MARINER FAQ

PEARLDIVER RESEARCH | MAY 2025

Table of Contents

Mariner Description

PearlDiver utilizes numerous data sets in-house to derive their research capabilities. This provides an overview of the key elements from each of the primary resources.

PearlDiver All Payer Claims Database (MARINER)

- Medical and Rx and from 2010 through December 30, 2023, derived from provider networks.
- Includes a patient volume of over 174 million patients with claims billed to all payer types including commercial insurance, Medicare, Medicaid, self-pay, etc.
- Data based upon disease state. Will include all records for any patient with a record(s) containing a diagnosis from defined state.
- Offers longitudinal tracking based upon distinct patient identifier.
- Researchable to physician level.
- Records derived from ICD-9/10 diagnosis codes, ICD-9/10 procedure codes, CPT codes, NDC codes, and prescription groupings.

Common Questions & Guidance

How should I cite the use of PearlDiver in my publication?

PearlDiver Mariner Patient Claims Database (PearlDiver Technologies, Colorado Springs, CO, USA).

How many patients are in Mariner?

Mariner currently includes nearly 170 million patients in total.

How is age handled in Mariner?

All "Age" fields are based upon the year the patient was born. Age is handled in three different options.

"Age" can be used as a filter, or statistical variable and is based upon the year the patient was born. "Age" represents the patient's age in years at the time of event as opposed to a range of years.

"Age Range" is also based upon the year the patient was born and captures the specified range of patient ages at the time of the claim. This is available as a variable in statistical queries, cross tab reporting, and the breakdown report.

"Age Code" allows you to use the code that corresponds to the "Age Range". This can be found in filtering.

Are there patients over the age of 80 in the Mariner data set?

Yes. Older age groups have significantly lower counts. To protect patient privacy, these groups are rolled into the lower age group of 70-74 for the years 2010-2013, 75-79 for the years 2014-2017, or 80-84 for 2018-2023. This is done so patient privacy is maintained without losing the ability to perform research on these patients.

Common Questions & Guidance

Why do the patient counts by individual demographic groups add up to more than the Total patient count in the Breakdown report?

Each row of the breakdown report will report on the claims of the referenced population that meet the demographic criteria of the respective row. Patients with more than one claim may meet the criteria of multiple rows and will be reported in each row. Use of the 'firstinstance' or 'lastinstance' filter(s) will by definition limit the patient claims to a single claim. Patient counts by demographic will then add up to the total patient count.

What payer types comprise Mariner?

Mariner contains all payer types including commercially insured, Medicare, Medicaid, self-pay, and government plans.

What are the definitions of the service locations available?

Clinic – Services provided to patients in an outpatient setting that is often a free standing or satellite facility. Urgent care facilities, renal dialysis centers and federally qualified health centers are all examples of clinics.

Inpatient - Services provided during admitted stay in hospital or long-term facility.

Office - Services provided to patients in office setting where one or more medical doctors, usually general practitioners (GP), receive and treat patients.

Outpatient - Services provided to patients in outpatient rehabilitation facility, hospital - outpatient surgery, skilled nursing – outpatient, ambulatory surgical center, and hospital - emergency room.

What is the best way to ask PearlDiver a research question or report a problem?

We are always happy to assist with your questions. The best method to send us a question is to select "Report Issue" from the query progress screen or from the Help menu while creating your query in Bellwether.

Using "Report Issue" allows you to write a message to us, and also automatically sends us a copy of your query, any results generated from that query, and information regarding the data set you are running the query in. This allows us to get started troubleshooting right away and eliminates most back and forth via email to collect all necessary information.

Common Questions & Guidance

How do I ensure I am capturing the largest sample of patients?

For procedures that can occur inpatient, include both the ICD procedure code as well as the CPT in your population definition. In a perfect coding world, both of these codes should show up in the form of a facility claim and a physician clam. But in the coding world the reality is that it is common for there to be only one of these. To maximize the sample size, include both.

Example:

population1 = {CPT-12345, ICD-9-P-1234, ICD-10-P-12345}

Are there modifiers for physician CPT codes in Mariner?

While CPT codes do not have modifiers in Mariner data, searching for laterality in the new data set is possible using ICD 10 code definitions for inpatient procedures, as well as ICD 10 diagnosis coding for diagnosis codes accompanying the procedure.

Bellwether also offers the Laterality Filter command, which will create three new buckets containing the records of a given bucket that match a laterality of left, right, or unknown based upon the definition of the diagnosis code in the primary field. These new buckets will contain cohort name and the suffix _lt, _rt, and _unk.

Does the Marinera data contain mortality?

The Mariner data sets do not contain information on mortality. The suggested approach currently is to use the sudden/unexpected death ICD 9 & 10 diagnosis codes. While these codes may not capture all patients passing outside of provider care, these patients have definitively expired. These codes include: ICD-9-D-7980, ICD-9-D-7981, ICD-9-D-7982, ICD-9-D-7989, ICD-10-D-R99.

One additional option is to include hospice treatment codes to capture patients who have entered end-of-life care. This is a close proxy to expiration. These codes include: CPT-T2042, CPT-T2043, CPT-T2044, CPT-T2045, CPT-T2046, CPT-G9751, CPT-G9812, CPT-G9852, CPT-G9855, CPT-99377, CPT-99378, CPT-G9473, CPT-G9475, CPT-G9476, CPT-G9477, CPT-G9478, CPT-G9479, CPT-G9524, CPT-G9525, CPT-G9687, CPT-G9688, CPT-G9690, CPT-G9691, CPT-G9692, CPT-G9693, CPT-G9694, CPT-G9700, CPT-G9702, CPT-G9707, CPT-G9709, CPT-G9710, CPT-G9714, CPT-G9715, CPT-G9718, CPT-G9720, CPT-G9723, CPT-G9725, CPT-G9740, CPT-G9741, CPT-G9761, CPT-G9768, CPT-G9857, CPT-G9809, CPT-G9858, CPT-M1022, CPT-Q5004, CPT-Q5005, CPT-Q5006, CPT-Q5007, CPT-Q5008, CPT-Q5010, CPT-S0255, CPT-S0271, CPT-S9126

Common Questions & Guidance

What is the best way to determine average length of stay (LOS) or filter for the appropriate LOS terms within a patient population?

Just as in previous data sets, LOS is only on the inpatient claim (ICD 9/10 procedure code). To ensure an accurate LOS, you will want your population that is to be filtered or reported on to contain only the inpatient claim. Should you have a mix of code types in your population bucket, it is recommended you first filter for the inpatient claims only. This would look like:

LOSgroup = filter base codetype = {ICD-9-P, ICD-10-P}

If I am researching cost, how can I break out the source of reimbursements?

When using the 'Totalcost' query, many researchers like to break out the results based upon the source of the claims. Typically this is separated into Rx, inpatient claims, outpatient/physician claims. To separate the reimbursement amounts out in this manner, you can run a 'Totalcost' query and then filter based upon these groupings.

Example: Total reimbursement within 90 days following a procedure

Kneetcwithrequired = totalcost Requireboth start = 1 day after {Requireboth} end = 90 days after

Kneein = filter Kneetcwithrequired chargetype = I
Kneerx = filter Kneetcwithrequired chargetype = R
Kneeother = filter out Kneetcwithrequired chargetype = {I,R}

Can the average time of patient coverage for a population be determined?

Yes, this can be done using the 'min' and 'max' commands. This would look like:

allmin = MIN bucketname
allmax = MAX bucketname
tbortho = TIMEBETWEEN bucketname {allmin}{allmax}

My code is stuck at x% despite rerunning it multiple times. What should I do?

If your code is stuck at a certain percentage, it's likely due to longer data processing times, especially if some parts of your code run faster while others handle more data and take longer.

Longer Data Processing: Some operations can take more time depending on the amount of data being processed. If your code is working with large datasets or performing complex calculations, it may take longer to complete certain tasks.

Optimize Code: Try making your code faster by simplifying tasks or breaking data into smaller parts.

In statistical functions, what is the difference between “patientsum”, “patientmean”, and “patientmedian”?

When patient-level data is used, all the records that a given patient has in a bucket are combined into one totaled record for that patient. The formula for combining the numeric values is controlled by which patient flag is used - patientsum will total the numeric fields on the given patient's records; patientmean will average the numeric fields on the given patient's records; patientmedian will find the median value for each of the numeric fields on the given patient's records. For any non-numeric fields, the first instance record's values are transferred to that patient's totaled record.

See the example below:

Raw Data

patient id	age	gender	reimburse	los
patient 1	30	F	15	2
patient 1	31	F	75	1
patient 1	35	F	60	1
patient 2	65	M	100	4
patient 2	65	M	75	2

Data Submitted to R

patientsum

patient id	age	gender	reimburse	los
patient 1	30	F	150	4
patient 2	65	M	175	6

patientmean

patient id	age	gender	reimburse	los
patient 1	32	F	50	1.333
patient 2	65	M	87.5	3

patientmedian

patient id	age	gender	reimburse	los
patient 1	31	F	60	1
patient 2	65	M	87.5	3

Please note that Bellwether will automatically take the first instance age value when patientsum is used.

Data Resource Comparison

Resource Comparison

	MARINER	Medicare SAF
Payer Types	All Payer Types	Medicare
Years Covered	2010 – 2023	2005 - 2016
Coverage	174 Million Lives	61 Million Lives
Claims		
Diagnoses	Yes	Yes
Procedures	Yes	Yes
Rx	Yes	No
Lab	No	No
Longitudinal	Yes	Yes
Billing Sources		
Facility	Yes	Yes
Physician	Yes	5%
Sites of Care		
Inpatient	Yes	Yes
Outpatient	Yes	Yes
Other	Yes	5%
Codes	ICD-9, ICD-10, CPT, NDC	ICD-9, DRG, CPT
Modifiers	No	Yes
Granularity	3-Digit Zip	Facility
Charges	No	Yes
Reimbursement	Actual/Estimated	Yes
Mortality	No	No
Race	No	2015-2016