

Healthcare Payments Database Program (HPD) Data Access and Release Webinar January 15, 2025 – Q&A Summary

Below is a summary of the questions and answers discussed during HCAI's webinar held on January 15, 2025. Also included are questions received live during the webinar, and many that HCAI was unable to answer in the time allotted. HCAI looks forward to seeing you at the next event! [Subscribe to HCAI's mailing list](#) to be notified of events and public meetings.

The Healthcare Payments Database (HPD) is California's All Payer Claims Database or APCD. The HPD is a research database comprised of healthcare administrative data, claims and encounters, generated by transactions among payers and providers on behalf of insured individuals. The webinar included an overview of the program and information on the process of applying for and accessing HPD Data using the recently opened application portal. Visit the [HPD Data Access and Release page](#) for more information.

Any additional questions or feedback may be sent to dataandreports@hcai.ca.gov.

General Question

1. What would HCAI recommend that requestors do for seeking technical assistance related to making a data request, developing proposals for external funding, and estimating the price of data specific to a project?

HCAI recommends that requestors submit their data request application as early as possible and pay the application fee. Upon receipt of the application fee, HCAI can provide technical assistance with finalizing the request, scoping the request for seeking external funding, and HCAI will make a preliminary decision regarding a request for a reduction to the price of the project, all leading to a more accurate scope of the project and accounting for the cost of the project.

Data Submission

2. Are insurance plans responsible for providing data to the HPD?

HPD data is collected monthly from health plans and insurers. Mandatory submitters are required by law to submit those data to the HPD System. Compliance is monitored by HCAI and enforced with support from the Department of Managed Health Care.

3. Which plans are exempt from submitting claims?

Self-insured ERISA-governed health plans, fully insured health plans below 40,000 covered lives, and federal health plans are not required to submit data to the HPD System. Also excluded: anything defined as not health insurance in California Insurance Code Section 106(b), supplemental insurance (including Medicare supplemental), stop-loss, student health insurance, vision-only, chiropractic-only, and discount plans.

4. The [HPD Snapshot](#) overview currently reports data through 2021. What is the expected cadence for the addition of more recent years?

Updating the HPD Snapshot to include data through 2023 is planned for release in early Q2 of 2025. The HPD Snapshot will be updated annually thereafter.

Non-Claims Payment (NCP) Data

5. Regarding non-claims service utilization, is this data currently missing for now? Is the utilization data available but without insurer/consumer paid info?

Utilization data (diagnoses, procedures, plan enrollment, and provider info) for capitated encounters (including Medi-Cal managed care) are available in the HPD alongside fee-for-service claims data. Non-claims payment amounts (including capitation payments) are not yet in the system. Regulations for collecting that data are pending approval with an estimate to begin data collection in late 2025 and 2026.

Data Request Process

6. Will there be documentation or a deck that outlines the Nonpublic Data application?

The HPD Data Use, Access and Release [regulations](#) define what is required as part of the nonpublic data application. The application can also be accessed following the creation of an [account](#) with the HPD data request portal.

7. What is, on average, the duration of the review and approval process for applications by HCAI?

HCAI does not yet have metrics on the average duration for the review and approval process. The approval process for HPD application can vary for each request type. HCAI's goal is to process and approve data requests in 120 days. However, this depends on multiple factors including responsiveness of the requestor, level of detail included in the application, type of data requested, and mode of accessing the data. Requests for commercial-only Standard Limited Datasets accessed via the secure data enclave will have the quickest turnaround time.

8. What is the process for requesting this data by a state government agency?

State agencies requesting HPD datasets will begin by completing an application in the [HCAI Data Request Portal](#) using the State Agency request form. Completing the request may require an Interagency Agreement and/or a Business Use Case Proposal, governed under the Interagency Data Exchange Agreement (IDEA). [View the IDEA](#)

[Guidebook here](#) which can be discussed with the HCAI analyst assigned to your request.

9. Do county agencies also go through a similar request process to state agencies?

County agencies go through a process similar to other standard requestors such as researchers and non-profit organizations.

10. Assuming there is a high bar for requests for research identifiable data for export, beyond a thorough review of system security, are there other considerations for this use case we should be mindful of?

Any requests for research identifiable data require both DRC and CPHS review. If a requester wants direct transmission of the data, the DRC will also need to approve that type of access. Requests for direct transmission will require justification for why the Secure Data Enclave will not meet the needs of the proposed data use.

11. If we need to have identifier of providers (hospitals and physician groups) in the dataset, do we need to request the Standard Limited Plus dataset?

Provider identifiers are available by request in the Standard Limited Data Plus (SLD+), Custom Limited, and Research Identifiable datasets.

12. Upon receiving authorization for a dataset, if another project can utilize the same dataset, is it necessary to submit an amendment notifying an expansion or modification of the project's scope?

Each distinct project is required to undergo a separate application review and approval process. An expansion or modification to an approved project will require a supplemental application to amend the existing approved project.

13. Is there a denominator file that accounts for every insured member regardless of whether they have a claim in a given calendar year?

Yes, the member eligibility tables include all subscribers and dependents enrolled per month in a health plan that submits data to HPD. Those enrollment records are included in the HPD dataset regardless of if the enrollee has a claim for that month or year.

Pricing

14. Are there opportunities to modify, update, or reduce our data request application, or adjust it based on the funding that may be available to a requester?

Data requests can be modified during the application process. Once a request is approved and final payment is made, there are no refunds for modifications to the approved request.

15. Would data pricing or the data request procedure be changed if only a geographically defined subset were needed (e.g. a county, or Covered California region)?

Data pricing for receipt of data within the Secure Data Enclave is based on the cost of administering the Secure Data Enclave and not the volume of data, an exception being when the volume of data to process or store requires additional compute power or storage space. Data pricing for receipt of the data via Direct Transmission is based on the files received and not the volume of data within the files. Subsets of data may be subject to the creation of custom files which carry an additional custom file creation cost.

16. For access to data in the Secure Data Enclave, does a separate license seat need to be paid for each team member who needs data access, or can a seat be shared between, for example a principal investigator and a data analyst?

Licenses within the Secure Data Enclave are specific to each individual user of the data and are not shareable.

Data Access

17. Regarding retention of the data, how long can we keep the data?

Data can be retained for the duration of the project defined and approved during the application process. Retention can be extended beyond the originally approved project duration by submitting a supplemental data request and receiving approval to extend the project duration.

18. What are the technical details behind extracting data? What tools are needed to interact with it. Are there hardware prerequisites like RAM, CPU, or storage?

The user will determine the size and file format of information they extract from the Secure Data Enclave, as authorized by HCAI staff. The file size of the approved extract will determine the amount of storage user will need on their local computer to receive those files.

19. Once data analysis or findings tables are exported from the Secure Data Enclave, do requestors need to have HCAI approval to use these data in external-facing products (e.g. meetings with non-DUA users, external campaigns)?

When there is a request or need to take data out of the Secure Data Enclave, HCAI will review the proposed exit of information. This requirement ensures any data product or information outgoing has been sufficiently deidentified in accordance with [California Health and Human Services Data Deidentification Guidelines](#) to minimize likelihood of reidentification. Additionally, HCAI will review the proposed download to ensure it aligns with approved use of the data, product creation, and dissemination considering potential scope creep and safety of healthcare providers, for example.

Before any data product or external-facing material is released, requestors are again required to submit draft public data products to HCAI for review to ensure continual compliance, as mandated by the Data Use Agreement and California Code of Regulations, Title 22, §97416.

Use Cases

20. Are there potential use cases for the data beyond care, treatment and payment? If it is intended for public or private analytics or research, what recommendations exist for obtaining global consent and providing privacy notice for covered entities?

The HPD program does not permit use cases for the data related to treatment, payment, or health care operations. Per California Health and Safety Code 127673.81(d), HPD data “shall not be used for determinations regarding individual patient care or treatment and shall not be used for any individual eligibility or coverage decisions or similar purposes.”

The HPD dataset contains member demographic and enrollment information, provider information, medical claims and encounters, and pharmacy claims and encounters. For determinations about the viability of any particular use case, submit an HPD data request. Requests to use HPD data for research must also be approved by the California Committee for the Protection of Human Subjects, which may provide further recommendation.

21. Is there the ability to use demographics for disparity assessments for Cal Hospital Compare?

To determine whether HPD would be helpful for hospital disparity assessments, one should submit an HPD data request and work with HCAI analysts to examine the detailed use case.

Data File Specifications

22. What are the earliest and most current dates of available claims in the database?

Data requests can be submitted for service dates in years 2018-2023.

23. Could you provide more information about the data quality? To what extent has HPD been validated for completeness? Is the data felt to be complete for all years from 2018 until present? Is the likelihood of data missingness likely to be similar for HMO-covered patients (commercial and Medicaid) as for others?

Refer to the [HPD March 2024 Legislative Report](#) (Chapter 5) which discusses the HPD data collection and validation processes, as well as quality and completeness. The legislative report also includes discussion of data completeness of key fields by submitter type (Medi-Cal, Medicare Fee-for-Service, and Commercial/Medicare Advantage).

24. Are data available on how representative the claims data are of inpatient discharges by payer? In other words, does missingness of hospital discharge claims vary by payer or is it fairly balanced between payer types (Medicare, medical, commercial, etc.)?

The [March 2024 HPD Legislative report](#) (Chapter 3) provides a general comparison of data represented in the HPD system to HCAI's emergency department and inpatient data. The data population and methodologies between these sources differ thus the evaluation provided is for initial evaluation purposes only. The comparison does not evaluate the missingness of hospital discharge claims by payer. However, Chapter 5 of the HPD Legislative Report provides an overview of completeness rates for key fields within the HPD as well as know data quality issues.

25. Do you have information on how complete the claims data are in terms of covering all inpatient discharges in the state? In other words, are all California inpatient discharges represented in the claims data?

HPD collects each member month of enrollment from submitters (including Medi-Cal). Month-by-month enrollment information is available in the dataset. Review the [March 2024 legislative report](#) for details on ED visits and inpatient discharges in the HPD data. Inpatient discharges for payers not submitting data will not be reflected in the HPD.

26. How often is the member eligibility file updated? More specifically, if we are interested in a question which requires monthly plan enrollment information for Medi-Cal recipients, would that be possible?

HPD collects each member month of enrollment from submitters (including Medi-Cal). Month-by-month enrollment information is available in the dataset.

27. How is HPD medical claims data different from Patient Discharge Data (PDD)/Emergency Department (ED)/ Ambulatory Surgery Center (AS) data? Are they linked with the same patient identifier?

Data collection requirements and specifications used by each program are different. The HPD program uses the [APCD-CDL v3.0.1](#) whereas the [Inpatient](#), [Emergency Department and Ambulatory Surgery](#) programs pull their specifications from many different national standards. There is not a common patient identifier between these data collection programs. [Chapter 3 of the HPD Legislative report](#) includes a comparison of the HPD system to inpatient discharges and emergency department visits which may provide useful information on the possibility of linking these data sources.

28. In the APCD common format, there is an employer identifier. However, in the "available elements" tab of the [HPD data request justification grid](#), there is no corresponding field. Is it possible to track members who are employees of the same employer? And is it possible to see members who may be enrolled in a family or multi-person plan?

Though the Employer Tax ID is a field in the CDL, it is not included in the HPD analytic dataset due to privacy concerns. Use of the Insured Group or Policy Number or Group Name may tie employees together. Research identifiable HPD datasets include both member and subscriber information which allows the user to locate the members under the same family or multi-person plan.

29. Does the research identifiable option allow you to see the IDs (i.e. NPIs) for providers?

NPIs are included as an option in the data element [justification grids](#) that must be completed when creating a custom research identifiable dataset. NPIs are also available in the Standard Limited Dataset Plus.

30. Does this have coverage for all of California and what type of facilities are included?

You can review the [HPD Snapshot](#) on the HCAI website to see how many Californians are included in the HPD dataset. The HPD dataset includes claims and encounters for procedures performed at facilities at which treatment is authorized under the health plans that submit data to HPD. Mandatory submitters for the HPD include, health care service plans, insurers licensed to provide health insurance, public self-insured plans subject to section 1349.2 and DHCS. HCAI also encourages voluntary submissions from private self-insured plans. For a full list of included and excluded coverage types please see the [HPD Data Release Website](#).

31. Is this data at an aggregated level or at the member level?

The HPD datasets are at the member level.

32. What portion of Medicaid is captured in the data? If it does not include all Medicaid recipients, does it represent a representative or a convenience sample?

Medicaid FFS and Managed Care claims and encounters are included in the HPD dataset. As seen on the [HPD Snapshot](#), over 13 million Medi-Cal members per year are included in the HPD dataset.

33. For the Standard Limited and Standard Limited Plus datasets, will the Managed Care Organization (MCO) plan names for Medi-Cal recipients be identifiable in the monthly enrollment data within member eligibility file? Additionally, are there monthly data available to track changes in insurance coverage?

No, the Healthcare Plan Code (which has a code unique to each MCO) is not included in the limited datasets. That field is available in the research identifiable dataset for each month of member enrollment in Medi-Cal. Submitters are required to submit accurate data on a monthly basis, including any changes; that information is present on the eligibility table.

34. Do the datasets include data that can be used to identify CalAIM related services, such as an individual receiving Enhanced Care Management or a Community Support?

The Aid Category Code is available in the Medi-Cal and may provide the detail necessary to identify such services.

35. What social determinants of health data might be available? Any other data elements possible, in addition to age, sex, location, like preferred language? How specific is location?

Refer to the [All-Payer Claims Databases Common Data Layout \(CDL\)](#) specifications, specifically the member eligibility file. The specifications include all demographic information that is collected for members, such as race, ethnicity, and Hispanic indicators. HCAI also collects members' primary language, along with other fields.

36. Is Integrated Delivery Network (IDN) data included (e.g. Kaiser)?

Yes, Kaiser submits data to HPD.

37. Will there be a dataset available to local health jurisdictions, similar to the HCAI ED and PPD datasets?

All datasets are available for local health jurisdictions to request. The Standard Limited Dataset may be most applicable to use cases common to local health jurisdictions.

38. Should there be a record for every outpatient encounter (i.e. office visit) for all patients insured by the type of insurer from which you collect data?

A record for every outpatient encounter will exist for every claim or encounter submitted by mandatory submitters. The volume of records for voluntary submitters will vary.

Enclave

39. Could you describe the Secure Data Enclave? How is it accessed and what tools are available? What are its limitations vs. direct transmission?

The Secure Data Enclave is a centralized service for 24/7 remote access to the HPD dataset. It is a secure, controlled environment hosted within HCAI's scalable infrastructure, allowing flexibility to move based on project needs. It ensures data protection in accordance with state and federal security and privacy rules.

The Enclave operates as a virtual machine accessed remotely from a user's computer, where HPD data is securely accessed. No data can be moved to a user's computer without administrator approval, or out of the Enclave onto the Internet. It offers data analytic tools such as Microsoft Office Suite software, SAS, R, and Python. Users can upload their data through the data request process, link to the HPD dataset, and perform analyses using the Enclave's scalable computing resources. Once data analyses are complete, de-identification extracts can be downloaded by users, pending HCAI staff review.

40. What can users expect in terms of Secure Data Enclave user support, system downtime, and review timeline for data being requested for download from the enclave?

User technical support will be available during normal business hours and will be handled through a ticketing system. The Secure Data Enclave will have high availability, except for regularly scheduled maintenance performed monthly during an established 4-

to-8-hour period outside of core business hours; or other necessary downtime as determined by the State.

41. Does the Enclave have Stata software?

Stata is not a base software package included in the base pricing for Secure Data Enclave licenses. Stata may be added to the Secure Data Enclave for \$880 per year.

42. Is standard SAS (not SAS Viya) included as part of the included software query tools?

SAS is not a software tool included in the base pricing for Secure Data Enclave licenses. SAS Viya may be added to the Secure Data Enclave for \$4,000 per year. Standard SAS software is not available in the enclave.

Data Linkage

43. How will HCAI link with other available data (e.g., CA Medicaid)? Will enhanced community supports be included in the dataset?

The HPD dataset already contains California Medicaid data. To link HCAI data to other databases, the applicant should clearly describe that need in the HPD data request. The HCAI analyst will work with the requestor to select the best options for linking those data.

44. Will identifiers be available to link patient data of an individual enrollee across HPD over multiple years?

Yes, there are HCAI-generated identifiers in the HPD dataset that consistently identify the same individual across years and across plans.

45. In the case of foster/adoptive youth, they often change names and social security numbers when adopted; would you still be able to track adopted youth longitudinally?

Identifying individuals longitudinally relies on name, address, date of birth, and other identifiers (like SSN). While it may still be possible to identify a person whose name, address, and SSN changes, it will certainly be less reliable without those elements.

46. Is it possible to link moms and babies in the limited dataset on the Secure Data Enclave?

The limited dataset does not include member information that can be used to identify and link moms and babies.

47. Is it possible to identify and link members sharing a family plan? If not, does the premium and plan type information apply to each member, with each member's data reflecting the full premium?

The limited dataset does not contain the subscriber ID, so it is not possible to link family members in the limited dataset. Premium amounts are only reported on the eligibility

record for the subscriber and include the full premium for all members covered under that subscriber.

48. Is it possible to add certain variables from DHCS to the HPD data through the current system or do researchers need to do linkages on their own?

To link HCAI data, the applicant should clearly describe that need in the HPD data request. Additional data from DHCS will need to be separately requested and approved from DHCS. The HCAI analyst will work with the requestor to select the best options for linking those data.

49. Is it possible to link the HPD data with the HCAI IP limited dataset? Does this linkage require approval?

HPD data can be linked with HCAI inpatient data. The request for HCAI inpatient data is currently a separate application and approval process. The linkage will require HCAI approval via the separate data request processes. HCAI plans to more fully integrate these datasets and request processes in the future.

50. Can you link outpatient pharmacy meds to inpatient records? Is it limited to certain insurer groups?

If the same individual obtains medication through a plan that submits data to HPD and has an inpatient visit paid under a plan that submits data to HPD, those records can be linked to the same individual. The only records in the HPD dataset are those submitted by health plans that are registered HPD submitters; if a claim or encounter is processed under an insurer that does not submit data to HPD, it will not be present in the HPD dataset.

51. Is it possible to identify patients who rely on durable medical equipment?

Claims that include durable medical equipment are available in the HPD dataset, and those claims will include a patient ID that can be used to link to other datasets and/or tables within the HPD dataset.

52. The limited dataset grid appears to have 5-digit zip code available; is this correct, not 3-digit, but the full zip code for linking?

The Standard Limited Datasets have 5-digit zip code available.

53. *Other specific questions about use cases and the viability of using the data for any specific project.*

To determine whether HPD would be helpful for your specific use case, one should submit an HPD data request and work with HCAI analysts to examine the detailed use case and linkage possibilities.