

# Mathematica Data Innovation Lab Transparency in Coverage Data Description

This is supporting information for the 2023 Health Data for Action (HD4A) Call for Proposals. To apply, or for more information about the funding opportunity, please visit [www.rwjf.org/cfp/hd4a4](http://www.rwjf.org/cfp/hd4a4).

At Mathematica, we use data, analytics, and technology to address pressing social challenges, from the effects of climate change on communities around the world to disparities in health care, education, and employment across the U.S. We're an employee-owned and mission-driven company, with a deep bench of expertise in both data and social science.

The Transparency in Coverage (TiC) rule came into effect July 1, 2022, requiring that most group health plans and health insurance issuers disclose price and cost-sharing information to participants, beneficiaries, and enrollees. The purpose of this rule is to “reduce the secrecy behind health care pricing with the goal of bringing greater competition to the private health care industry.”<sup>1</sup> Although the wealth of information has the potential to transform the healthcare landscape, gathering and deciphering all the data presents a major challenge for the industry. There is about 630 terabytes of data files<sup>2</sup> available online. The data files in their raw form are not ready analysis; rather, they require external reference data and industry expertise to make effective use of the data. However, few organizations have the resources to download, ingest, and analyze them in a meaningful way.

Leveraging expertise in data architecture, provider-payer contracts, medical claims, health care data sources, Mathematica had built a centralized database that downloads, processes, enrich, and analyzes health plan price data. Under the HD4A CFP, Mathematica will provide researcher with access to vast, research-ready data on negotiated prices through our database. Our data is linked across payers, plans, and providers. Researchers will have access to most up-to-date negotiated rates at the provider-billing code level for commercial insurance plans issued by national insurance carriers (such as United Healthcare, Aetna, Elevance) and regional carriers (such as local BCBS plans). The price data cover nearly all employer-sponsored insurance plans, as well as individual and group plans offered in the health insurance marketplace.

We make sure that the data are research ready by performing multi-faceted validation checks to identify and clean data anomalies. For example, we check for spurious combinations of billing codes and specialties using real-world data from Medicare, Medicaid, and commercial claims, and remove inappropriate rates (e.g., colonoscopy rates for OB-GYN practices) from the dataset. In addition, data may include several options for merge-on variables, such as provider details available in NPPES<sup>3</sup>, group-practice, hospital and system-level information, and billing-code level data such as Medicare rates. Additional linkages are possible to Covid RDB data and to Medicare claims at the NPI-billing code-billing class level. Specific examples of readily available augmented data include NPI type, hospital CCN and ASC flag.

We will work with our counterpart to restrict the amount of data to relevant plans, providers, and regions for their research question. Each tailored research dataset will cover specific plans, regions, specialties providers, and services of interest. As detailed in the data dictionary, billing codes include multiple types, such as MS-DRG, CPT, HCPCS, etc. Rates cover both institutional and professional arrangements, as well as both in-network rates and allowed amounts.

Proposals may focus on a range of topics, including, but not limited to, the following: rate comparisons between commercial and Medicare rates for specific billing codes and plans, rate comparisons in between plans for specific billing codes, rate comparisons within a plan across different providers, rate comparisons across plans in a specific region for distinct billing codes.

**Data Dictionary:** The data dictionary is available [here](#).