

Transitioning from Leaf to Data2Evidence - 2 Minutes

Andrew Deonarine

August 26th, 2026



**Icahn School
of Medicine at
Mount
Sinai**

What is AIR·MS?

AIR·MS is a research platform that is composed of:



A very fast, integrated database including electronic health record data, Pathology and Radiology metadata



A research workspace where you can use familiar tools to work with AIR·MS data



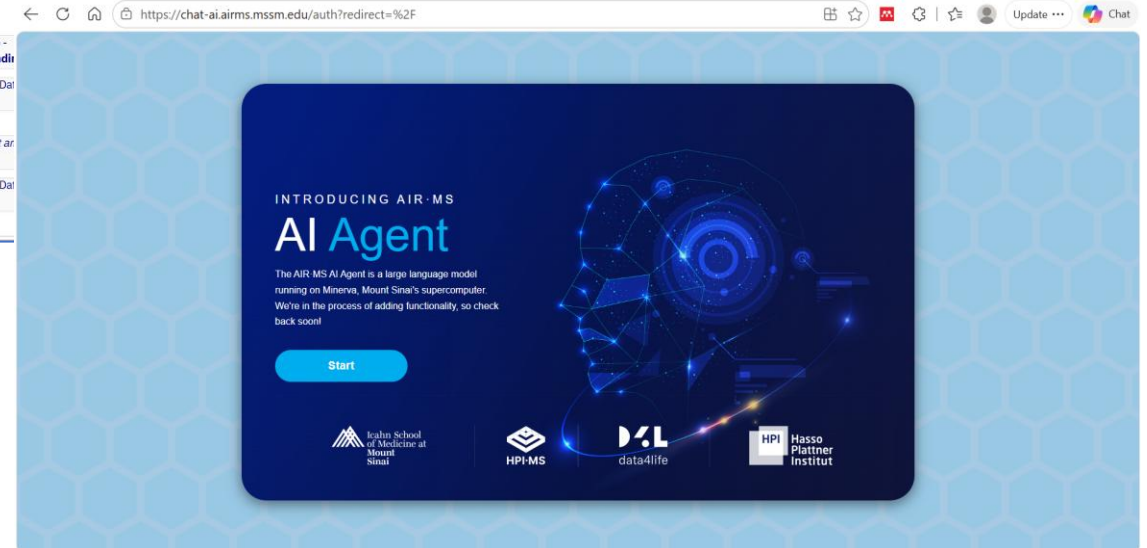
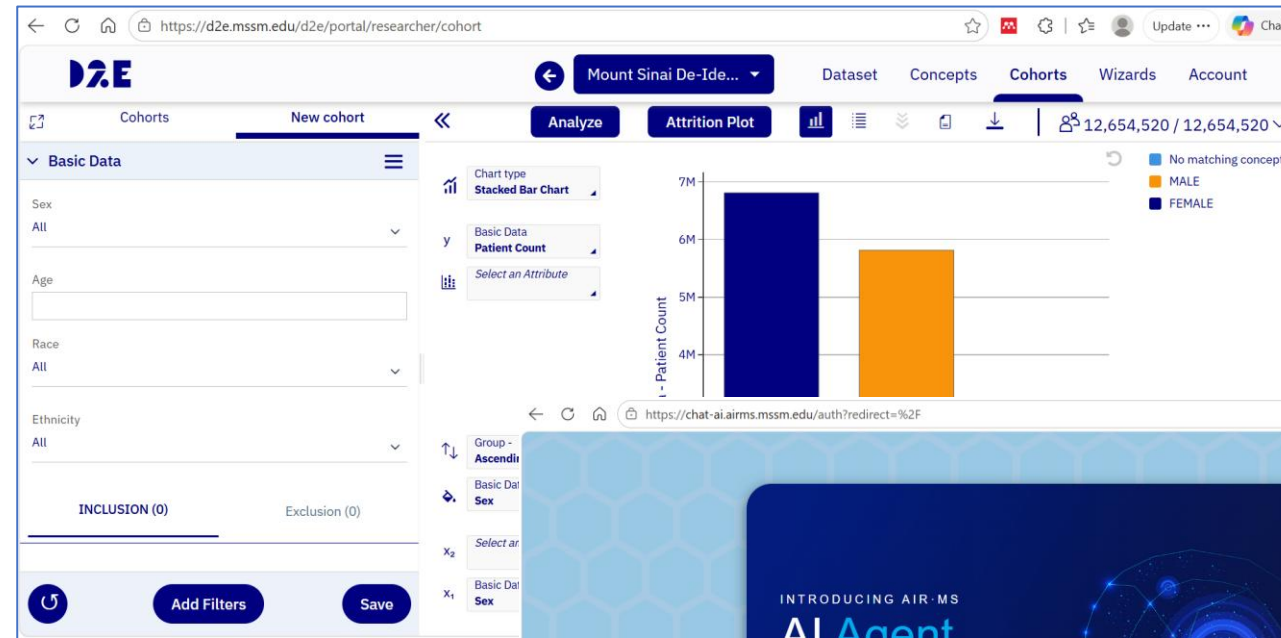
An **Application Tier** to host a growing number of applications – including cohort building tools and annotation apps

What Can AIR·MS Do?

Explore millions of clinical records in seconds using AIR·MS

AIR·MS contains millions of records and billions of measures derived from Mount Sinai electronic health record data

With a few clicks, you can start exploring the data!



What Can AIR-MS Do?

Three Ways to Access the Platform:

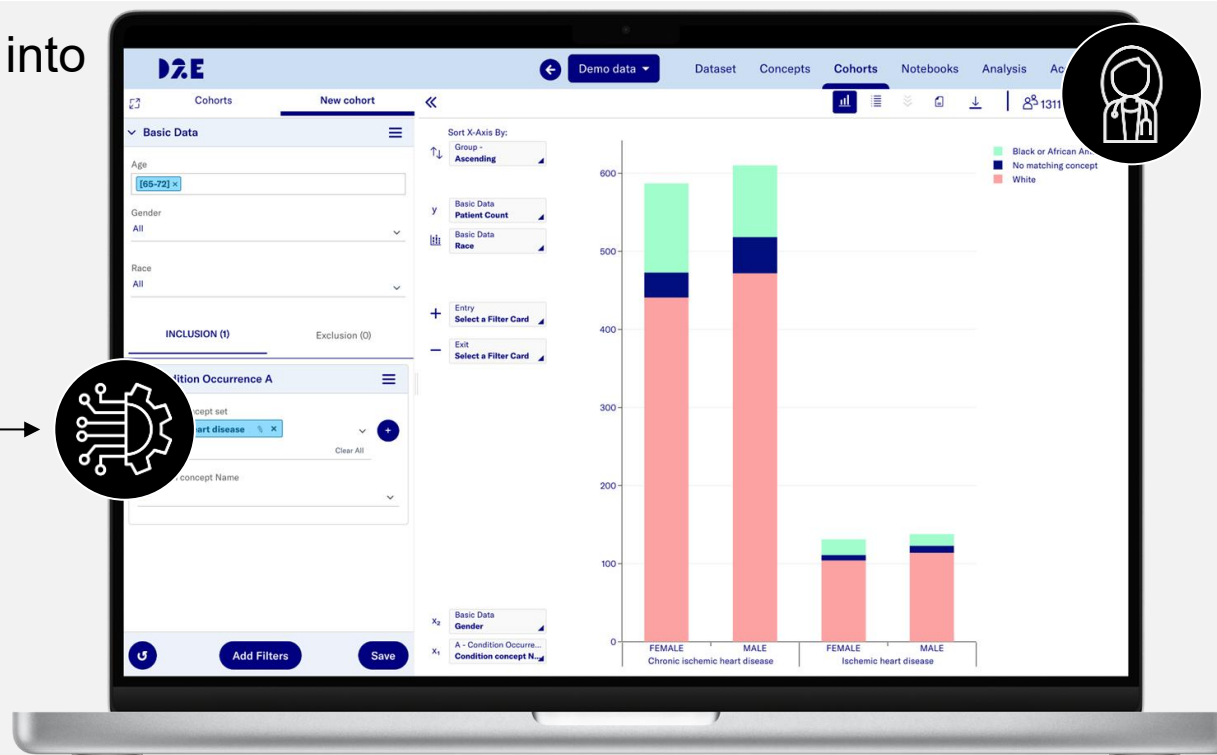
**Data2Evidence
Cohort Query Tool**

**Direct Database
Access (Using Python/R)**

**AIR-MS
Agent**

Search across multiple datasets and reduce work to test hypotheses, gain insight

Extract and combine data into interoperable dataset



Offer a one-stop platform to researchers

- Interactive dataset exploration
- Customized cohort creation
- Data quality checks for datasets and cohorts
- Collaborative analysis with other researchers

<https://d2e.airms.mssm.edu>

Step 1 – Accept the Disclaimer

Leaf

leaf.mssm.edu

leaf v3.11.4 ITHS Institute of Translational Health Sciences ACCELERATING RESEARCH. IMPROVING HEALTH. NATIONAL CENTER FOR DATA TO HEALTH

Patient data restricted to De-identified mode only

I want to find information for

Quality Improvement Research

I have an Approved IRB

No Yes

Planning to publish your results? Great! Please cite this manuscript to ensure we can continue making Leaf even better:
Nicholas J Dobbins, Clifford H Spital, Robert A Black, Jason M Morrison, Bas de Veer, Elizabeth Zampino, Robert D Harrington, Bethene D Britt, Kari A Stephens, Adam B Wilcox, Peter Tarczy-Hornoch, Sean D Mooney. Leaf: an open-source, model-agnostic, data-driven web application for cohort discovery and translational biomedical research. Journal of the American Medical Informatics Association. 2020;27(1):1-10.

Data2Evidence

d2e.mssm.edu/d2e/portal/researcher

Disclaimer

applies to all uses, including data being used in preparation of a project, or for purposes of research that is considered not federal regulated human subjects research.

6. For any identified (i.e., containing PHI) Datamarts, data sets or reports made available to me, I will exclude any subjects personally known to me or co-investigators except in a formal provider/patient relationship.

7. Data supplied for projects with IRB approval shall not be re-used or re-disclosed without explicit permission from the IRB.

8. I certify that I understand and agree to abide by the guidelines of the PPHS, the rules and regulations of the Mount Sinai Medical Center, and all applicable federal and state laws and regulations.

9. I understand that all access is audited, and that unauthorized access or inappropriate usage of data may result in disciplinary action up to and including termination.

A De-identified Data Set with the following limitations:

- Patients with an age of 89 and older are masked with a value of "Greater than 89".
- Cohorts less than 10 are masked
- Dates in the dataset are shifted equally per patient based on the date shift value. This ensures that the relative distance between dates in the patients' chronology remains intact.

Logout Accept

Step 2 – Now Click on Data => Cohorts => D2E

The image displays four sequential screenshots of the D2E web application interface, numbered 1 through 4, illustrating the workflow from Data to Cohorts.

Screenshot 1: The main dashboard shows a link to "Mount Sinai De-Identified Clinical Data – Leaf Config" circled in red. Below the link, it states: "Click here to explore de-identified data derived from the clinical care processes at the Mount Sinai Health System. This interface is designed to use vocabularies familiar to users of the Leaf Cohort Query Tool. Please note that it does not currently support using saved concepts sets for cohort search, and requires the user to specify each concept individually in the search interface. By July 1, we will deploy a feature that supports saving concepts for cohort search." Metadata includes: Total subjects: 12,654,520, Date: 2023-06-08, Version: 5.3, Data model: omop5-3.

Screenshot 2: The "Dataset" page for "Mount Sinai De-Identified Clinical Data – Leaf Config" is shown. The "Concepts" tab in the top navigation bar is circled in red. The page includes a description of the dataset and a metadata table.

Resource type	Dataset
Dataset ID	15bde2eb-97c3-4bad-962c-4ed2f6ff1da9
Updated date	2023-06-08

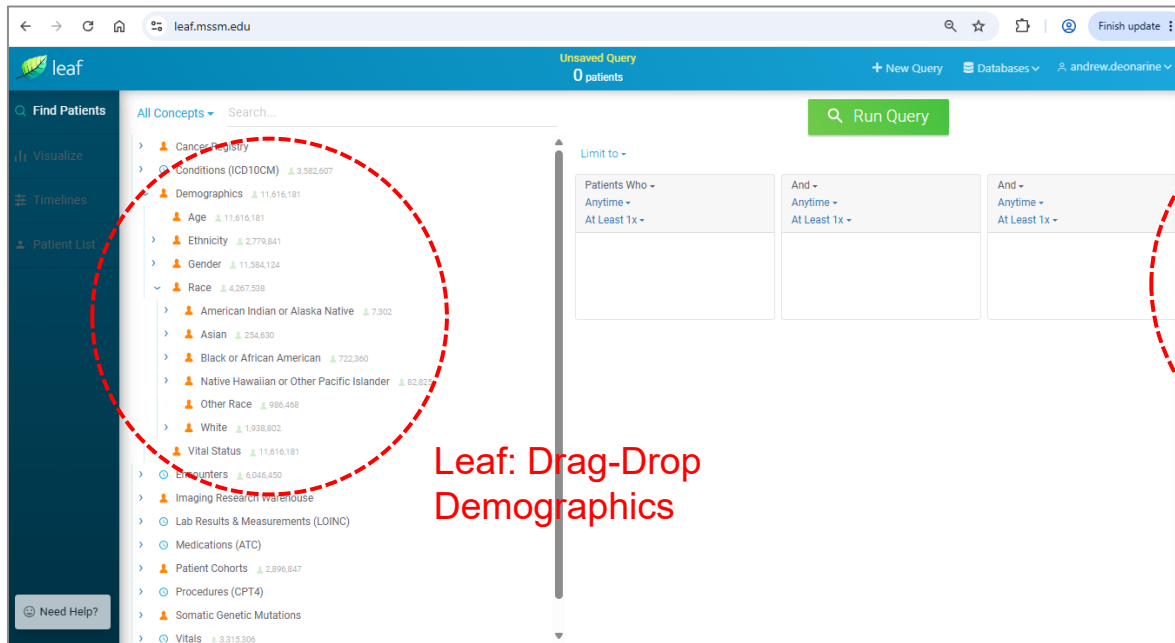
Screenshot 3: The "Cohorts" page is shown. The "Create Cohort:" button is circled in red. The button is labeled "D2E".

Screenshot 4: The "New cohort" page is shown. The "Basic Data" section includes filters for Sex, Age, Race, and Ethnicity. The "INCLUSION (0)" and "Exclusion (0)" counts are displayed. The "Analyze" button is highlighted. A bar chart titled "Basic Data - Patient Count" shows the distribution of patient counts by sex (FEMALE, MALE, No matching concept).

Sex	Count
FEMALE	~6,800,000
MALE	~5,800,000
No matching concept	0

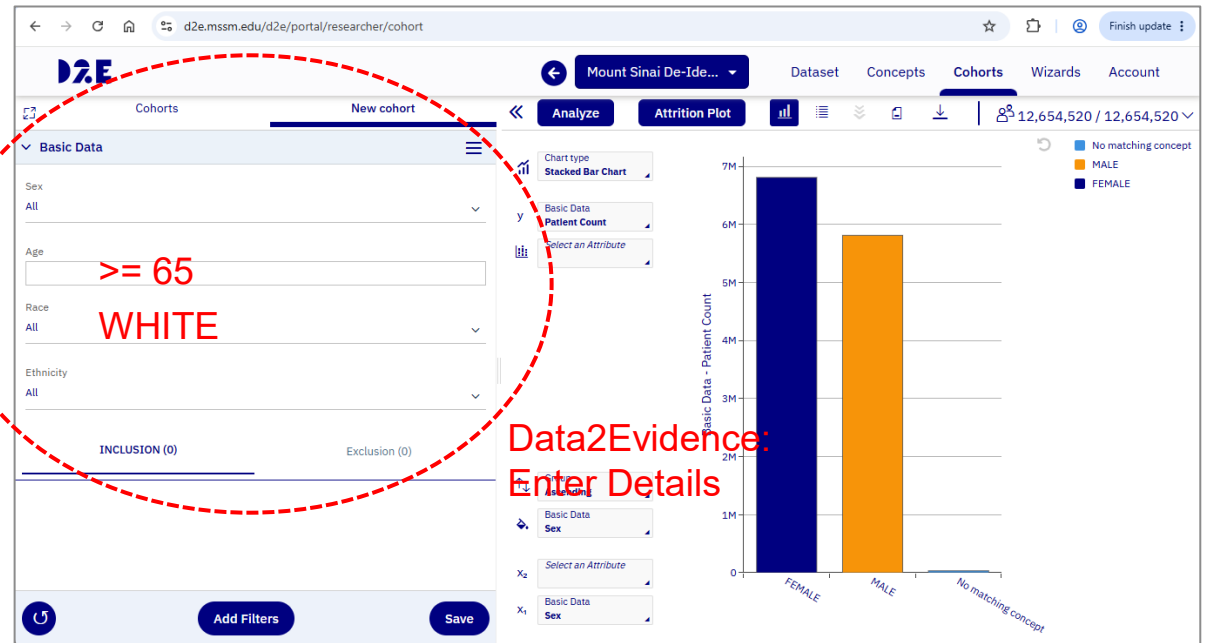
Step 3 – Set your Demographics

Leaf



Leaf: Drag-Drop Demographics

Data2Evidence



Data2Evidence: Enter Details

Sex	Patient Count
FEMALE	~6.8M
MALE	~5.8M
No matching concept	~0.2M

Instead of dragging and dropping items in Leaf (left), you can directly edit information (ex. ≥ 65 for age, "WHITE" for race) into Data2Evidence (right). The graph will automatically update.

Step 4 – Set the Rest of Your Query

Leaf

Leaf: Drag-Drop Other Filters

Data2Evidence

Data2Evidence: Click "Add Filter" and Enter Details

Once your demographics are set in Data2Evidence, you can then click "Add Filters" to add diseases, drugs, procedures, etc. (note: there are mapping limitations, see next slide). This is like dragging and dropping in Leaf (left figure).

Key Points

- Note that we are currently building the data mappings in AIR·MS, so you may search for some procedures (ex. Ophthalmology-related procedures), drugs (glp1 agonists) and other information and not see any results. This is a known issue and we're working to add new mappings on an ongoing basis.
- You can save and re-run your previous queries quickly and easily when you click on the initial "cohorts" screen.
- We are also adding "Wizards", which will perform basic calculations such as incidence, prevalence, readmission, etc. This is is being continually updated so please check back soon.

Data2Evidence: Getting Started

<https://labs.icaahn.mssm.edu/minervalab/air%C2%B7ms-training/>

Quick Start

- Click [here](#) to view a helpful video on how to get started with using Data2Evidence.

Data2Evidence Cohort Query Tool Training

Training slides and recordings from our two most recent sessions are now available here:

Session 1 ([SLIDES](#)) ([VIDEO RECORDING](#))

Session 2 ([SLIDES](#)) ([VIDEO RECORDING](#))

Data2Evidence Documentation

This document provides a description of the features of the platform and explains how to use it: [Data2Evidence User Guide](#)

Thank you



Icahn School
of Medicine at
**Mount
Sinai**