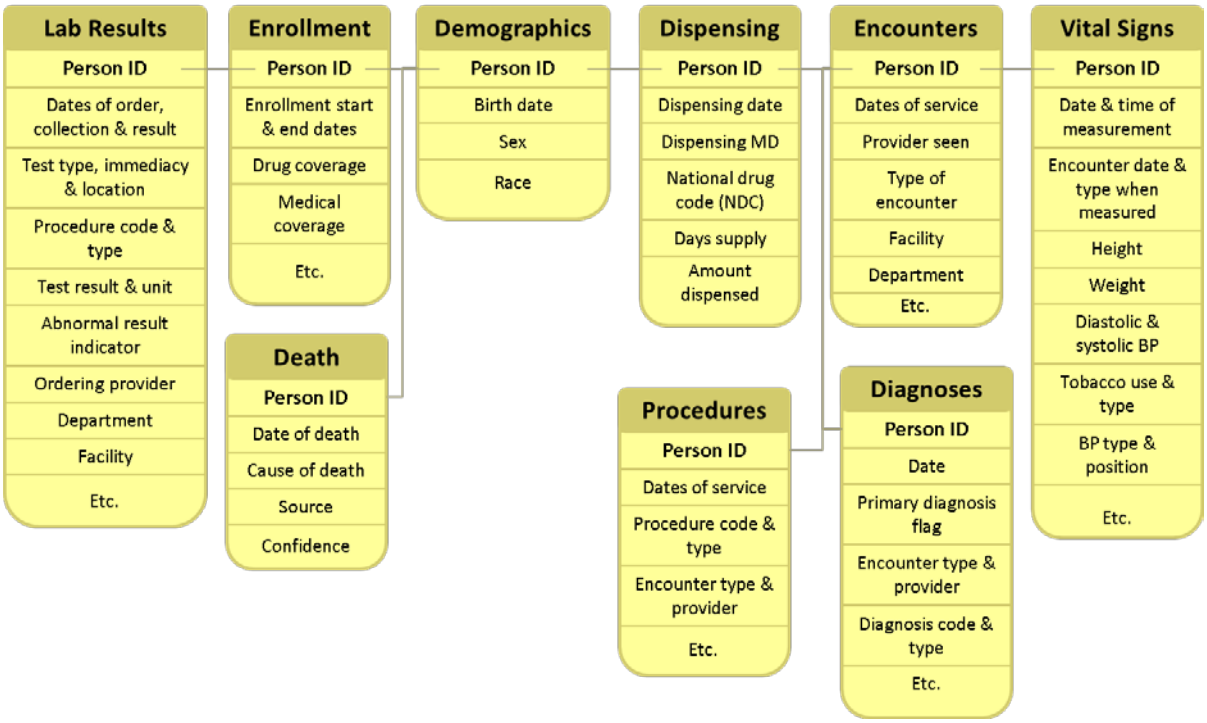
Figure 1. Unique Individuals by Age Range



Data elements

a) The data is maintained as a common data model, shown schematically in Figure 2. A full description of the data model is here: www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Common-Data-Model.pdf

Figure 2. Common Data Model



b) Completeness of data

- Only medically attended events in the "regular" health care system are captured; events that are not captured include out of hospital death, over-the-counter medication use, and immunizations provided in community-based immunization clinics.
- Inpatient treatments and procedures are currently identifiable if they are itemized within the overall hospital bill.
- Some data partners do not create every table; for example, vital signs are available for only a subset of individuals.

Examples of ways to use the data

Various ways of using the data are listed below, in increasing order of cost and complexity. To make most efficient use of the data resources, we recommend that all questions begin with the simplest type of query and progress in order of complexity. This will ensure that investigators gain experience with the data and characteristics, such as sample size, before developing more complex questions.

a) Counts of people with diagnoses or procedures, stratified by age, sex, and year: These are useful to assess trends in medical care utilization, generate background rates of conditions, and identify new users of medical products. These are the quickest and least expensive queries.

Example: Counts of incident and prevalent exposure to 24 cardiovascular therapy agents by individuals less than 18 years of age is here: http://mini-sentinel.org/assessments/medical_products/details.aspx?ID=257

b) Cohort identification and descriptive analysis: These identify cohorts using complex inclusion and exclusion criteria comprised of combinations of diagnoses, procedures, and treatments. It is then possible to generate rates of specified outcomes during "at-risk" periods.

Example: Rates of first diagnosis of kidney stones following exposure to anti-epileptic drugs is shown here: http://mini-sentinel.org/assessments/medical_events/details.aspx?ID=260

c) Comparative analyses, using propensity score matching to control for confounding: These programs provide descriptive analyses, including a typical "Table 1" that describes the cohort, the rates of specified outcomes for the two groups that are compared, along with unadjusted and adjusted rate differences and hazard ratios.

Example: A propensity-score matched comparison of the rates of angioedema among new users of angiotensin converter inhibitors vs. new users of beta-blockers is here: www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Known-Positives-ACEI-Angioedema.pdf

d) Other**:** Customized analyses can be developed to address questions not answerable with the standard programs described above. It is possible to review full text medical records for instance to confirm rare outcomes and their exposure status. These are time and labor intensive, compared to the prior types of queries.

Example:
www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Rotavirus-and-Intussusception-Report.pdf

This evaluation was also published in the New England Journal of Medicine:
Yih et al., Intussusception Risk after Rotavirus Vaccination in U.S. Infants. N Engl J Med 2014; 370:503-512 www.nejm.org/doi/full/10.1056/NEJMoa1303164

How the system works

Data Partners participate in collaborations on a project-by-project-basis. All data requests must be submitted to the NIH Collaboratory DRN CC using the **NIH Collaboratory DRN request form** (See Appendix B). The NIH Collaboratory DRN CC will conduct an administrative review to determine appropriateness of the query/request. This may involve clarification of the request with the requestor, and determination of appropriateness with relevant Data Partner Site PIs. Once a request is approved by the NIH Collaboratory DRN CC, the CC will initiate the request, manage the request process, and provide the results to the requestor. In addition, the NIH Collaboratory DRN CC tracks all network activities and ensures network functionality.

Distributed querying is typically accomplished through the following steps:

1. The requestor completes the NIH Collaboratory DRN request form and sends it to the NIH Collaboratory DRN CC;

2. The NIH Collaboratory DRN CC reviews the request, asks for any necessary clarification, determines best query mechanism, and determines whether the request is appropriate for the DRN;
3. The NIH Collaboratory DRN CC works with requester to develop the specific query to be distributed to the Data Partners and distributes the query to the Data Partners using the NIH DRN Query Tool;
4. Each Data Partner determines whether or not to answer the request and whether or not an agreement with the requestor is needed;
5. Data Partners use their local datasets to obtain results;
6. Data Partners securely send results to the NIH Collaboratory DRN CC using the NIH Collaboratory DRN Query Tool;
7. The NIH Collaboratory DRN CC reviews the results and submits the response to the requestor. Results are often aggregate counts, without confidential or proprietary data. The level of data sharing is determined, in advance, as part of the collaboration agreement, and all query responses can be reviewed by the Data Partner before they are released.

Authorized requestors include NIH Collaboratory leadership and the Data Partners, and other individuals/organizations designated by NIH Collaboratory leadership. All requestors must adhere to the responsibilities and expectations outlined below.

Requestors are responsible for:

- Completing request forms
- Clearly describing the nature of the request and the intended use of the findings, including grant or other funding applications
- Responding to requests for clarification from the NIH DRN CC and the EHR Core
- Working with NIH DRN CC and Data Partners to execute necessary agreements/contracts

Requestors are expected to:

- Use results provided only for the stated and approved purpose. This may include a requirement to keep results confidential or to distribute only upon approval.
- Abide by any other limitations on use, issued by the EHR Core, the NIH Collaboratory DRN CC, and/or the Data Partner providing and reporting data

The Data Partners are described in Table 1 below.

Table 1: Summary of Data Partner Data

	HealthCore, Inc.	HealthPartners Institute for Education & Research	Harvard Pilgrim Health Care	Aetna	Group Health Research Institute	Humana: Comprehensive Health Insights, Inc.	OptumInsight, Inc.	Ochsner Health Systems
Years of data available	2006-2013	2000-current	2000-2013	2008-2013	2004-2013	1997-2013	2008-2013	2011-present (partial data available prior to 2011)
Total enrollees in research database: all years	40,150,000	2.5 million	3.7 million	40.1 million	3,545,000	15.4 million	45.5 million	945,830 (as of 2013)
Total enrollees in research database: most recent year	15,108,000 (as of 2012)	970,000	955,000	22.1 million	715,000	7.2 million	22.5 million	514,284
Age distribution (%): Most recent year								
0-18 years	25.5	30	28	25	23	6	24	20
19-44 years	34.2	33	33	45	30	11	39	28
45-64 years	29.6	29	31	24	32	18	29	29
65+	10.7	7	8	6	15	65	8	23

Note: White columns describe a summary of Data Partner data available through the NIH Collaboratory Supplement; gray columns represent Data Partners who currently participate in the NIH Collaboratory DRN but are not participating in this Supplement

APPENDIX A: Query Examples

These examples are taken from the experience of the FDA Mini-Sentinel program. All current participants in the NIH Collaboratory DRN are part of Mini-Sentinel.

a) Counts of People with Diagnoses or Procedures, Stratified by Age, Sex, and Year: These provide incident or prevalent counts and rates of people with specified diagnoses, procedures or treatments. They are simple and inexpensive to perform and are therefore usually the best first query of the distributed data system, since the results can guide the development of subsequent queries using more sophisticated DRN programs. The example shown here uses summary tables to identify people with a condition of interest.

i. Progressive Multifocal Leukoencephalopathy

Query: The Mini-Sentinel Distributed Query Tool was used to obtain a description of counts and prevalence of one diagnosis code for Progressive Multifocal Leukoencephalapathy in the Mini-Sentinel Distributed Database. The queries were run against the ICD-9-CM 3-Digit Diagnosis Code Summary Table, and queries were run using data from the inpatient setting. The report provides Progressive Multifocal Leukoencephalopathy events per patient per year, age group, and sex in the inpatient setting and includes information from 18 Data Partners.

Result: The annual number of patients with an inpatient diagnosis consistent with Progressive Multifocal Leukoencephalopathy varied principally with size of the population under observation. In 2012 there were a total of 87 individuals. The table shows the age and sex distribution and prevalence rate.

Prevalent cases of Progressive Multifocal Leukoencephalopathy in 2012

Age (years)	Males	Prevalence rate per 10,000 enrollees	Females	Prevalence rate per 10,000 enrollees
0-21	1	0.01	0	0
22-44	16	0.14	8	0.07
45-64	29	0.31	18	0.18
65+	6	0.16	9	0.20

The full report is here: http://www.mini-sentinel.org/assessments/diagnoses_and_medical_procedures/details.aspx?ID=282

b) Cohort identification and Descriptive Analysis: These programs provide substantial flexibility in identifying cohorts of interest and linking the individuals to specified outcomes.

i. Identifying long term bisphosphonate users and assessing their fracture rates

Query goal:

1. Identify individuals who were continuously exposed to bisphosphonates for at least 3 years.
2. Assess the risk of both hip fracture and "fractures of interest" (principally subtrochanteric fractures).

A reusable "Cohort Identification and Descriptive Analysis" program in the NIH Collaboratory Distributed Research Network library was used to identify all new users of alendronate, risedronate, and ibandronate, and to characterize the frequency of subsequent events, including fracture of interest, esophageal cancer, hip fracture, non-vertebral fracture, or osteonecrosis of the jaw in the Mini-Sentinel Distributed Database. The population covered were members of four large health plans who had both medical and pharmacy coverage. We thus believe the data are complete both for exposure to these pharmaceuticals and for the outcomes of interest. Bisphosphonate exposure was determined from dispensing records (National Drug Codes) and outcomes were assessed via diagnosis codes (ICD-9-CM).

Seventy-eight scenarios were examined with different exposures, events, minimum episode durations, and exposure extension periods. The report includes counts of individuals, durations and amounts of exposures, and numbers and rates of outcomes. These are provided for each bisphosphonate and for multiple age and sex groupings. A list of ICD-9-CM diagnosis codes used for fracture of interest, esophageal cancer, hip fracture, non-vertebral fracture, and osteonecrosis of the jaw can be found in the full report's appendix (see link below). The time window for this request was January 1, 2006 to December 31, 2013.

Result: At the time this query was conducted, we estimated that there were approximately 22,000 current alendronate users in the Mini-Sentinel Distributed Database (MSDD) who had been exposed for 3 to 5 years. Approximately 9,000 people enter this cohort each year. This figure was obtained by extrapolating data through 2013, assuming steady state initiation of alendronate, and projected to the full Sentinel population. Fracture counts and rates are shown in the table.

Fractures in long term alendronate users*

Fracture type	Exposed people	Person time (yrs)	Fractures	Rate/ 10K yrs
Hip	34,428	138,386	725	52
Femoral fractures of interest	34,672	140,020	339	24

* New users of alendronate, continuously exposed for at least 3 years

The full report is here: www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Modular-Program-Report_MSY6_MPL1R_Select-Bisphosphonates-Select-Outcomes-of-Interest.pdf. Data partners performed this analysis to demonstrate capability.

ii. Assessing hemolysis rates among immunoglobulin recipients

Query:

1. Identify individuals exposed to each of several immunoglobulin products.
2. Assess the risk of hemolysis within one or ten days of administration.

A reusable program was used to investigate use of several immunoglobulin (Ig) product groups (subcutaneous Ig, other branded intravenous immunoglobulin (IVIg), other IVIg, and intramuscular Ig) and diagnosis of hemolysis events on the same day as (1 day risk window) and within 10 days (10 day risk window) of Ig injection. Exposure to immunoglobulins was determined through HCPCS and ICD-9-CM procedure codes. Hemolysis was identified through ICD-9-CM diagnosis codes. The query was run against the Mini-Sentinel Distributed Database (MSDD) for the time period of January 1, 2006 through December 31, 2012. The request was distributed to 18 Data Partners; most of the observations described below occurred among organizations that participate in the NIH Collaboratory Distributed Research Network. This report presents results for incidence counts of new Ig users, new lookup periods, total lookup period duration (days), number of users with an event, eligible members, and member-years only.

Result: There were 47,164 new users of immunoglobulins. Of these 329 were assigned an ICD-9-CM code consistent with hemolysis within one day of administration; 434 were assigned one of these codes within 10 days. Results were provided separately for each of 13 Ig preparations, and stratified by age, sex, and year of administration. Event rates are shown in the table.

Incident hemolysis codes among immunoglobulin recipients

	Exposed people	Exposure periods	Persons with event	% new users with event
1 day risk window	47,146	303,574	329	0.70
10 day risk window	47,146	303,574	434	0.92

The full report is here: http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_%20Modular-Program-Report_Immunoglobulin_Hemolysis_MSY4_MPR48_Report1.pdf

APPENDIX B: NIH Collaboratory DRN Request Form

NIH Collaboratory DRN Query Request Form

Background

This form should be reviewed, completed, and submitted by any requestor who wishes to submit a standard request/query via the NIH Collaboratory Distributed Research Network (DRN). The NIH Collaboratory DRN Coordinating Center (CC) will review the query for feasibility and to provide guidance on developing query specifications. This may involve clarification of the request with the requestor, and determination of appropriateness with relevant Data Partners. The NIH Collaboratory DRN CC will work with the requestor to select the most appropriate mechanism for each request. Once a request is approved by the NIH Collaboratory DRN CC, the CC will initiate the request, manage the request process, and provide the results to the requestor. (Please refer to section 4.2 of the NIH Collaboratory DRN Governance document for more details.) Distributed querying is typically accomplished through the following steps (Figure 1):

1. Requester submits completed request form to CC for approval

The requestor develops a question and sends it to the NIH Collaboratory DRN CC, using the NIH Collaboratory DRN request form.

The CC will review the request and, if necessary, notify the requestor of any questions.

Requests may include:

a. Data Partner and/or data source search
b. Data completeness and data characterization activities
c. Questionnaires and targeted questions
d. Analyses with aggregate results
e. Analyses with person-level data

2. CC selects appropriate mechanism for responding to request

Once the request is approved by the NIH Collaboratory DRN CC, the CC will select, with the requestor, the most appropriate query mechanism.

3. CC forwards request to potential collaborating Data Partners for review

Once the appropriate mechanism for responding to the request is determined, the CC will forward the request to potential data partners for their review. The potential Data Partners will review all requests and respond to the CC within five working days regarding their participation (i.e., whether or not they will answer the request) and whether or not an agreement with the requestor will be needed. This stage of the process may involve email and/or telephone discussions as the CC attempts to clarify any questions with the Data Partners.

4. Data Partners run query

Data Partners who agree to participate will respond using appropriate local resources.

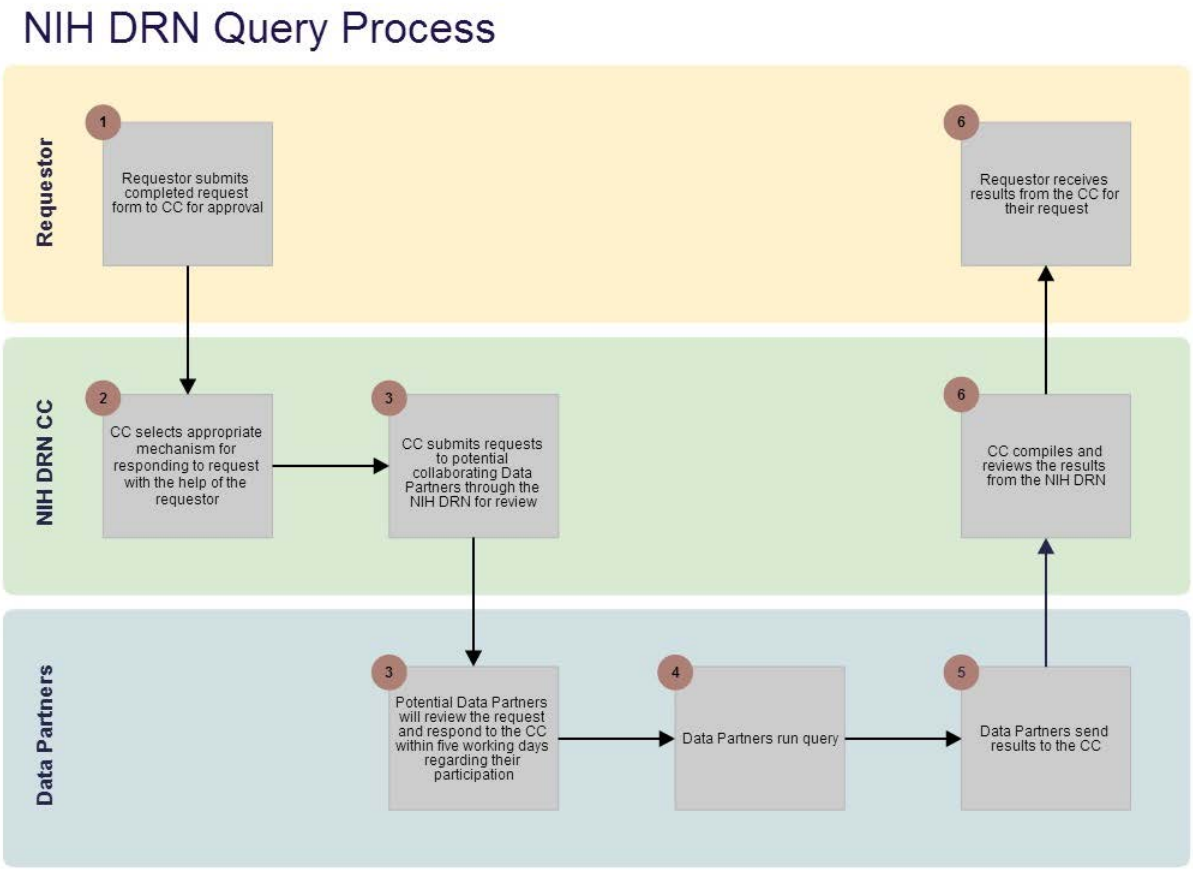
5. Data Partners send results to the CC

Once results are obtained, the Data Partners will securely return those results to the NIH Collaboratory DRN CC using the NIH Collaboratory DRN.

6. CC reviews and submits response to requestor

Upon receipt of responses, the CC will compile and review the results and submit the response to the requestor.

Figure 1.



To Be Completed by Requestor:

Please complete the remainder of this form and submit it to support@popmednet.org to initiate the request process. Provide as much of the requested information as possible. If you are unsure about codes or other details, provide as much description as possible and the Coordinating Center (CC) will work with you on details. The CC will follow up with questions as necessary. If you do not include a date for when you would like the requested information, your request will be considered low priority.

1. Provide the following general information about yourself and your timeline.

Today's Date	
Name	
Affiliation	
Mailing Address	
Telephone Number	
Email Address	
Date Information Needed	

2. Describe the specific information you hope to obtain from this request.

Example: "I would like to know the number of patients aged ≥25 who were newly diagnosed with condition X during 2008–2012 and received procedure Y within 6 months after diagnosis. I would like to see this information stratified by diagnosis year, gender, and age at diagnosis for at least [X number] of sites."

3. Provide a summary of your proposed research, if appropriate.

Include questions you hope to answer, hypotheses you want to test, etc. Example: "We are investigating the feasibility of a pragmatic clinical trial to investigate the comparative effectiveness of infection control interventions."

4. If known, why type of query request are you proposing? Describe the nature of your request (e.g., summary table queries, modular program requests, EHR Support for Public Health [ESP]).

5. Provide specific details about your request, if known.

If specific codes are not known, please provide a description of what is being requested. Insert any proposed table shells/outlines at the end of this section.

A. Diagnosis codes (ICD-9) and clinical conditions

B. Procedure codes (ICD-9, CPT, and/or HCPCS) and descriptions

C. Date range and type (e.g., diagnosis date vs. procedure date)

D. Age range and type (e.g., age as of when)

E. Drug names (brand and/or generic) and route of administration (if applicable)

F. Desired strata (e.g., specific age groups)

G. Other notes

For CC Use Only

Request ID	
Date Received	
Date Approved	