

Health Care Cost Institute 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.

The Health Care Cost Institute's (HCCI) multiyear data, which is de-identified and compliant with the Health Insurance Portability and Accountability Act (HIPAA), includes the health care claims for over 50 million members per year with health insurance through their employer for the years 2012 to 2021.

These data—contributed by Aetna, Humana, and Blue Health Intelligence—consist of fully adjudicated, paid, de-identified medical and pharmacy claims, including: patients' and providers' geography, provider specialty and encrypted ID, service dates, diagnostic (ICD-9 or ICD-10) and procedure codes (MS-DRG, CPT, HCPCS), prescription national drug codes (NDC), and allowed amounts—which are the actual amounts paid to health care providers by an insurer—plus any co-payments, deductibles, or co-insurance paid by the insured person. Members are assigned a unique identifier and can be tracked over time within an insurer.

The HCCI data is available to researchers in two different views in order to maintain the statistical de-identification of the data. The views of the data are relatively similar; Data View 1 has more detailed patient demographic information (year of birth) with geographic information about the member's CBSA of residence, while Data View 2 has more detailed geographic information (five-digit ZIP codes of patients and providers) and member age band (e.g., 18-34). The data dictionary has information on the variables available and which variables are available in each data view. The demographic information for members is limited to age (either year of birth or age band) and sex. Neither racial/ethnic nor socioeconomic information about the insured patient is included.

The license for the HCCI data allows data access for one year, with options to extend or renew for additional costs that would be the responsibility of the research team. During this year, up to two users per research team may access the data through HCCI's secure, virtual data enclave, hosted by NORC at the University of Chicago. The secure data enclave has a Citrix infrastructure that turns researchers' computers into "remote terminals."

Researchers who wish to merge external data files into the HCCI data may upload them to the secure data enclave with approval from HCCI. Providers are identified in the HCCI data by encrypted (de-identified) National Provider Identifiers (NPIs). External data files that have NPIs can be encrypted for merging with the HCCI data for an additional fee for which the research team will be responsible. Any planned merges and/or comparisons between the HCCI data and any other data sources need to be clearly specified in research proposals.

For the purposes of Health Data for Action, HCCI will license data to academic institutions and other non-commercial, policy research organizations. Researchers based at or working on ongoing projects affiliated with health insurance plans are not eligible. Researchers based at consulting companies, provider organizations, or research institutions affiliated with those organizations are not eligible.

Successful applicants' institutions will have to sign a standard license agreement with HCCI for data access, and successful applicants will have access to the HCCI data for 12 months after the initial entry into the data enclave. After the 12-month access period, teams may extend or renew their data license at current costs, for which the researcher will be responsible. If researchers choose not to renew their data license, the analytic files created inside the data enclave will be held for 12 months. HCCI will provide basic user support (e.g., answering questions about the data, helping researchers use the data enclave), but will not provide any technical assistance (e.g., cutting data files).

Proposals using HCCI data could focus on a broad range of topics, including, but not limited to, the following:

- Health care spending, utilization, and prices
- Trends in employer-sponsored insurance
- Trends in private insurance markets
- Consolidation, integration, and market competition

The following examples are illustrative. No particular interventions or issues are prescribed, and we encourage other topics and research questions that are not listed. Possible research questions might include:

- How does health care spending and use vary across populations with different chronic conditions?
- What are the trends in hospital and other health care prices over time, across geographic areas, and across kinds of patients?
- Are communities with high levels of Accountable Care Organization (ACO) penetration different with respect to utilization or prices from similar communities with minimal ACO penetration?
- How do higher or lower levels of community-level spending on public health or a social determinant of health correlate with health care utilization and/or spending?
- Do price transparency regulations or mandates appear to have influences on health care prices, and if so, for which services?
- How does provider consolidation influence health care prices?
- How do health policy changes (e.g., No Surprises Act) impact health care prices?
- What factors are associated with variation in physician reimbursement both cross-sectionally and over time?

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