

Overview of the Multi-Payer Claims Database (MPCD)

Genesis of the MPCD

The MPCD project is one of a number of initiatives related to comparative effectiveness research (CER) funded by the American Recovery and Reinvestment Act of 2009. The Act provided \$1.1 billion to build the necessary infrastructure and capacity to support CER. Approximately 25% of the \$400 million allocated to the Office of the Secretary within the Department of Health and Human Services (HHS) was appropriated to data infrastructure projects such as the MPCD. Within HHS, ASPE was tasked with managing the MPCD project in partnership with the Centers for Medicare and Medicaid Services (CMS).

Purpose of the MPCD

The project represents a public/private partnership with the goal of consolidating access to longitudinal data on health services financed by public and private payers, to help facilitate CER. Under the current implementation contract, use of the MPCD will be limited to CER applications. Once the MPCD enters its operation and maintenance phase, it is envisioned that the database will continue to be used primarily for CER-related inquiries; however, other uses may be considered subject to the approval of data contributors.

Inclusion of data from multiple sources should allow for adequate coverage of priority patient populations, less common medical conditions and health care interventions, and geographic areas. A number of states have developed or are in the process of developing all-payer claims databases (APCDs). These databases vary in the types of claims they collect, but they generally include data from private insurers, Medicaid, and Medicare Advantage plans. However, they often do not incorporate Medicare fee-for-service claims due to the additional cost of acquiring these data from CMS. The MPCD will serve as a centralized data access point where qualified researchers will be able to request data from state APCDs, which may be supplemented with other types of data such as Medicare fee-for-service claims and private health plan data for states that do not currently have APCDs. Ideally, patients will be able to be tracked over time as they move between different payers. Hence, in addition to allowing for comparison of health intervention outcomes between the elderly and younger age groups (e.g. outcomes for antihypertensive treatments) and evaluation of practice variations across different regions of the country, the MPCD also has the potential to make identifying and requesting relevant data more efficient for researchers.

Overview of project phases

The project includes a series of four contracts to design and implement the MPCD. The first contract was awarded to Avalere to develop a basic framework for how the MPCD should be designed based on input from a wide range of stakeholders with experience in building, administering, and using claims databases. This phase of the project was completed in April 2010. The second two contracts were pilot projects that explored methods to integrate data from different sources (awarded to Vexcel) and to develop a user interface prototype for a web-based portal to analyze data (awarded to Thomson Reuters). These phases of the project are nearing completion and are informing the design of the MPCD.



In the current phase of the project, which began in January 2011, OptumInsight (formerly Ingenix) was awarded a contract to implement the MPCD. This is the largest and most critical phase of the project. The MPCD project management team is in the process of negotiating with potential data partners about the data they can contribute to the MPCD. The technical design plans and data access models are being developed in tandem with these discussions to address the needs of potential partners.

Who will contribute data to the MPCD?

Given their status as parties to the contract, it is expected that CMS and OptumInsight will contribute some portion of their data to the MPCD. Aside from OptumInsight and CMS, the project team cannot list other contributors, because we are currently formalizing these arrangements. In addition to numerous private data partners, such as commercial health insurance companies and data aggregators, we are actively pursuing states that have all-payer claims databases in place or that are nearing completion. The goal for the initial implementation stage is to include two additional data partners in the MPCD. Other potential partners, including states nearing completion of their APCDs and contributors that require additional time to establish data use agreements with the MPCD, will be pursued in later phases of the project.

How will the database function?

Access to the data

As currently envisioned, the MPCD will include three tiers of user access. In the first tier, visitors to the web portal will gain free access to public use files (PUFs) that contain aggregated data that is not considered sensitive in any way. The second tier consists of a centralized repository of standard analytic files (SAFs), de-identified using a variety of methods, and drawn from a portion of the total claims available from data contributors. In addition to only including a sample of the overall claims, the number of available data elements (variables) in the SAFs may also be limited to those posing minimal risk to data partners or the values of sensitive elements may be masked or distorted. After registering on the web portal, users will be able to submit requests for the SAFs that will be reviewed by a Data Stewardship Board, discussed in further detail below.

For studies that require additional data that are not available in the first two tiers of access (e.g. studies on rare conditions), researchers will have the option to submit a request for a customized data extract from the entire set of claims available from data partners. These requests represent the third tier of data access. In this tier, after a customized data extract request is submitted by a researcher via the web portal, the request will initially be evaluated by the Data Stewardship Council to ensure that it meets the basic criteria used for third tier approval. After this review is complete, the request will be routed to relevant data partners for their approval.

Centralized vs. distributed data

To accommodate data contributor preferences and concerns, the MPCD will utilize a hybrid data storage model with centralized and distributed components. Some partners may prefer to have the MPCD manage their data due to limited resources. Hence, these partners may decide to contribute their data to a centralized repository housed at the Buccaneer Data Center where the CMS Chronic Condition Warehouse data are currently stored. Other partners with greater resources may prefer to



retain physical control of their data and provide access to it on a per request basis. The distributed model allows for this flexibility. Under either option, no data will be released to requestors until the appropriate level of approval is granted. Tier 2 requests will only require approval by the Data Stewardship Board, while Tier 3 requests will also require approval by data contributors.

For both the centralized and distributed components, the only personally identifiable information (PII) elements that will ever need to be transferred between data contributors and the MPCD processing facility are the state, year of birth, and gender associated with a claim. These are necessary filter criteria for data requests. The remaining PII fields will not be transmitted, because the processing center will execute an MPCD-provided one-way hashing algorithm against the PII elements in the data contributor's environment. Since the MPCD processing center will not maintain the mappings between clear text values and the hashed values, the decrypted PII cannot be re-identified upon arrival at the processing center.

In order to facilitate the matching of individuals across data sources all partners will be required to use the MPCD-provided hashing algorithm. After the successful execution of the linking algorithm, each unique individual will be assigned an MPCD individual identifier. This identifier will serve as the link for the hashed PII information for an individual that may appear across multiple sources. For distributed data sources, once the hashed PII information from a data contributor is assigned/mapped to an MPCD individual identifier, a combination of the MPCD individual identifier and the corresponding hashed version of the data contributor-specific identifier (e.g. health insurance subscriber number) will be pushed back into the MPCD appliance located in the data contributor's environment. This is done to facilitate the ability to pull the claims associated with an individual across the data contributors.

Privacy and Security

The MPCD will include stringent privacy and security safeguards and will be designed to comply with provisions in FISMA, HIPPA, ARRA, and with CMS IT security and privacy policies. As mentioned above, all data made available to requestors will be de-identified.

The hashed PII information that is stored in the MPCD for both the centralized and distributed data sources will only be used for linking individuals across data sources. These data will not be included in any of the data extracts that are produced from the MPCD. The hashed PII information will be stored in a separate location in the database with access privileges limited to the program that executes the "individual linking" algorithm. To minimize the risk of inferential identification through linkage attacks, the data will undergo a statistical review and when necessary the data will be modified to protect patient anonymity, such as by masking small cell sizes and modifying dates of service. As an added precaution, once the data have been loaded onto the web portal, users will have a limited amount of time to download the data before it is removed from the portal and subsequently destroyed.

The project may also test a data enclave access approach. Data enclaves are controlled and secured environments where authorized users can perform analysis using sensitive data that are protected from external exposure. A variety of data enclave approaches currently exist ranging from working on data at an off-site facility (i.e. a location other than a researcher's office) with secured computers that do not allow the data to be copied and transferred out the facility, to accessing the data using a secured computer at a researcher's worksite. We are currently assessing the merits of various data enclave approaches.



In addition to making the MPCD a secure environment to protect patients, we also recognize the importance of protecting data contributors from unauthorized access to their data. Data contributors will retain ownership of their data and can revoke MPCD access to their data at any time. As mentioned above, data contributors must approve all tier three custom data extracts when researchers require access to more comprehensive and sensitive types of data. We acknowledge that some potential data contributors may be concerned about OptumInsight's role in developing and operating the MPCD during the implementation phase, given they may compete with OptumInsight or its parent company, UnitedHealth Group. Hence, it is important to make clear that the UnitedHealth Group (including OptumInsight) is prohibited from transferring data from Buccaneer to storage facilities operated by UnitedHealth Group, unless the data originates from UnitedHealth Group or the organization has received appropriate approval to access data through the same process available to all data requestors. In addition, OptumInsight is prohibited from performing analyses on data residing within the MPCD for purposes other than maintaining the database and processing data requests. The software application used to standardize and merge the data will reside at Buccaneer, and data contributors will have the ability to audit/monitor all actions performed on their data using this software. As members of the Data Stewardship Council, contributors will have the authority to cease and inspect the processing of their data if any unusual activity is detected.

MPCD governance structure

The project includes a number of advisory bodies that will provide recommendations to the contractor as it develops and implements the MPCD. A Leadership Council composed of policy leaders and data experts across multiple agencies within HHS is helping to guide the project. The project also includes an advisory Governance Board convened by AcademyHealth that consists of individuals representing researchers, private payers, providers, consumers, states, and HHS. This board will provide advice and input to the project team on key design and operational decisions. Once the initial set of data partners has been established, the contractor will convene a Data Stewardship Council that will include representatives from the data partners. This board will be responsible for reviewing and approving data requests, as well as making recommendations on the data distribution process.

Timeline and testing

The exact date for making the MPCD available to the public is yet to be determined, but will occur before September 2013. Ahead of this date, the MPCD will be tested extensively by selected researchers within HHS and from the University of Washington. The contract allows for 100 queries from individuals within HHS prior to its public release. An expert panel of researchers within HHS familiar with claims data has been assembled to test the system beginning in February 2012, to provide the project team with feedback on how the system can be improved. The Centers for Comparative and Health Systems Effectiveness at the University of Washington will also conduct end-user testing.



Sustainability of the project

It has not yet been determined who will operate and maintain the MPCD once the period of performance under the current contract is complete. However, the federal government will retain ownership of the MPCD infrastructure upon completion of the contract. It is a priority of the federal government that the MPCD be a sustainable resource available for future investigators involved in CER. The Governance Board will play a key role in evaluating which sustainability options appear most promising.

Future plans (post-implementation)

As the title of the project suggests, the MPCD will initially include claims data, since these data are most readily available. However, the MPCD infrastructure is being designed to accommodate data with additional clinical detail from other sources such as EHRs, should such data become available in the future and be deemed feasible to include in the MPCD. In addition, the project team plans to include some value-added analytic tools in the post-implementation phase such as episode grouping, risk adjustment, and quality measures.

