Multi-Payer Claims Database /

Task 12: Summary Report and Recommended Design Option

May 2010

Avalere®

This report is based on research conducted by Avalere Health under contract to the Department of Health and Human Services (HHS), Washington, DC (Award No. HHSP233201000162G). The findings and conclusions in this document are those of the authors, who are responsible for its contents, and do not necessarily represent the views of HHS. Therefore, no statement in this report should be construed as an official position of HHS.

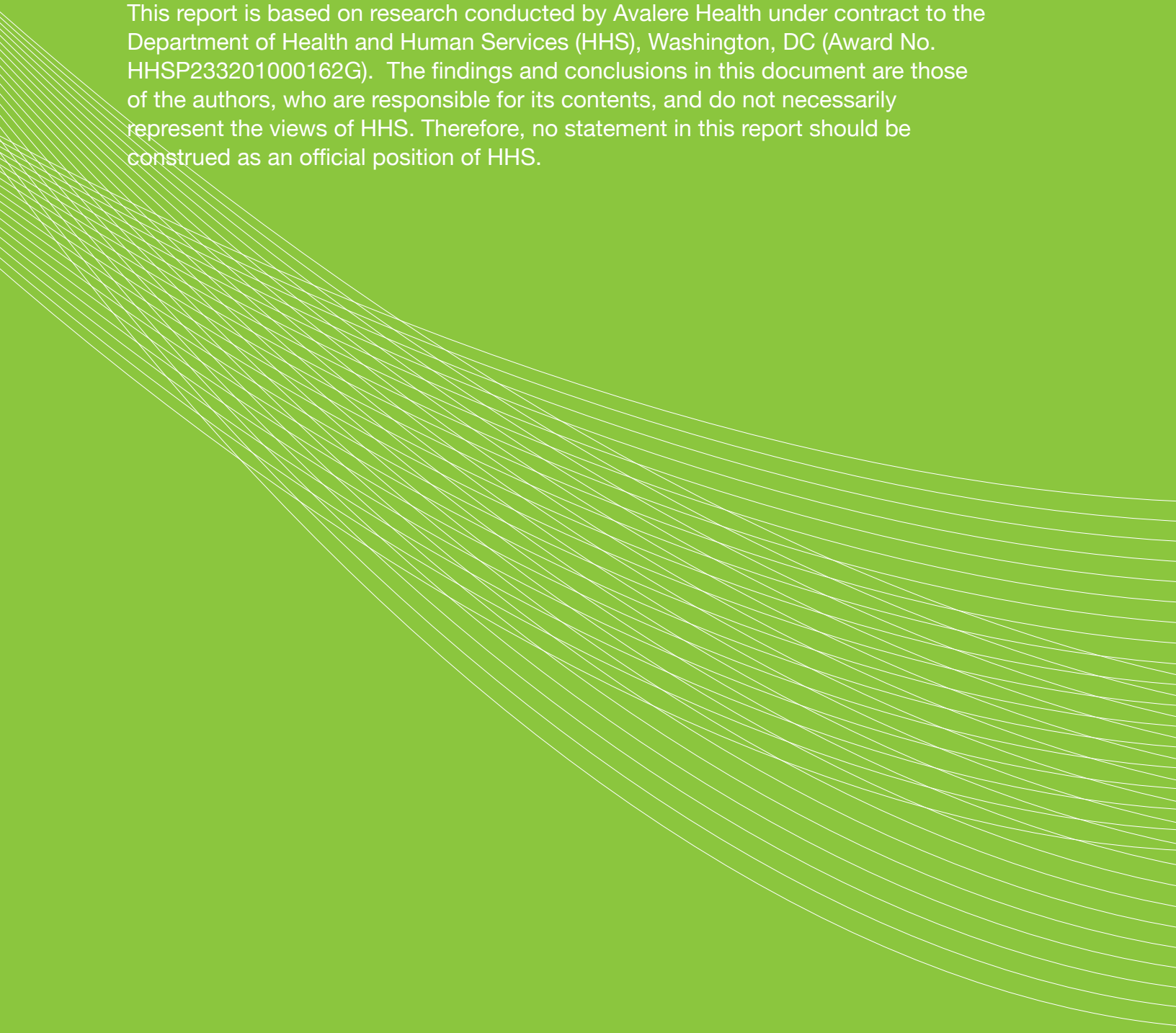


Table of Contents

Executive Summary

With funds available through the American Recovery and Reinvestment Act of 2009 (ARRA), the U.S. Department of Health and Human Services' (HHS) Assistant Secretary for Planning and Evaluation (ASPE) is developing a multi-payer claims database (MPCD) to support the growing effort in comparative effectiveness research (CER). HHS contracted with Avalere Health to develop three to six strategies for creating, operating, and maintaining the MPCD for CER purposes and to recommend one strategy for implementation. The results of this work will be used to inform the second stage of this project— the implementation phase—which will be conducted under separate contract.

Key findings from each work stream as well as our final recommendation and implementation plan are summarized in this paper. All work was completed over the course of fourteen weeks from January 21, 2010 to April 30, 2010.

To inform the design of the MPCD, we sought input from a wide range of stakeholders who have experience building, administering, and/or using existing claims databases. Discussions with over 50 stakeholders covered a range of relevant issues. They included how a MPCD can be used for CER; other valuable and common uses of claims databases; the most common user requirements (including data elements and linkages); and stakeholder recommendations for the structure and governance of the MPCD. In addition to these discussions, we conducted an analysis to identify key barriers and devise mitigation strategies for addressing them. We defined key barriers as issues that could significantly compromise the creation, operation, and maintenance of the MPCD effort if left unaddressed. We also conducted case studies of existing databases. The purpose of the case studies was to inform the various strategic dimensions of our design recommendation.

Based on this work, we found that a MPCD would have an incremental advantage over existing claims data sources due to its ability to link disparate sources, enabling research on broader populations that better reflect real-world clinical settings than do clinical trials. Not every comparative research question of importance is appropriate to answer via clinical trials. A MPCD presents a viable option for exploring a subset of those questions. Furthermore, because several payers already have efforts underway to collect claims data and have thus evaluated the benefits and limitations of using such data for research, the opportunity exists for the broad use of a MPCD in the short term, relative to other data sources such as patient registries and electronic medical records.

An array of barriers, ranging from operational to legal, could interfere with the implementation of a MPCD. Two barriers in particular—an inability to build a viable network of partners and violation of patient privacy protections—present the most precarious challenges for HHS and its contracting partners in the next phase of the project. Those barriers require considerable skill in navigating relationships with multiple stakeholder types, as well as an established understanding and mechanism for

addressing privacy concerns presented by HIPAA regarding the protection of patient data and the release of data for research purposes.

We categorized potential design options primarily by data source, i.e. where the claims data would come from. We present four models—state-based, plan-based, employer-based, and a hybrid. After careful consideration of pros and cons of each MPCD model, we recommend a hybrid approach with a state-based component in which the HHS core team works with a private data aggregator to link private data with state data and federal claims data. Ultimately, we believe a hybrid approach with a state-based component, in which a private data aggregator leverages its existing multi-payer claims and incorporates state-based data, is the most promising path forward. It would utilize the existing expertise of a private data firm that has navigated relationships with multiple payers, plans, and/or states in order to build its own datasets.

Implementation cannot stop at the construction of the database if the MPCD is to be successful. The effort must be tied to dedicated user training and education campaigns. It must also include periodic evaluation of the MPCD's impact and future direction. Through these activities, mechanisms for potential expansion will emerge. The MPCD should be initially built with the intent of future growth, allowing for incorporation of additional data to expand the types of research that can be conducted.

We note in conclusion that because CER funding through ARRA is the first step in a long-term national CER investment that was begun with the passage of the Patient Protection and Affordable Care Act, the potential uses for a MPCD are even more acute and the need to implement it with strong stakeholder support is especially important. The recommendations that we put forth from a strategic perspective—including a sample implementation plan—should be viewed as a guide, not as a blueprint. There are several issues that HHS must monitor closely with its contracting partner to ensure success, including the trade-offs associated with different models of data ownership and access.

Introduction

Comparative Effectiveness Research as a Strategy for Improving the U.S. Healthcare System

For many years, thought leaders and a range of healthcare stakeholders have identified comparative effectiveness research (CER) as a promising vehicle for improving Americans' quality of care and maximizing health care resources by identifying what works best and for whom.[1] A lack of sufficient data comparing therapies, preventive measures, and delivery system strategies is a major barrier in understanding and measuring value in the healthcare system. Consequently, a limited understanding of what constitutes true healthcare value prevents the nation from effectively addressing the dual problems of rising healthcare costs and inadequate delivery of high-quality care to patients.

A confluence of factors, including lack of real world data comparing treatments and care strategies, leads to almost half of adult patients in the U.S. receiving sub-optimal care.[2] CER can help patients and clinicians understand which option best fits an individual's needs and preferences. In addition to showing which treatments and strategies work best on average, it has the potential to identify which interventions work best in patients that are underrepresented in current clinical research (e.g., racial and ethnic minorities, children, and the elderly).

Despite the identified potential benefits of CER, steps needed to achieve lasting improvement include—the development and execution of a robust research agenda, the development of new data infrastructures, and the translation of research into clinical and patient decision support. These have not yet been realized. They require both broad stakeholder participation and long-term investment.

Driven by the increased interest in CER, the federal government committed to fund and support the infrastructure required for a new CER enterprise in the U.S. through a $1.1 billion investment in the American Recovery and Reinvestment Act of 2009 (ARRA). Of the $1.1 billion, the Act allocated $400 million to the Office of the Secretary of the Department of Health and Human Services (HHS), $400 million to the National Institutes of Health (NIH), and $300 million to the Agency for Healthcare Research and Quality (AHRQ). ARRA created the Federal Coordinating Council for Comparative Effectiveness Research (the Council) to advise the Secretary of HHS on how to invest its $400 million portion.[3]

The Identified Need for a Multi-Payer, Multi-Claims Database to Foster CER

In its June 2009 *Report to the President and the Congress*, the Council noted four major categories of potential CER investment and activities:

- Research
- Human and scientific capital
- CER data infrastructure
- Dissemination and translation of CER

The Council cited several examples of successful government efforts in three of those categories (conducting research, building human and scientific capital, and disseminating research), including those of AHRQ, NIH, and the Department of Veterans Affairs (VA).[4] Due to the immediate need for an improved CER data infrastructure, the Council recommended that this category be the primary destination for the Office of the Secretary's ARRA funding.

Based on feedback from public listening sessions, the Council noted that the types of data sources that would be required to constitute a new data infrastructure for CER include patient registries, distributed practice-based data networks, longitudinal electronic health records data, and the aggregation of existing administrative and claims

data. To date, the Office of the Secretary has acted to obligate funds to several HHS agencies to move forward on myriad projects encompassing these data source types.[5]

HHS identified a multi-payer, multi-claims database (MPCD) as one specific resource that could leverage existing claims data to conduct robust CER. Comments by several stakeholders during the Federal Coordinating Council's public listening sessions affirmed the value of a data source that brought to bear claims data from a range of public and private payers for the purpose of conducting robust analyses.[6] Several issues raised at these listening sessions informed the decision by HHS, acting through the office of the Assistant Secretary for Planning and Evaluation (ASPE), to move forward in calling for a strategic design to implement a MPCD at the national level. These points include the following.

- Because knowledge on what works best for which patients is a pressing and continuous need in medicine, it is practically impossible to answer these questions solely on the basis of large-scale randomized controlled trials. Though such trials the "gold standard" of clinical evidence, they are time- and resource-intensive and may be inconclusive for clinical transformation.
- A national claims data source leveraging existing information may represent a more immediate opportunity than investment in large-scale registries or electronic medical records. Furthermore, a claims database could be set up to eventually incorporate the clinical data contained in these non-claims data sources.
- A MPCD would allow for comparison of benefits and harms to populations in a real world setting, and can also evaluate the implications for sub-groups typically under-represented in clinical trials—thereby addressing two of the most pressing goals of CER.
- Although there are a range of existing databases housed within states and the federal government that could facilitate CER, each database has marked limitations. For example, the Centers for Medicare & Medicaid Services (CMS) repositories and warehouses are limited to Medicare and Medicaid populations, while current state-based all-payer claims databases are limited in geographic scope and variability of infrastructure design
- Data sources developed by private companies typically include a broader and more demographically diverse patient population than government datasets. However, these sources are cost-prohibitive for frequent use by many researchers.

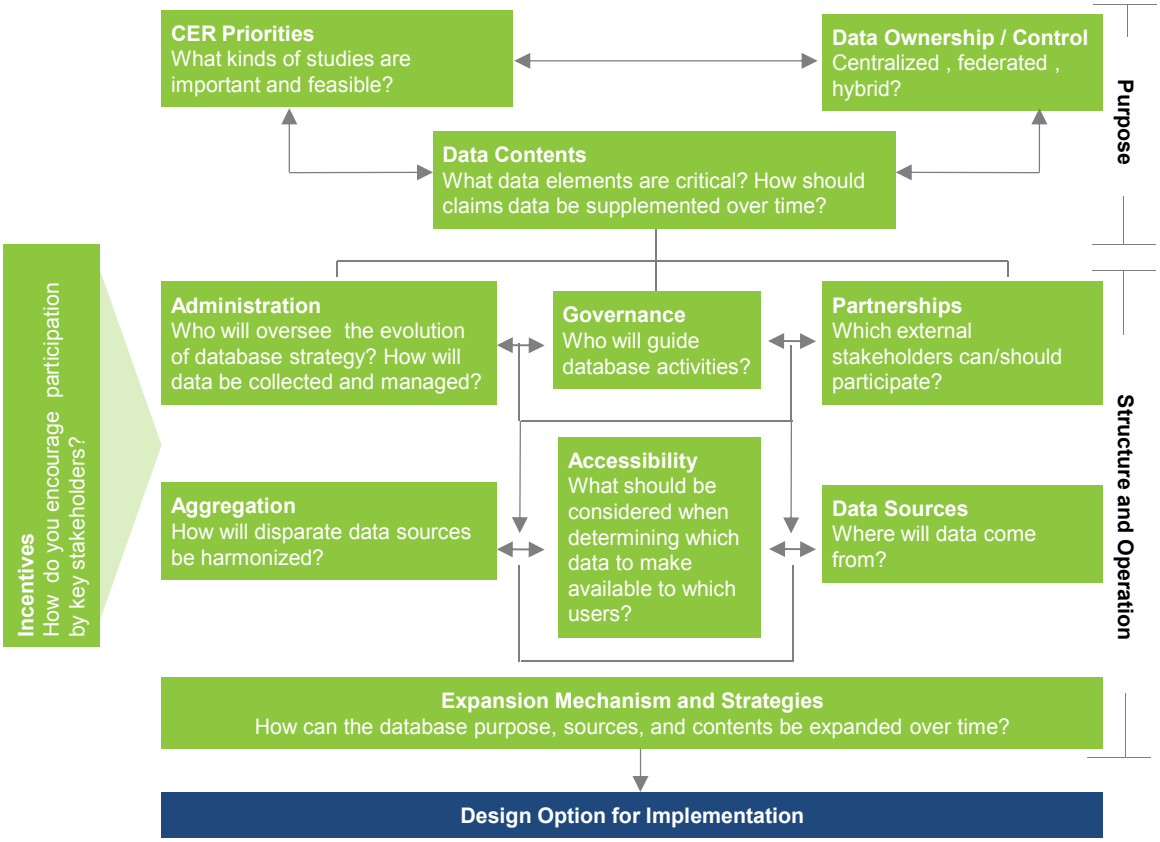
Recognizing that a range of potential options existed for implementing a MPCD, ASPE contracted with Avalere Health to develop three to six strategic designs over the course of 14 weeks, and to recommend one in particular. This recommendation will then be expanded through an initial implementation plan developed by Avalere. Our recommendation, though not binding, will inform the ultimate development of the MPCD. This Final Report presents our analytic approach, the key insights that informed the final set of MPCD models and the recommended option.

Approach and Methods

Initial Steps and Strategic Framework

As one of the first major tasks of the project, we worked in conjunction with HHS to develop a design framework that would structure our recommendation and implementation plan. The framework was intended to be a living document throughout the project that guided each subsequent task. We frequently modified the framework throughout the course of the project to reflect findings from stakeholder discussions and Avalere's secondary analyses. The framework contains 11 design dimensions (Figure 1) that represent the major considerations informing the purpose, structure, and operation of the MPCD.

FIGURE 1 Strategic design framework.



To arrive at a well-informed range of design options and a recommended approach, we considered the dimensions of the strategic framework as we conducted the various work streams including the case studies, identification of strategic, operational and legal barriers/challenges, and stakeholder discussions. These work streams provided key takeaways on the dimensions of the strategic framework that will warrant consideration as HHS and its potential contracting partners move forward in the next phase of MPCD implementation.

In synthesizing the learnings from each work stream and arriving at an ultimate recommendation, we consulted with several external experts on a range of issues. For example, legal experts helped identify statutory barriers to MPCD implementation and prioritize reasonable strategies to address them, while consultants from End Point Corporation ("End Point") provided expertise on the potential technical requirements for creating an MPCD.

Case Studies

Using the strategic framework, we analyzed key attributes of selected databases, files, and initiatives (collectively referred to as "database case studies"). The purpose of this exercise was to understand how the different types of aggregated data models that already exist in the research environment are being used, and how they could serve as potential models for the MPCD design option. We, in consultation with the HHS project team, identified fourteen existing databases or initiatives to include for review. Of the fourteen databases we analyzed (listed in Table 2 below), half comprised federal- and state-based initiatives, while the other half comprised private-sector initiatives.

We obtained information from a variety of sources including database documentation (such as data dictionaries when available) and primary interviews of sources involved in administering, overseeing, or otherwise involved with the databases.

TABLE 1 Database Case Studies

Federal/State-Based Data Initiatives	Private Sector Initiatives
CMS / Standard Analytic Files (SAFs)	Blues Health Intelligence
CMS / Medicaid Analytical Extract (MAX)	HMO Research Network
CMS / Chronic Conditions Warehouse (CCW)	IMS Health
CMS / Integrated Data Repository (IDR)	SDI/Verispan
AHRQ / Healthcare Cost and Utilization Project (HCUP)	Wolters Kluwer
FDA/CMS / Sentinel Initiative	Thomson Reuters/MarketScan
States / Regional All Payer Healthcare Information Council (RAPHIC)	Ingenix

Analysis of Strategic, Operational and Legal Barriers/Challenges

As a critical component of this project, we sought to understand the challenges that have faced others in establishing existing MPCDs. Therefore, we identified key barriers, or potential issues that could significantly compromise the effort to create a national MPCD, if left unaddressed. Equally important, we identified a range of mitigation strategies for each of the barriers identified. To identify key barriers and devise mitigation strategies for addressing them, we first considered the questions regarding such an endeavor. The questions dealt with defining an initial set of CER questions, securing partner participation, balancing inclusiveness and transparency, addressing

legal restrictions, mitigating operational challenges, and ensuring financial sustainability. As source material, we used presentations by thought leaders given at MPCD conferences, as well as state legislation and regulations. We reviewed federal laws, including the Health Insurance Portability and Accountability Act (HIPAA) and the Trade Secrets Act, as well as relevant CMS regulations. Additionally, we consulted thought leaders from states with MPCDs, and representatives from CMS and AHRQ for their insights into relevant MPCD barriers. We used our own subject matter expertise with claims databases and, as described below, we obtained information from our stakeholder discussions and incorporated their insights as well. As a final component of our approach, we consulted legal experts to identify statutory barriers and to prioritize reasonable and sound strategies to address them.

Stakeholder Discussions

We conducted 52 discussions with leaders of each of the types of stakeholders shown in Table 2 below. Stakeholders were identified in consultation with the HHS project team. The purpose of these discussions was to seek input from a wide range of stakeholders who have experience building, administering, and/or using existing claims databases. These discussions focused on how an MPCD can be used for CER, other valuable and common uses of claims databases, the most common user requirements (including data elements and linkages), and stakeholder recommendations for the structure and governance of the MPCD.

Because we conducted the stakeholder discussions in parallel to the completion of the case studies and the analyses of the barriers tasks described above, the team was able to integrate relevant stakeholder insights into those tasks as well. The breadth and depth of these conversations, along with the general enthusiasm with which many stakeholders agreed to participate in the project, allowed these discussions to serve as meaningful qualitative data points to inform our strategic design recommendations.

TABLE 2 Type and Number of Stakeholder Discussions

Stakeholder Type	Number of Discussions
Researchers and academics	7
Private data analysts	7
Government data sources	7
Public policy users	7
Employers and healthcare purchasers	5
Health plans	7
Provider organizations and practitioners	8
Life sciences manufacturers	2
Patient and consumer groups	2

Recommendations/ Implementation Plan

The final part of our approach entailed developing four strategic design options for consideration by the HHS project team and a recommendation of one option that, in our view, would be the best for implementation given the goals and timelines attendant to the database. Analyses of case studies and barriers, as well as stakeholder insights all served to shape our final recommendation. Furthermore, we conducted other analyses to supplement our primary methods as needed. For example, when discussing the potential uses of a MPCD, we joined stakeholder recommendations with a high-level internal analysis of the top 50 topics identified in the Institute of Medicine's June 2009 *Initial Priorities for Comparative Effectiveness Research* report to determine the extent to which a MPCD would provide value over existing data sources for a significant number of these topics.[7]

Finally, in putting forth our recommendation we consulted with database architecture experts at End Point, who assisted in clarifying the technical design choices.

To support HHS and other contractors who may be involved in the implementation phase, our final deliverable included an Implementation Plan that laid out recommendations on specific aspects of the MPCD's construction, contents, and governance as well as proposed timelines.

Key Findings

Purposes of a Large MPCD

Potential Advantages over Existing Claims Data Sources The ultimate goal of establishing a MPCD is to create a database that: 1) is nationally representative, 2) can track patients across settings and payers, 3) is conducive to linking multiple public and private data sources, and 4) would be available to researchers and other end-users to conduct comparative studies.

A central question that must be answered before a large investment is made in a MPCD is whether and how it would present the research community with a significant advantage over existing public and private claims and administrative data sources that are being used to conduct analyses. Based on stakeholder discussions, we garnered several common insights that articulate the potential value-add of a MPCD.

- A MPCD would link Medicare, Medicaid, and private data to improve the ability of researchers to track patients as they move across payers over time. Researchers cited the inability to track patients as major limitations to current research efforts.
- A MPCD would complement current clinical trial and other research efforts by more easily identifying appropriate patient populations and unique sub-populations to be included in clinical research.

- It could be designed in such a way that many more researchers and other end users could access it than is currently the case with private data sources that are cost-prohibitive. Expanded academic access could be realized through institutional subscription models.
- A MPCD could lead to both increased sample size and increased data completeness if data from multiple sources is appropriately linked and harmonized. Current one-payer claims sources are too often either broad but not granular (in terms of the patient information) or vice-versa.
- It would allow for comparisons encompassing both broad geographic areas and individual providers or facilities. Current data sources are often perceived as insufficient, especially at capturing geographic practice variation.

The identification of these potential advantages represented stakeholder support for the concept of a MPCD and allowed for a deeper probe into the question of whether the types of comparisons that are typically thought of under the purview of "CER" would benefit significantly from a MPCD.

Despite the identification of the potential advantages described above, an overarching theme among stakeholders was the notion that the true value and success of a MPCD would be largely dependent on the types of research questions that would be pursued. While many end-users were attracted to the idea of an MPCD, they were quick to note that the potential advantages of data harmonization and capturing larger patient populations would only be realized if the topics researchers pursued could directly benefit from those features. Furthermore, while private and public stakeholders whose datasets could potentially contribute to the MPCD noted the potential advantages of a multi-payer source, several noted that robust participation would be directly predicated upon the types of questions HHS seeks to answer through the endeavor.

Optimal Types of Research Questions Addressed Stakeholders noted that the definition of CER put forth by the FCC[8] implied four major categories of comparative research.

1. Comparisons of diagnostic, prevention, and treatment strategies
2. Approaches to healthcare financing, payment, and delivery
3. Analyses of geographic and utilization variations
4. Measurement of provider performance quality

Appendix A contains a non-comprehensive sample of stakeholder-identified potential research questions identified in each of these four categories. A large number of stakeholders believed that, based on the applications of current administrative and claims sources, a MPCD could most often and easily be used to address the third and fourth of these categories while there is greater uncertainty over the ability to address the first and second of these categories with claims data alone. However, a large portion of research needs that are typically considered as "CER" fall under these categories. For example, when the IOM produced its list of top 100 CER priority topics in June 2009, several topics comprised comparisons of diagnostic, prevention, and/or treatment

strategies. Therefore, several stakeholders, particularly academic researchers, questioned whether a MPCD—despite providing the advantages over existing data sources described above—would optimally address the topics that are under the purview of HHS through its ARRA funding commitment.

Two findings suggest that a MPCD would provide an advantage in addressing high-priority CER topics. First, some stakeholders did suggest specific instances in which a MPCD could be used to address topics related to prevention, diagnosis, and treatment strategies as well as approaches to payment, delivery and financing. For example, one health plan discussant reported that the expanded data on a large population could improve "approximate" outcomes information for high-cost therapeutic areas and rare diseases. A clinical practitioner also stated that while claims databases are more generally useful for hypothesis generating than hypothesis testing, this is not universally true, and research conducted with these databases can sometimes be sufficient to impact clinical practice, particularly around subgroups of patients.

Second, an Avalere analysis of the top 50 of the IOM priorities showed that a majority of these could be researched with claims data. Eleven of these topics could potentially be addressed sufficiently through existing administrative and claims data sources while 18 could benefit from a MPCD either as a primary data source or used in conjunction with other clinical data such as that contained in electronic medical records. Generally, we found that for a significant number of priority topics, there is at least some incremental value of a MPCD over the existing research landscape. This finding complements the potential uses laid out by stakeholders.

Furthermore, it should be noted that the IOM topics were presented in such a way that each topic comprises several potential research questions that could be answered using a variety of data sources, including a MPCD. For example, one IOM topic is "Compare the effectiveness of upper endoscopy utilization and frequency for patients with gastroesophageal reflux disease (GERD) on morbidity, quality of life, and diagnosis of esophageal adenocarcinoma." If a researcher wanted to know based on that topic whether higher endoscopy utilization leads to higher rates of adenocarcinoma diagnosis in a particular geographic region, that could be accomplished using current one-payer claims data sources. However, to understand how endoscopy utilization varies across geographic regions and GERD sub-populations, a MPCD would allow for appropriate identification of patient populations and variation. This illustrates the general notion that the range of potential CER questions is sufficiently large and continually expanding to ensure that demand for a MPCD would always exist in the research community.

The true value of a MPCD may not be fully realized until the claims data contained in it is joined with other types of data. Stakeholders routinely named plan benefit design information, demographic information, lab results, and additional clinical fields from electronic medical records as possible data types needed to eventually link with claims data and foster a broadened scope of CER. The potential to eventually link to other data types confers an additional advantage upon the prospect of a MPCD that is nationally

representative and widely available. Appendix B contains data elements identified by stakeholders as potential useful information for CER; while some of these elements are readily available in standards claims forms, some are not.

Ultimately, based on our analyses and extensive stakeholder conversations, we believe that a MPCD serves both an immediate purpose in assisting HHS in addressing the CER infrastructure expansion it seeks to enact through ARRA, and the long-term purpose of supporting a robust research enterprise. Eventually, the MPCD could support not just "CER" as it is traditionally thought of (i.e., comparisons of treatments), but also the broader view of CER put forth by the FCC and the Patient-Centered Outcomes Research Institute that was established by health reform legislation signed into law in March 2010. Whether it is ultimately used by the research community as a primary data source or a complementary analytic tool to existing clinical and administrative information, a MPCD clearly has the potential to garner support from many potential end-users to meet their needs.

Major Barriers to Success

Overview of Major Barriers For purposes of creating, operating, and maintaining a MPCD, we defined barriers as issues that could significantly compromise the MPCD effort if left unaddressed. Throughout our research and analysis, we identified nine barriers that will likely confront the development, maintenance, and operation of a MPCD. These barriers are the following.

- A misalignment of CER priorities and database capabilities
- An inability to build a network of partners
- Closed governance and insufficient management support
- An absence of (or inappropriate) standards to control data accessibility and use
- Inadequate provider protections
- Violations of patient privacy when accessing, warehousing and releasing data
- An absence of (or ineffective) methods to link data
- Failure to harmonize variables across different payers and settings of care
- Mismanagement of costs or inadequate funding

Table 3 below contains a summary of these barriers and descriptions of the mitigation strategies that are aimed to address them.

TABLE 3 Barriers and Associated Mitigation Strategies

Barrier	Mitigation Strategies
What would keep a potential partner or user from supporting / joining this initiative?	
Misalignment of CER Priorities and Database Capabilities	• Define target end-users and priority applications • Build channels for research community feedback • Invest in user education
Inability to Build a Network of Partners	• Seek to define partnership in a way that does not undermine the supplier's business model • Define incentives that encourage participation ("carrots") • Consider penalties ("sticks") for non-participation • Support user training and outreach • Provide sustained funding to academics to stimulate demand for data
Closed Governance and Insufficient Management Support	• Leverage (but do not burden) federal assets and expertise • Prioritize core group of stakeholders to participate in governance/management • Promote transparency of initiative
What legal and confidentiality factors, if unaddressed, could undermine this initiative?	
Absence of (or Inappropriate) Standards to Control Data Accessibility and Use	• Establish tiered access to correspond with appropriate applications and users • Articulate and transmit explicit evaluation criteria for data access requests • Define clear processes for data access
Violation of Patient Privacy	• Determine existing/needed legislative authority to establish MPCD • Identify CER questions that can be answered with de-identified health information—or limited datasets • Use combined approach—seek legislative authority to obtain protected information and pursue limited dataset option • Incorporate state-specific privacy rules in development of the MPCD
Inadequate Provider Protections	• Make strategic choices about what data to release and what aggregation levels should be used • Establish rules that govern the release of study results from private researchers
What operational challenges could limit the potential for the MPCD?	
Absence of (or Ineffective) Methods to Link Data	• Determine common variables for linking across payers • Mandate uniform encryption methods to be used by all contributors • Determine subset of claims that could be used for longitudinal analyses, if data can't be linked across different payers/providers • Develop longer-term strategy to ensure claims can be linked among payers/providers
Failure to Harmonize Variables across Different Payers and Settings of Care	• Determine uniform set of variables that can be collected among different payers • Consider strategies to include data elements from non-claim sources

Spotlight on Key Barriers and Mitigation Strategies Two barriers in particular pose the most significant challenges to this initiative: the inability to attract a viable network of partners and violation of patient privacy protections. Ineffectively navigating these barriers will significantly compromise the success of the MPCD. Specifically, if the right partners with the right datasets are unwilling to participate in the MPCD, it will not evolve from concept to implementation. Furthermore, if potential partners are not confident that the privacy of the patients in their datasets will be protected as required by law, they will not participate, and the MPCD will fail to launch. The patient privacy protections also have important implications for MPCD operations that must be followed to sustain the database. We focus on these two barriers for the remainder of this section, as they are key to shaping our recommendation.

Inability to Build a Network of Partners Moving the MPCD enterprise from concept to reality will require a committed network of partners who share the same vision and understand and accept collective responsibility for its success. Forging partnerships among a myriad of stakeholders that have different and sometimes conflicting interests will pose a fundamental and early challenge. It will be important to have partnerships with data contributors as well as with end users, such as academic and nonacademic researchers.

HHS should identify the right incentives and inducements to attract, secure, and sustain the right partners for the MPCD. Also, the initiative requires adequate consideration and incorporation of participants' viewpoints into the construct and operation of the MPCD. To address these barriers, HHS should employ certain direct and indirect economic incentives to bring data contributors and technical experts to the table. HHS should ensure partnership arrangements do not undermine the data contributor business models. Possible incentives to encourage participation could include preferred data access, access to CMS data, customized benchmarking, and compensation where appropriate.

HHS should build proficiency and confidence in the MPCD, using research ambassadors who will serve as spokespeople and trainers to generate end user interest in and capacity to utilize the MPCD for research. The initiative should also include the development of training modules for case studies that show real world application of the data. The intent of these efforts will be to encourage use of the MPCD and achieve early successes with its use.

Through stakeholder discussions, we learned that a key determinant to encourage data contributors to participate is the ownership model—that is, whether the database uses a centralized or distributed (or federated) model approach. Two aspects of these models differ—data ownership and control. Under a distributed model, the original data owners manage their data behind firewalls, protected by their own privacy and security procedures. Advantages of a distributed approach include potentially greater willingness for data owners to participate, as owners retain control of their data and have the right to opt out of particular analyses. However, efficient continued participation from data

contributors in a distributed model may work best with a clear, narrow focus for analyses. As research questions vary more widely or require greater degrees of patients' health information (PHI), a distributed model may lead to duplicative effort for the researcher or data aggregator in soliciting agreement from multiple data contributors. In contrast, advantages of a centralized approach are characterized by a more streamlined coordination of data dissemination.

Violation of Patient Privacy Protections Successful navigation of patient privacy depends on three elements: gaining access to patient data in a way that does not violate privacy laws; protecting the patient data once acquired; and complying with data release restrictions.

- **Privacy Laws:** As the holders of their beneficiaries' and PHI, healthcare payers (including Medicare) and providers are subject to privacy laws and regulations that restrict their ability to release this data. Specifically, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and implementing regulations, known as the Privacy Rule, restrict health plans, healthcare providers, and healthcare clearinghouses, referred to as "covered entities", from using and releasing PHI. Furthermore, some state laws are more restrictive than HIPAA. If a state law provides individuals with greater privacy protections, then the covered entities subject to the state law must comply with the state law. In addition to HIPAA, CMS must also adhere to the Privacy Act of 1974, which requires federal agencies to protect against any anticipated threats or hazards to the security or integrity of records, which could result in "substantial harm, embarrassment, inconvenience, or unfairness to any individual on whom information is maintained". To comply with the Privacy Act, CMS limits the disclosure of PHI to that which is necessary to accomplish the intended purpose of an agency activity, such as improving the Medicare program's operations, or ensuring beneficiaries quality health care.

HIPAA does allow covered entities to release PHI to their business associates. Business associates perform certain functions on behalf of the covered entity such as data processing, utilization review, and consulting. To gain access to PHI data for the MPCD, ASPE should seek a data aggregator with existing business associate contracts with several health plans, employers and providers. Using a data aggregator with many existing business associate contracts will also help in complying with the state laws that supersede HIPAA requirements, as they will already be subject to and know how to operate under these regulations. To integrate Medicare data into the MPCD, the data aggregator should establish a business associate agreement with CMS. To get this agreement it will be important to for the data aggregator to demonstrate that the MPCD fits within the purpose of an agency activity.

- **Protecting Patient Data:** HIPAA's Security Rule requires covered entities and their business associates to implement appropriate administrative, technical, and physical safeguards that protect against uses and disclosures not permitted by the Privacy Rule and that limit incidental uses and disclosures. Therefore, in choosing a

data aggregator, HHS should place high consideration on a candidate's ability to safeguard the data in a manner compliant with the HIPAA Security Rule.

- **Releasing Data:** HIPAA requires covered entities only to disclose the minimum amount of data necessary to accomplish the intended purpose of the research request. Therefore, datasets with PHI should only be released as a last resort, when de-identified information or limited datasets will not fulfill the purpose of the data request. Furthermore the requestor of PHI data must have the approval by an independent review board.

The minimum necessary rule should also guide the design of the accessibility model that dictates which users get access to which level of data for what purposes. For example, the MPCD accessibility model may dictate that only government researchers can gain access to PHI data with an IRB approval, while academic researchers cannot have access to this data. However academic researchers may have access to limited datasets if they sign a data use agreement, while the general public does not.

Balancing utility for CER with patient privacy protections and protections for data contributors should be considered early and should be an ongoing discussion between the federal government and its partner organizations. Most databases we examined as part of our case studies, use tiered access and structured data use agreements. The data elements released depend on the researcher, research purpose, or level of data contribution. In particular, private data vendors commonly offer customized rather than "one size fits all" solutions in the data elements they release in order to balance researchers' data needs with protections for patients and data contributors. For example, to protect proprietary pricing information, a data vendor might provide either charge data or payment data, according to researcher need, but not both. With respect to patient protections, a vendor might provide dates of service paired with less detailed geographic information, while still meeting requirements for statistically de-identified data for HIPAA compliance.

Strategies to increase efficiencies in the data use agreement process can support CER or other mission-driven research priorities. For example, plans to increase efficiency in a federal database initiative we examined included building a portal that identifies certain classes of questions for pre-approval and enables predetermined authorized users to submit queries or create data files for analysis.

To facilitate the use of limited datasets, we suggest that HHS require a data aggregator to design some limited datasets that will be readily available in response to anticipated common requests that fit into each of the research priorities of the MPCD. These "off the shelf" datasets will be a key tool in disseminating MPCD products to users to foster the demonstration of early research successes.

Potential Models for an MPCD

To develop design options for the MPCD, we focused on data sources, while considering that data could be aggregated with Medicare and Medicaid data from CMS. As part of our database case studies, we observed that different databases draw on various combinations of contributing data sources including, but not limited to, employers, payers, providers, plans, switches or clearinghouses, and pharmacies. The trade-offs associated with different data sources have implications for the population, range of services, and capacity for longitudinal research in the resulting database. For example, employer-based claims data tend to include carve-outs such as pharmacy benefits and mental health, in addition to medical record information in some cases, but do not include data on the uninsured. In contrast, switches or clearinghouses contain robust prescription drug utilization data and include data on the uninsured, but typically lack a defined eligible population to serve as a denominator for analyses.

We identified four options for consideration:

- a state-based approach,
- a health plan-based approach,
- an employer-based approach, and
- a hybrid approach.

The hybrid approach is intended to leverage existing relationships data aggregators have with current contributors; thus data could encompass a range of sources such as health plans, providers, employers, TPAs, PBMs, etc.

The state-based approach entails collecting data from a small number of states, possibly three initially, that currently have all-payer, all-claims databases. The mandatory nature of these initiatives allows for data to be relatively complete and reflect a full range of health care services provided.

The health-plan based approach includes collecting data from existing initiatives such as the HMO Research Network, Blue Health Intelligence, and United; the underlying premise for this option was to attract a manageable range of health plans that are diverse enough in benefit design and cover a large number of lives.

In contrast, the employer-based option involves collecting data from large employers already involved in such initiatives, including data from the Federal Employees Benefit Health Plan program from OPM. Within each of these three options, a contractor would be needed with primary responsibility for aggregating and standardizing data from a variety of contributors, creating standard output files, filling customized data requests, and developing supporting data documentation such as data dictionaries and file layouts.

The hybrid approach, involves the solicitation of a data aggregator that can leverage existing business associate agreements to contribute data that reflects a range of patient demographics, health plans, health services, and geographic areas. Central to the approach is a contractor that would be responsible for all the data aggregator responsibilities described above, as well as responsibility as a data contributor providing a significant amount of data to the MPCD. Our analysis of the pros and cons of each are described in the table below.

TABLE 4 Pros and Cons of Potential Options

Options	Pros	Cons
State-Based	• Comprehensive representation of number of lives within a state • Working towards standardization of numerous data contributors	• Misalignment of purpose—purpose of state initiatives is not to roll up into a nationally representative database for CER • States with current APCDs not nationally representative
Health-Plan Based	• Includes range of services covered under a health plan	• Lacks carve out services such as mental health • Change in health plan can impact ability to longitudinally track patients
Employer-Based	• Includes a range of services employees receive, including carve out services • Ability to track employees across plans and over time • Interest in claims data most closely aligns with other payers such as CMS	• Inherent bias in data from large employers with comprehensive health care coverage • Large number of small employers will be difficult to manage • Change in employers can impact ability to longitudinally track patients
Hybrid	• Combinations of different data sources can provide information on a full range of services and maximize number of lives • Use of a data aggregator could benefit from established relationships	• Need to ensure eligibility information is available and claims are adjudicated and paid • Need to ensure data aggregator can use data for identified CER purposes

As the table above illustrates, there are pros and cons to each approach. For example, the state-based approach contains the most comprehensive claims data in terms of health plans due to its legislatively mandated nature. However, state APCDs were not constructed with a national CER purpose in mind. Additionally, the ability to track patients longitudinally is impacted by the source option. For example, an employer approach can track employees for as long as they are employed by participating companies. Whereas, a health-plan based approach can track employees as long as they remain a beneficiary within the participating plans. Last, the type of services collected can differ depending on the source (i.e., services included within a benefit structure versus carve-outs).

The Recommended Option

Hybrid Approach with State-Based Component

To narrow in on a chosen option, we considered the political and technical feasibility, as well as the ability to demonstrate significant progress within the ARRA stipulated timeframe. Additionally, we informed consideration of these factors based on a comprehensive review of the stakeholder insights, barriers, and the database case studies. Based on this analysis we recommend a hybrid approach with a state-based component. This recommendation attempts to mitigate several challenges such as the ability to build on existing private-sector relationships, contracts, and aggregation platforms, while also allowing for the ARRA investment to be diversified and not limited to a single solution.

The hybrid approach aims to mitigate the challenges associated with building partnerships, protecting patient privacy, and aligning incentives of participants. Advantages of the hybrid approach are listed below.

- **Leverage existing relationships.** As we previously discussed, building relationships is a key challenge to bringing partners to the table for this initiative. To address this challenge, the hybrid approach is intended to obtain data from a data aggregator that can leverage its existing relationships with data contributors.
- **Mitigate data fatigue.** The option is also ideal in terms of trying to mitigate data fatigue that numerous stakeholders reported having, including the states and health plans.
- **Utilize existing standardization expertise.** A data aggregator is likely already addressing standardization issues to some extent within the current construct of its aggregated data. This approach will minimize the extent to which data sources have to be "newly" harmonized. Also, the time required to negotiate and achieve standardization can be significantly reduced by leveraging a data aggregator's expertise and previous efforts to standardize various data sources.
- **Align incentives.** Both CMS and a data aggregator could benefit from accessing and leveraging each other's data. For example, CMS could potentially be able to better understand the medical histories of its beneficiaries, while, a private data aggregator could integrate Medicare data into its offerings, enriching their database.
- **Leverage expertise to protect patient privacy.** The hybrid approach is designed to benefit from a data aggregator's existing expertise in this area while also benefiting from the existing trust data contributors place in private data aggregators to protect patient privacy as they provide PHI data. Finally, existing private data aggregators already operate within the constraints of the business associate contracts and HIPAA, so the data safeguards are already established.

With respect to investing in a state-based component, the mandatory nature of these databases makes them attractive in terms of completeness within a given state and provides an opportunity to learn from existing standardization approaches of a wide

range of health plans as well as other types of sources. In terms of incentives, most of the state APCD initiatives could benefit from obtaining Medicare data. Also, investing in state initiatives could provide needed resources to states to augment APCDs with non-claims data and advance the type of analyses that can be done. CMS could benefit from obtaining managed care data for its beneficiaries within particular states as well as other potential gaps such as drug data under a Retiree Drug Subsidy program. Additionally, state's APCDs collect SSNs so the ability to construct longitudinal files is attainable with these data as well. Finally, state initiatives also benefit from use of data aggregators with experience safeguarding and releasing data that could potentially be leveraged as well.

Aggregation/Ownership and Control

Central to the hybrid approach, is the concept that the data aggregator will be able to contribute a large amount of data based on its existing business model. To this end, it is assumed that the chosen data aggregator will have aggregation methods in place for its existing data contributors. Key to our recommended approach will be the contractor's ability and flexibility to include CMS data with its existing data. As CMS complies with rules governing release of its data, we recommend that certain Medicare and Medicaid data elements be centralized so that standard files/output can be created in the most streamlined manner possible and for pre-determined CER purposes. In terms of other customized requests that may require more PHI data, a decentralized approach could be used so that CMS retains control over how its data are being used. We believe an approach such as this regarding the ownership and control model incorporates the strengths of each option—either centralized or distributed—while mitigating the limitations. Therefore, the selected contractor will have to be able to accommodate both models.

Data Elements

Minimum data elements from claims and enrollment files needed to fulfill the purpose of the database must be defined at the outset. For the claims-based case studies that serve a variety of analytic purposes, including CER, data elements include patient demographics, enrollment information, health plan benefit information, and a range of medical, drug, and financial information. The level of protected health information (PHI) collected by a data aggregator will have implications for research capabilities. Most private data vendors use PHI to link patients across health plans or types of claims, and then de-identify to varying degrees before disseminating. Certain PHI data will be needed as it will facilitate the creation of longitudinal files that link claims across different provider types and possibly health plans. Thus, we recommend that the chosen data aggregator has access to PHI data.

Over time, the database should be expanded to include additional elements to enhance research capabilities, including non-claims data such as lab test results, race / ethnicity data, or medical records. The "stepping stone" nature of the initial MPCD is vital. As noted by many stakeholders, to adequately address some of the most pressing

comparative research questions, claims data will need to eventually be supplemented with these additional data points. Thus, we also recommend that the data aggregator have experience augmenting claims data with non-claims data such as lab results.

Governance and Management

The core team responsible for leading the database initiative must possess not only technical ability to execute, but also a reputation for credibility and objectivity. Many initiatives in both the public and private sectors use advisory groups to enable stakeholder input and guide privacy issues. Using a small core team with larger advisory boards can serve to balance the need for inclusiveness with the need for a small decision-making group. This collaborative approach represents a particularly crucial factor for success among databases that rely on voluntary data contribution. To forge successful partnerships with potential data contributors, for both the initial and expansion phase of a database, the core team must provide incentives for participation that meet the needs of potential partners.

The need for a strong core team that is located within the federal government is buttressed by insights provided by stakeholders. Many highlighted the need for the federal government to have a strong role in a multi-claims database effort. For example, employers stated that a strong federal role would be necessary to overcome stakeholder resistance. The most important function of the core team may be to foster trust among public and private sector participants. Over the course of our discussions, it became apparent that long-standing personal relationships can play a significant role in data holders agreeing to participate in efforts such as a MPCD. For example, one employer with past experience with a data collection effort for quality improvement included a multi-stakeholder governing board of members that worked in quality improvement for a long time and knew each other for many years. The discussant noted that the relationships and trust were crucial for understanding shared goals and coming to agreements.

We recommend that a core team in HHS be responsible for: leading the initiative and driving the vision, purpose and capacity of the MPCD; overseeing the data aggregator; and, promoting transparency of the initiative and communication strategy. In support of the core team, we propose the use of three advisory groups—a user group, a contributor group, and a privacy group. We recommend that these advisory groups be responsible for making recommendations to the core team on a variety of planning and operational issues described in the table below. We have also identified potential members of each group.

TABLE 5 Advisory Group Roles

Advisory Group	Purpose	Potential Members
User Group	• Represent the views of end users • Recommend accessibility model to the Core Working Team • Identify needed variables to address purpose of MPCD • Participate in biannual meetings to provide input on how the database can continue to meet its intended purpose and expand to meet evolving CER needs • Serve as ambassadors for initiative with intent of advancing MPCD utility	• Patients • Providers • Plans • Employers • CMS • States • Academic researchers
Data Contributor Group	• Represent the views of data contributors • Recommend accessibility model to ASPE and the Core Working Team • Recommend how use of data could be expanded to meet additional CER needs • Work with data aggregator to standardize data • Vet output file constructs that would meet research needs	• CMS • Representatives from data contributors (specific health plans, states, employers, etc.)
Privacy Group	• Provide support to ensure patient privacy protections are addressed • Provide statistician de-identification expertise to the User Group as it recommends an accessibility model	• Privacy lawyers • Experts specializing in the statistical de-identification methods

Partnerships

The initial partnerships established will be with a chosen contractor and CMS. However, HHS and the chosen contractor should ensure data contributor views are incorporated into the initiative as well. If the true source of the data is alienated from the process, the sustainability of the effort could be jeopardized. Therefore, if the data source is predominately health-plan based, HHS should include these stakeholders as part of either the stakeholder and/or the user group. Also, through fostering this communication, HHS should make deliberate efforts to communicate the purpose of the MPCD and associated data needs as well as seeking to understand what types of information could benefit the actual data contributors, not just the data aggregator. The purpose of establishing these relationships through user and/or data contributor groups will be not just to ensure that enough data can be brought to the MPCD initially, but to sustain the participation over time and potentially add contributors if needed.

Accessibility Model

A variety of files and tools should be made accessible to a range of users. A successful accessibility model will:

- Maximize access to data, while ensuring appropriate commercial and personal protections;
- Be flexible as data needs change over time; and
- Be easy to navigate to encourage use.

To ensure files and tools meet the range of user needs, we recommend that the accessibility model be developed with input from key user groups such as researchers. The model should define different products and target audiences such as those described below.

TABLE 6 Accessibility Model

Product	Description	Target Audience
Website	An internet accessible application that enables users to compare the quantity and cost of healthcare services of hospitals and providers	General public
Public User Files	"Raw" de-identified datasets which have not been manipulated prior to use	General public requesters[1]
Standard Outputs	Readily available, "off the shelf", datasets that are constructed in anticipation of common CER topics; for example, separate datasets may exist to meet the needs of users studying preventable readmissions and those studying geographic variation of certain diseases; depending on the user, these files may be de-identified, limited, or be research identifiable	Government researchers Academic researchers Private researchers
Custom Datasets	Datasets that are constructed in response to users' specific data requests when the standard outputs are not sufficient to meet their research needs; depending on the user, these files may be de-identified, limited, or be research identifiable	Government researchers Academic researchers Private researchers
Benchmarking Datasets	Benchmarking datasets would allow comparisons between each contributor and the aggregated inputs of the other contributors; preferred access datasets may aggregate information for each contributor category, i.e., Medicare, private insurers, etc.	Database contributors

[1] These requestors will likely be researchers who can manipulate large datasets.

In terms of setting prices for the various products, we recommend that HHS solicit a proposed price structure/schedule from potential contractors. In so doing, potential contractors will be able to identify what type of financial incentives will facilitate their participation in the MPCD effort. We also recommend that the contractor be responsible for monitoring the use of websites and other output such as standard analytic files; if demand is low, the accessibility model should be reassessed as well as the initiative's public relations efforts as discussed in more detail below.

Expansion Mechanisms and Strategies

Before the MPCD is expanded, we recommend an evaluation be conducted after the first year of implementation. The evaluation should address who is using the data and for what purposes, how CER priorities have been addressed with the data, how the MPCD has integrated with other efforts, including the Patient Centered Outcomes Research Institute mandated by the Patient Protection and Affordable Care Act, and what areas may need improvement. This assessment will inform the nature of future expansion. For example, depending on what type of data the aggregator can bring to the initiative, focus may be on acquiring other contributors in the future. Additionally, depending on the diversity of the type of data the aggregator can contribute, focus on expanding data elements such as EMR or other sources such as registries could be the focus.

User Education and Training Campaigns

User education and training campaigns should accompany the MPCD to ensure broad acceptance and use. Objectives should include helping end users clearly understand database's initial capabilities and building proficiency and confidence in the database by end users.

To achieve these objectives, the contractor should be responsible for developing a range of user education material including data documentation such as file layouts and data dictionaries. Key to the success of this endeavor will be the identification of researcher "ambassadors" who can serve as spokespeople and trainers to help peers understand capabilities, in addition to gaining feedback from their peers regarding the MPCD. Also, the contractor should be responsible for developing training modules, while ambassadors could create case studies that showcase real-world data applications. As an incentive for ambassadors to participate, preferred access could be granted along with a focus on some of the highest IOM CER priorities.

Key Takeaways

The key takeaways for each dimension of the strategic framework are summarized below.

Key Takeaways for Strategic Framework Elements		
Framework Element	**Takeaways**	**Lesson Learned For Recommended Strategy**
CER Priorities What kinds of studies are important and feasible?	The value of an MPCD lies in its large sample sizes, geographic representation, and capture of longitudinal information on a wide range of individual patients	Clearly communicate this specific benefit as part of overall value proposition
	An MPCD could be used to inform or complement clinical research	Clearly define intended purposes and additional uses of the database when linked with clinical data

Key Takeaways for Strategic Framework Elements		
Framework Element	**Takeaways**	**Lesson Learned For Recommended Strategy**
	An MPCD could be a valuable resource for priority CER topics, but usefulness is dependent on particular research question	Communicate a value statement that defines a clear and differentiated purpose
Data Sources Where should the data come from?	Each distinct data source (states, employers, plans) contains unique barriers that may be prohibitive	Solicit a data aggregator to bring a large amount of data to the MPCD using existing relationships with data contributors; data could be from a variety of sources
Incentives How do you encourage participation by key stakeholders?	Direct and indirect economic arrangements are required to bring participants to the table	A range of options exist; HHS should consider the feasibility of pursuing each (e.g., ability to provide compensation where appropriate, ability to confer preferred data access and access to CMS data)
Aggregation How will disparate data sources be harmonized?	Private data analysts are well versed in aggregating a range of data from different sources, payers, and health plans	Select a contractor that has aggregation methods in place and the ability to include Medicare, Medicaid, and state data
Data Ownership/ Control Centralized, federated, hybrid?	Most stakeholders support a distributed/ federated model as the most feasible, though several described a centralized approach as the "ideal"	Consider a unique approach that combines the strengths of each model
	Choosing between centralized and distributed ownership may require trade-offs between level of control and level of access	Practical levels of control and access that attract broadest possible group of participants may be achieved through a combined approach
Data Contents What data elements are critical? How should claims data be supplemented over time?	Stakeholders called for a wide range of data elements, including many not currently captured in claims	Think about linking administrative claims data to additional data sources as the database evolves
	Additional data that should eventually be incorporated include benefit design info, lab results, EMRs, demographics	HHS' contracting partner must implement the MPCD in such a way that it can flexibly be built upon over time

Key Takeaways for Strategic Framework Elements		
Framework Element	**Takeaways**	**Lesson Learned For Recommended Strategy**
Governance Who will oversee database activities?	A core team is needed to govern database activities and oversee the overall vision, through fostering stakeholder cooperation	Team should be a cross-section of HHS agencies
Administration Who will oversee the evolution of the database strategy? How will data be collected and managed?	Various advisory groups can assist the core team and foster buy-in	Different advisory groups are needed to represent the views of experts on distinct MPCD functions
Partnerships Which external partners can/should participate?	Employers and researchers expressed the most enthusiasm for this effort; potential data contributors (including health plans, private data analysts, provider organizations that have data, and employers) expressed the most concern about this effort and were often described by other stakeholders as a potential barrier to an MPCD effort	Leverage existing stakeholder interest and mitigate the concerns of potential private-sector data contributors to encourage their participation; this will entail providing clarity about potential new business lines the MPCD could offer and explaining how this effort will not impede existing business lines
Accessibility What should be considered when determining which data to make available to which users?	Maintaining patient privacy and confidentiality of data related to an organization's commercial interests (i.e., identification of costs, clinician, hospital, or health plan) are vital concerns and major potential barriers in building an MPCD	Assess the levels of data contributors may be willing to provide and gather their input on how to maintain privacy and confidentiality while maximizing access to data
	Successful navigation of patient privacy depends on gaining access to patient data in a way that does not violate privacy laws; protecting the patient data once acquired; and complying with data release restrictions	Utilizing a data contractor/ data source with mechanisms in place to address privacy concerns is the most promising path forward
Expansion Mechanism and Strategies How can the database, purpose, sources, and contents be expanded over time?	A data aggregator has the potential to bring a large amount of data to the MPCD that allows for a range of representation from different populations, geographical areas, payers, and plans	Conduct an evaluation prior to expansion to address who is using the data and for what purposes; determine what needs to be expanded—data sources and/or data elements

Implementation Plan

To implement the MPCD in the ARRA specified timeframes, we developed a high-level implementation plan that has four primary phases—Pre-Kick Off, Planning, Infrastructure-Building, and Evaluation and Expansion. The table below summaries key tasks and timeframes associated with these four phases.

Table 7 Key Tasks and Timeframes

Phases	Key Tasks	Timeframes
Pre-Kick Off	• Agree on purpose • Finalize delineation among stakeholder board roles and responsibilities and contractors • Select a contractor	Prior to and up to contract selection
Planning	• Develop plans necessary for the administration and governance of the MPCD initiative	Week 1 to Month 5
Infrastructure - Building	• Implement plans to create the MPCD and ensure that the data is collected and ready to be accessed appropriately	Month 6 to Month 12
Evaluation and Expansion	• Conduct an evaluation to ensure the MPCD is aligned with the intended purpose and continues to expand and grow to meet a broader set of needs	Year 1 to Year 2

Within each of these four phases, we developed more specificity to accomplish two goals.

- Ensure the MPCD is up and running within a year from the time the contract is awarded
- Ensure there is an evaluation period to inform expansion decisions

Table 8 Details on Pre-Kick Off Phase

Pre-Kick Off		
Task	**Role**	**Timing**
Establish core working team	HHS	Pre-contract award
Validate / agree upon MPCD Charter	Core Team	Pre-contract award
Finalize Advisory Group Functions and Terms of Participation	Core Team	Pre-contract award
Establish parameters for state awards	HHS	Pre-contract award
Award Contract (Non State)	HHS	Kick off
Release RFP for State Awards	HHS	Kick off

Table 9 Details on Planning Phase

Planning Phase		
Task	**Role**	**Timing**
Begin discussions on revising Business Associate agreements with CMS	Contractor & HHS	2 weeks from award of contract
Begin discussions on revising Business Associate agreements with existing data partners	Contractor	2 weeks from award of contract
Call for nominations for advisory groups	HHS	1 month from award of contract
Grant Applications for State Awards Due	Interested States	1 month from RFP release
Finalize State Awards	HHS	2 months from RFP release
Nominations for advisory groups due	Public	2 months from award of contract
Announcement of advisory group members	HHS	3 months from award of contract
Finalize Business Associate agreements with data sources	Contractors & others	3 months from award of contract
Convene Initial Meeting of advisory group members	HHS	4 months from award of contract
Identify potential "ambassadors" for user training	Contractor	5 months from award of contract
Craft user education materials (awareness)	Contractor	5 months from award of contract

Table 10 Details on Infrastructure-Building Phase

Infrastructure-Building Phase		
Task	**Role**	**Timing**
Create common data dictionary between data contributors; Define "standard data elements" for submission	Contractor (consultation with Core Team and advisory groups)	6 months from award of contract
Finalize accessibility model	Contractor (consultation with Core Team and advisory groups)	6 months from award of contract
Map out user training program	Contractors & Ambassadors	7 months from award of contract
Launch user education "campaign"	Contractor	8 months from award of contract
Present data collection and aggregation plan	Contractor	8 months from award of contract

Infrastructure-Building Phase		
Task	**Role**	**Timing**
Collect Medicare data and link with its existing health plan, employer, provider, or other types of data	Contractor	10 months from award of contract
Launch user training program	Contractor & Ambassador	12 months from award of contract

Table 11 Details on Evaluation and Expansion Phase

Evaluation and Expansion		
Task	**Role**	**Timing**
Evaluate how the MPCD has been used to address the CER priorities	Third-party	1 year and 6 months from award of contract
Articulate priorities for expansion based on cost-benefit analysis of options	Core Team & Contractor	1 year and 6 months from award of contract
Develop phased plan for expansion	Contractor	1 year and 8 months from award of contract
Implement phased expansion	Contractor	1 year and 10 months from award of contract

Conclusions

The CER funds in ARRA—which are to be obligated and spent within two years of enactment—appear to be an initial "down payment" on a long-term federal commitment to CER. On March 23, 2010, President Obama signed the Patient Protection and Affordable Care Act into law, greatly expanding access to healthcare and setting the foundation for long-term improvement in the quality of care through the enactment of a broad range of delivery system reforms. Among these strategies is the establishment of the Patient-Centered Outcomes Research Institute, a non-profit, non-governmental entity that will have dedicated public- and private-sector financing and participation through 2019 and is expected to eventually become a central coordinating hub of CER activities in the United States.[9]

While the research agenda and preferred methodological standards of the PCORI are yet to be determined, the various infrastructure enhancement projects that HHS has undertaken through ARRA funding will likely serve as a valuable foundation for future national CER activities. Among these infrastructure enhancements, the MPCD represents an important step in harnessing the power of public and private claims data to provide insights that lead to better care. While stakeholder insights and other analyses illuminate several of the potential challenges for a large multi-claims database, the prospects of gaining insight on the comparative effectiveness of therapies, delivery, and

financing strategies give the MPCD a valuable potential advantage over existing data sources.

Although many different potential models of a MPCD exist, each with distinct pros and cons, we believe a hybrid approach – in which a private data aggregator works to link private data with Medicare, Medicaid, and state data – represents the most promising and feasible approach. The private contractor that is ultimately chosen to lead this effort must not only have the technical expertise to shepherd this large-scale aggregation effort, but must also possess relationships with a range of stakeholders whose support will be necessary to achieve success.

Appendix A. Potential Research Questions

Note: While all research questions below may not be addressed using claims alone, discussants made these suggestions working with the assumption that other data, such as electronic health records, plan information, or specific provider information, may become available in the MPCD in the future.

Benchmarking and Reporting

- How does performance of physicians employed by a hospital compare with that of independent physicians?
- How do years of experience impact clinician performance?
- Why are some outlier physicians "so good" despite having more or less experience than other high-performing clinicians?

Approaches to Payment, Delivery, and Financing

- What is the optimal ratio of doctors to patients in a hospitalist unit?
- What forms of medical homes and accountable care organizations (ACOs) best reduce annual per capita care spending and disability-inducing quality defects?
- What approach to prescription drug tiering would best reduce annual per capita care spending and improve worker productivity?
- What is the likelihood of complications of a certain treatment in different settings?

Prevention, Diagnosis, and Treatment Strategies

- What is the relative effectiveness of drugs within a specific therapeutic class?
- What is the relative safety of different interventions for a specific condition?
- What is the relative effectiveness of different routes of administration (e.g., oral vs. intravenous)?
- What is the relative cost effectiveness of medical interventions compared to surgical interventions for a specific condition?
- What is the relative clinical and cost effectiveness of collecting a thorough patient history and physical exam compared to conducting a series of high-tech diagnostic tests?
- What are the combinations of drugs that present the greatest risk for particular patients with particular conditions?
- What are the characteristics (e.g., race) that make patients more or less likely to benefit from a certain treatment?

B. Data Elements

Over the course of the stakeholder discussions, many discussants identified specific data elements that would be valuable within a multi-payer claims database to support CER and other research, shown below. While researchers and end-users requested that the MPCD capture these particular elements in order to conduct longitudinal studies, cost of care analysis, and other research, the MPCD would not need to release all of these elements in order to be valuable.

While several of these elements are currently standard fields found on claims, or easily obtained from supplementary sources such as eligibility files, stakeholders also indicated a desire to link to or include additional data on clinical outcomes and more specific patient, provider, and plan information. In some cases, the desired data elements are already available but stakeholders felt that the data could be improved (e.g., through more standardized reporting) to better support research. In particular, stakeholders commonly cited the following as elements that should be added or improved in a MPCD: lab test data, eligibility information, and identification of provider information.

Please note that the list below is not intended to be comprehensive but merely illustrative of the range of components stakeholders frequently suggested would be useful in a MPCD.

Standard Elements

Category	Potential Elements
Patient Demographics	• Date of birth/age • Gender • Geographic location
Patient Identifier Information	• Unique patient identifier
Medical Information	• Procedure and diagnostic codes • Whether diagnosis is present on admission • Dates of service • Diagnosis related group (DRG)(inpatient)
Provider Information	• Place of service • Physician ID • Physician characteristics (e.g., specialty, location) • Facility type and location • Granular facility information (e.g., department)
Health Plan Features	• Plan type (fee-for service, health maintenance organization, preferred provider organization)
Financial Information	• Payment amounts • Patient copayments, deductibles, coinsurance • Facility charges within a DRG payment

Category	Potential Elements
Drug Information	• National drug code (NDC) • Quantity and days supplied

Elements Where Some Data is Available but May Be Improved

Category	Potential Elements
Race and Ethnicity Data	• Either self-reported or reported by clinician[1]
Enrollment Information	• Eligibility information[2]
Provider Information	• National provider identifier (NPI)

Supplementary Elements

Category	Potential Elements
Lab Results	• Body mass index (BMI) • Blood pressure • Lipid profile • Hemoglobin A1c (HbA1c)
Additional Medical Information	• Link to National Death Index (NDI) for cause of death
Additional information on plans and payment approaches	• Benefit design • Information on presence of new delivery or payment approaches, such as medical homes and ACOs
Severity of Illness/Patient-Reported Outcomes	• Disease severity/staging • Pain scores
Functional Measures	• Functional assessment measures obtained from patient assessment instruments in post-acute care settings
Patient Characteristics	• Patient income • Marital status • Family history

1 Reporting methods should be standardized to ensure comparability in order for this variable to be useful.
2 Stakeholders reported varying difficulty in obtaining eligibility information from different public and private payers.

Endnotes

[1] G.R. Wilensky, "Developing a Center for Comparative Effectiveness Information," *Health Affairs* 25, no. 6 (2006): w572-w585; U.E. Reinhardt, "An Information Infrastructure for the Pharmaceutical Market," *Health Affairs* 23, no. 1 (2004): 107-112; Congressional Budget Office, "Research on the Comparative Effectiveness of Medical Treatments," December 2007.

[2] E.A McGlynn et al., "The Quality of Health Care Delivered to Adults in the United States," *New England Journal of Medicine* 348, no. 26 (2003): 2635-2645.

[3] H.R. 1, American Recovery and Reinvestment Act of 2009, signed into law February 12, 2009. Available online at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h1enr.pdf.

[4] Federal Coordinating Council for Comparative Effectiveness Research, *Report to the President and the Congress*, June 30, 2009. Available online at http://www.hhs.gov/recovery/programs/cer/cerannualrpt.pdf.

[5] Presentation by Patrick Conway, expert advisor to Assistant Secretary for Planning and Evaluation (ASPE/HHS), Institute of Medicine Roundtable on Value and Science-Driven Care, April 1, 2010. Available online at http://www.iom.edu/Activities/Quality/VSRT/2010-APR-01/Agenda.aspx.

[6] Transcripts and public comments from the Federal Coordinating Council for Comparative Effectiveness Research's listening sessions are available at http://www.hhs.gov/recovery/programs/cer/comments/2009/apr-june/index.html.

[7] The Institute of Medicine (IOM) issued a report in June 2009 as mandated by the American Recovery and Reinvestment Act of 2009. Based on public input and expert opinions, the IOM released its recommendations for Top 100 priority research topics. The list of topics is available at http://iom.edu/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx.

[8] CER is the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in "real world" settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances. To provide this information, CER must assess a comprehensive array of health-related outcomes for diverse patient populations and sub-groups. Defined interventions compared may include medications, procedures, medical and assistive devices and technologies, diagnostic testing, behavioral change, and delivery system strategies. This

research necessitates the development, expansion, and use of a variety of data sources and methods to assess comparative effectiveness and actively disseminate the results.

[9] H.R. 3590, Patient Protection and Affordable Care Act, signed into law March 23, 2010. Available online at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h3590enr.txt.pdf.

Corresponding Author:
Nora Hoban, nhoban@avalerehealth.net

Contributors:
Riaz Ali
Valerie Barton
Shamonda Braithwaite
Alejandra Herr
Nora Hoban
Dinesh Kumar
Sara Sadownik
Alana Tucker

Avalere Health LLC
1350 Connecticut Ave. NW
Suite 900
Washington, DC 20036

www.avalerehealth.net