

# Why PicnicHealth Data?

## Comparing PicnicHealth data with existing research databases

### Introduction

Recognizing the need for accelerating medical product development and modernizing the drug approval process, the [21st Century CURES Act](#) was enacted in 2016 and augmented the ability for the FDA to incorporate real world data and patient perspectives into the drug approval process. While randomized clinical trials are still the gold standard, observational data is increasingly utilized to supplement clinical trial design, as synthetic controls in single arm trials and for label expansion.

Currently, most existing real world databases are either incomplete (i.e. claims data, system-based EHR data) or expensive and time-consuming to maintain (i.e. registries). PicnicHealth fills the gap by combining insights from electronic health records, claims data, and patient reported outcomes (PROs) for any patient from any facility within the U.S. By providing disease-specific structuring of narrative text sections of records, PicnicHealth is able to unlock the wealth of data that has historically been inaccessible to researchers. On average, PicnicHealth collects five to seven years of retrospective longitudinal patient data for each patient and can continue to collect prospective data as needed.

### Existing Databases

The most common and readily available real world data comes from administrative claims data. While claims data can provide an overview of patients' diagnoses and healthcare utilizations, they lack the clinical depth needed to truly understand a patient's disease course. There is little data on patient symptoms and test results, and ICD-9 and ICD-10 codes captured are reliant on varied provider billing practices.

An additional source of real world data comes from standard electronic health records (EHR). These are often limited to specific facilities and thus provide only a partial snapshot of a patient's care. In addition, EHR data is often only derived from the structured portions of the records, such as problem lists and medication lists, which lack context and are not always complete or kept up to date by clinicians. While natural language processing (NLP) has advanced in recent years, using NLP alone to extract narrative texts is not yet sufficient for producing transparent and regulatory grade clinical data.

Many researchers turn to registries and traditional observational databases for more consistent and longitudinal data, yet these are expensive to build and maintain, are typically isolated to specific disease areas, and can only provide predetermined data elements.

# PicnicHealth Data

## Overview

PicnicHealth addresses the shortcomings of existing real world data sources by offering a flexible and scalable platform that harmonizes both the structured and narrative sections of patients’ records, allowing for research-quality data that generate valid real world evidence. PicnicHealth retrieves medical records, including texts and original DICOM images, from all facilities where a patient has received care. The data abstraction process is both machine-guided and human-curated, which enables PicnicHealth to utilize natural language processing (NLP) to scale while ensuring data integrity. Table 1 compares key characteristics of PicnicHealth data and other sources of real world data.

|                       | Claims Data                             | Standard EHR Data  | Observational Database / Registry   | PicnicHealth Data   |
|-----------------------|---|--|---|---|
| Population            | Insured patients only                   | Patients seen at one or more specific facilities                 | Any patient   | Any patient   |
| Visits                | All billed visits                       | Visits from providers within specific facilities                 | Visits to participating providers or facilities only  | Visits from all providers in any U.S. facility that interacted with the patient   |
| Time span             | Variable, average 2-3 years             | Variable   | Variable  | 5-7 years retrospective, indefinite prospective follow-up   |
| Patient details       | Basic demographics                      | Demographics and descriptive social and family history, etc      | Variable, depending on the disease and study needs  | Standard EHR data and histories converted from narrative text to structured format  |
| Disease assessment    | Coded diagnosis                         | Problems lists, vitals, labs, pathology, imaging diagnoses, etc. | Can capture conditions of interest, does not contain complete provider notes, lab reports, or images                        | Standard EHR data, DICOM images, data from narrative texts converted to structured data, including symptoms progression, pathology details, physician assessments, etc. |
| Treatments & outcomes | Billed procedures, filled prescriptions | All procedures, all filled and unfilled prescriptions            | Can capture specific provider & patient reported outcomes, but does not allow for new data elements after database is built | Standard EHR data, drug dosages, responses to treatments, and patient reported outcomes   |

Table 1 - Comparing PicnicHealth data with existing databases.



Life sciences and health systems partners work with PicnicHealth to identify specific patient cohorts of interest. Onboarding criteria can be selected based on disease types, severity levels, treatments history, etc. Partnerships with patient advocacy groups and providers allow PicnicHealth to recruit patients with both common and rare diseases from every location within the U.S.

Patients consent to sharing their de-identified data, and in return for contributing to research, they gain free access to the PicnicHealth timeline which provides an easy to use, complete longitudinal view of their care. By empowering patients to control their own health data and share it with caregivers and providers, PicnicHealth enables patients to improve the care that they receive.

### Patient Reported Outcomes (PROs)

While electronic medical records are rich in data, they often lack objective measures of the patient experience. Unlike traditional EHR databases or registries with fixed data variables, PicnicHealth can dynamically generate the data elements collected by conducting surveys on patient symptoms, treatments, and quality-of-life changes in real-time. Survey methods and questions can be tailored based on research needs. Continuous engagements with patients ensure ongoing prospective data capturing. PRO results from participating patients and caregivers are de-identified and linked back to the medical records.

### Claims Data Linkage

PicnicHealth is able to provide medical records data that is linked to medical and pharmacy claims databases. Where closed claims information is available, claims data can provide confidence for the accuracy and completeness of EHR data. Claims and EHR linked datasets are available to research partners to enable a better understanding of health service utilizations and actual prescription fill rates which are not available through EHR records alone.

### Mortality Data

PicnicHealth integrates data from the Social Security Administration Limited Access Death Master File (LADMF), which is considered the most complete source of mortality data in the US. Patient's mortality status can be abstracted

through medical record reviews. In specific use cases, patient families may submit death certificates for more timely data reporting. Additional commercial data sources may be available as add-ons.

## Summary

PicnicHealth is a faster and less expensive alternative to traditional observational datasets, and it is more customizable than disease registries. PicnicHealth data combines the depth and breadth of both structured and unstructured EHR data, claims data, original DICOM images, and patient reported outcomes, to provide a more longitudinal and comprehensive understanding of a patient's care journey.