

# Expert Meeting on the Potential Value of Centers for Medicare and Medicaid Services Data as a Resource for National Institute on Aging Studies

The National Academies of Sciences, Engineering, and Medicine Building  
2101 Constitution Ave. NW, Washington, DC, Room 118

May 4, 2016

*Revised 8/19/2016*



This meeting summary was prepared by Anneliese Ebersole, Rose Li and Associates, Inc., under contract to the National Institute on Aging, National Institutes of Health (HHSN271201400038C Requisition No. 4096425). The views expressed in this document reflect both individual and collective opinions of the meeting participants and not necessarily those of the National Academies of Sciences, Engineering, and Medicine, National Institute on Aging, National Institutes of Health, or U.S. Department of Health and Human Services. The following individuals are gratefully acknowledged for reviews of earlier drafts of this report: Julie P. W. Bynum, Michael Chernew, Constance Citro, Martin Gaynor, John Haaga, Chandra Keller-Allen, Rose Li, Krisztina Marton, Allison Oelschlaeger, Nancy Tuveson, and Alan Zaslavsky.

## Table of Contents

<b>Acronym Definitions</b> .....	<b>iii</b>
<b>Meeting Summary</b> .....	<b>1</b>
<b>MEDICARE CLAIMS DATA FOR RESEARCH</b> .....	<b>2</b>
The Challenges of Using Medicare Data for Research .....	3
Available Resources for Linking Survey Data to Medicare Claims .....	4
Future Strategies to Facilitate CMS Linkages .....	4
<b>MEDICAID DATA AS A RESEARCH RESOURCE</b> .....	<b>5</b>
Medicaid Data Sources.....	5
The Challenges of Using Medicaid Data for Research.....	6
The Linkage of Medicaid to Other CMS Sources and Survey Data.....	7
<b>USING PRIVATE HEALTH INSURANCE CLAIMS DATA FOR RESEARCH</b> .....	<b>7</b>
Potential Data Sources of Private Health Insurance Claims.....	7
Linking Private Insurance Claims Databases with External Datasets .....	9
<b>THE CMS PERSPECTIVE ON POTENTIAL OPPORTUNITIES FOR RESEARCH</b> .....	<b>9</b>
<b>IMAGING DEMENTIA—EVIDENCE FOR AMYLOID SCANNING</b> .....	<b>11</b>
<b>USING MEDICARE CLAIMS WITH ADD-INS</b> .....	<b>11</b>
<b>Panel Discussion</b> .....	<b>13</b>
<b>Appendix 1 Workshop Agenda</b> .....	<b>15</b>
<b>Appendix 2</b> .....	<b>17</b>
<b>List of Meeting Participants</b> .....	<b>17</b>

## Acronym Definitions

Acronym	Acronym Definition
APCD	All-Payer Claims Database
BHI	Blue Health Intelligence
CAHPS	Consumer Assessments of Healthcare Providers and Systems
CBES	Children’s Health Insurance Budget and Expenditure System
CCW	(CMS) Chronic Conditions Data Warehouse
CHIP	Children’s Health Insurance Program
CMS	Centers for Medicare and Medicaid Services
DME	durable medical equipment
DRG	Diagnosis-Related Group
DUA	Data Use Agreement
FFS	Fee for-Service
HCCI	Health Care Cost Institute
HEDIS	Healthcare Effectiveness Data and Information Set
HIC	health insurance claim
HIPAA	Health Insurance Portability and Accountability Act
HRR	Hospital Referral Region
ID	identification
IDEAS	Imaging Dementia-Evidence for Amyloid Scanning
LDS	Limited Data Set
MA	Medicare Advantage
MAX	Medicaid Analytic eXtract
MBES	Medicaid Budget and Expenditure System
MCBS	Medicare Current Beneficiary Survey
MDS	Minimum Data Set
MedPAR	Medicare Provider Analysis and Review
MMIS	Medicaid Management Information System
MMLEADS	Medicare-Medicaid Linked Enrollee Analytic Data Source
MSIS	Medicaid Statistical Information System
NHS	Nurses’ Health Study
NIA	National Institute on Aging
NPI	National Provider Identifier

<b>Acronym</b>	<b>Acronym Definition</b>
OASIS	Outcome and Assessment Information Set
PDE	Part D Prescription Drug Event
PECOS	Medicare Provider Enrollment, Chain, and Ownership System
PET	positron emission tomography
PPO	Preferred Provider Organization
PSID	Panel Study of Income Dynamics
ResDAC	Research Data Assistance Center
SEER	Surveillance, Epidemiology, and End Results Program
SES	socioeconomic status
SMEC	Summary Medicare Enrollment and Claims
SNF	skilled nursing facility
T-MSIS	Transformed Medicaid Statistical Information System
VRDC	(CMS) Virtual Research Data Center

## Meeting Summary

The National Institute on Aging (NIA) commissioned the Committee on National Statistics of the National Academies of Sciences, Engineering, and Medicine to convene an expert meeting on May 4, 2016, to discuss the value of using data from the Centers for Medicare and Medicaid Services (CMS) and private health insurance claims as a research resource for NIA-funded studies. Research made possible by linkages to CMS data facilitates CMS purposes as well. Participants examined the role that CMS data can play in enhancing NIA behavioral and social research, the research opportunities that exist, and the limitations associated with these data. They also discussed examples of common and novel uses of CMS data, the different types of data available to researchers, as well as the feasibility of linkages to other data infrastructure studies.

The NIA funds several longitudinal, population-based studies that are used by researchers from a broad range of disciplines to evaluate the health of an aging American population. These studies provide rich information on health behaviors and risk factors, as well as self-reported health. There is potential to study participants selectively based on particular health events and to follow downstream consequences. The NIA currently encourages many studies to include the collection of biomarkers so that the range of outcomes to be investigated may be expanded.

Many NIH-funded studies—such as the Health and Retirement Study, National Health and Aging Trends Study, and the Popular Study of Income Dynamics—currently provide linkages to other sources of information (e.g., Medicare, Medicaid, or private health insurance data). Such linkages leverage existing investments, add valuable exposure and outcome measures, and maximize the value of these population-based studies. The purpose of this meeting was to explore strategies for making the process of creating linkages to CMS data easier for both existing and potential NIH-funded studies, and to identify other potential research uses of CMS data.

Several suggestions emerged that could facilitate broader data access and use across all available data resources containing information about the health of the U.S. population.

- Facilitate data access agreements at a higher level (e.g., center, institution, or funding agency) rather than at the investigator or individual study level. Start-up costs to data access can be prohibitive to individual investigators. For example, resources could be pooled to make linkages of basic data available to multiple grantees, rather than have different research groups individually negotiate for data access.
- Encourage replication by investigating similar research questions with multiple strategies, approaches, and investigators to effectively move the science forward.
- Document and share successes, challenges, and solutions that researchers have experienced when trying to link datasets.
- Provide guidance to new investigators to focus on specific health events or conditions and to work with data warehouses to obtain specific information, rather than mine comprehensive datasets for as much information as possible.

- Develop an open source platform for researchers to share protocols, code, data, or experiences with linkages.
- Because socioeconomic status (SES) and income data are not available in CMS datasets, researchers should be encouraged to obtain this information via self-report from study participants.

Suggestions also emerged for possible approaches for the CMS to facilitate its relationship with the research community.

- Streamline and simplify the process for granting data access to researchers via the CMS Virtual Research Data Center (VRDC). Consider structuring access at a higher level (e.g., all NIA or NIH grantees) rather than by individual research project.
- Improve operational efficiency by streamlining and simplifying the process for granting the data.
- Streamline the ResDAC user experience and examine the flow of the data request and delivery to improve operational efficiency.
- Develop a process for researchers to ask questions and interact with CMS representatives directly to troubleshoot data access issues.
- Provide outreach to new investigators to educate them about the available data so they are able to write effective research applications incorporating the use of CMS data. CMS and others who have experience using these data are encouraged to plan and implement training sessions, webinars, and online materials to disseminate information about the opportunities, challenges, and best methods for using these data.
- Make Medicare Advantage data accessible to researchers because these data are critical to understanding publicly financed programs.

The meeting agenda and list of participants are included as appendices 1 and 2. The remainder of this report summarizes the highlights from each session and key discussion points.

### **MEDICARE CLAIMS DATA FOR RESEARCH**

Julie Bynum, Dartmouth Institute for Health Policy & Clinical Practice, and the Geisel School of Medicine

Medicare data have historically been the only complete, consistent source of data on health care spending in the United States. Medicare consists of Parts A, B, C, and D, which include hospital, medical, and prescription drug insurance. Medicare Advantage (Part C) is a plan offered by a private company that contracts with Medicare to provide Part A and Part B benefits. Medicare Advantage includes health maintenance organizations (HMOs), preferred provider organizations (PPOs), and private fee-for-service (FFS) plans. Medicare eligibility was expanded in 1972 to include younger individuals with permanent disabilities who receive Social Security Disability Insurance and people of all ages with end-stage renal disease or permanent kidney failure requiring dialysis or a kidney transplant. In 2014, Medicare covered more than 55 million Americans (about 16 percent of the U.S. population). Approximately 17 percent of the total Medicare beneficiaries were disabled.

Medicare data are an important resource for researchers for several reasons: (1) data collection is tied to physician billing; (2) the data include a large, diverse population and enables researchers to investigate questions about the entire United States; (3) the data allow for the ability to investigate long-term outcomes at a low cost; (4) the data are ideally suited to investigate outcomes that are relevant to clinical care; and (5) linking Medicare data to large, longitudinal studies could leverage the value of the longitudinal studies and enable them to continue.

Linking Medicare administrative data with large national studies, such as those funded by the NIA, provides important and diverse research opportunities. Research can be conducted at different levels of analysis including patient and physician characteristics, the type of service provided, medication use, specific diagnoses, or the population level. One example of successful data linkage for research is the Nurses' Health Study (NHS), whose participants are all Medicare beneficiaries. Bynum and colleagues examined a subset of participants with objective cognitive functioning measured longitudinally over 10 years. Linking NHS and Medicare data effectively combined information about cohort utilization and services received with self-reports of functional status.

Other examples of how linkages with Medicare claims data have been useful to researchers include studies of:

- The influence of patient characteristics on rates of mammography screening by race
- Geographic patterns of dementia drug use across Hospital Referral Regions (HRRs)
- Treatment patterns by physician specialty, controlling for physician and patient characteristics
- Regional variation in diagnostic practices
- The influence of social networks on diffusion of prescribing practices
- Variability in end-of-life expenditures
- Adverse events associated with combination or multiple prescription drug use

### **The Challenges of Using Medicare Data for Research**

The Medicare data files contain information on the enrollment status and health care utilization and expenditures of Medicare beneficiaries. There are a number of limitations to the data. For example, the Medicare data cover acute care and disease care; they do not include long-term care, and the home care information that exists in the data is not equivalent to long-term support for home care. Another issue is missing data for some subpopulations. About 30 percent of the Medicare population is currently in managed care, and there is differential uptake across regions and significant national variation in this population. Limiting study populations to FFS can lead to selection issues. Determining the accuracy of a diagnosis is another challenge. Patients must present to the health care system and have their diagnosis recognized. This is particularly problematic with dementia patients. Physicians often do not make accurate diagnoses, and the diagnostic codes used do not always reflect current diagnostic categories.

Medicare data files that have been used by researchers include the following:

- Part A (hospital insurance benefits) covers inpatient hospital care, skilled nursing facility (SNF) stays (not long-term care), and hospice care. The Medicare Provider Analysis and Review (MedPAR) file contains Part A data.
- Part B (physician/supplier file) covers physician services (including professional fees and labs), outpatient care, durable medical equipment (DME), and some home health care. The DME file contains claims data submitted by DME regional carriers. The Carrier File contains claims data submitted by noninstitutional providers and largely consists of physician claims records.
- Part C (Medicare Advantage) includes encounter data on the services provided to enrollees. The CMS is in the process of making sure the data are complete and accurate before making them available for use by the research community.
- The Medicare Part D Prescription Drug Event (PDE) File contains a summary of prescription drug costs and payment data. Information is available about drug events, which is not necessarily what was prescribed. All Medicaid beneficiaries and only 50 percent of Medicare beneficiaries are enrolled in Part D.
- The Summary Medicare Enrollment and Claims (SMEC) file includes Medicare cost and claims data from multiple Medicare services files.
- The Medicare Current Beneficiary Survey (MCBS) captures expenditures and sources of payment for all services used by Medicare beneficiaries, including co-payments, deductibles, and non-covered services.
- The Long-Term Care Minimum Data Set (MDS) contains items that measure physical, psychological, and psychosocial functioning. It contains nursing home assessment data including everyone in a Medicare-qualifying nursing home irrespective of payment source. It does not include information about payment and other services.

### **Available Resources for Linking Survey Data to Medicare Claims**

There are two main sources of information for researchers interested in linking survey data to Medicare claims. New users are most commonly referred to ResDAC, which provides free assistance to academic and nonprofit researchers interested in using MCBS data for research purposes.<sup>1</sup> The CMS Chronic Conditions Data Warehouse (CCW) is another important tool that provides researchers with Medicare and Medicaid beneficiary, claims, and assessment data.

### **Future Strategies to Facilitate CMS Linkages**

There is no compiled documentation of the procedures and challenges associated with linking CMS data to survey data. It is currently challenging for investigators to learn from the experience of other researchers who may have already succeeded in linking CMS data to their data. Participants suggested creating a crosswalk of linked CMS and survey data as a public good for the research community, to share approaches to creating linkages across datasets.

---

<sup>1</sup> See <https://www.resdac.org>.

Another challenge involves data use agreements (DUA), which tend to discourage cross-institutional work. Continued access to resources for multiple investigators who may be building off the same data is very important and sometimes a challenge with DUAs. The CMS is investigating ways in which it can better support the research community while maintaining privacy and data security. The VRDC can potentially serve as a holder of data not proprietarily owned that investigators agree to share. The NIA has encouraged researchers with NIA funding to make their code available to other researchers. However, researchers may not be willing to share code and datasets that they have spent years developing. One concern is that if data or code are shared with another investigator, questions may arise about its creation or use. Fielding these questions can be a time-consuming burden. However, participants pointed out the possibility of tiered levels of data access. For investigators trying to create a new measure, this process might be part of their intellectual property. One way to improve consistency across studies is for ResDAC or CMS to provide code, not only specifications, for standard measures (e.g., for defining emergency room visits) to benefit the public good.

### **MEDICAID DATA AS A RESEARCH RESOURCE**

Thomas MaCurdy, Stanford University

In 2014, Medicaid provided health care coverage for nearly 70 million low-income Americans. Medicaid is the main payer of nursing home care and long-term care services, and also the largest source of public funding for mental health care. It is administered by the states and is financed by federal and state funds. Health care through Medicaid is delivered through private FFS and managed care programs.

Seniors and individuals with disabilities comprise one-fourth of Medicaid enrollees, but account for two-thirds of Medicaid expenditures. Medicaid is administered by the states and funded jointly by federal and state governments. States have considerable discretion to tailor Medicaid programs to the unique needs of their low-income populations. Because of the considerable variation in state Medicaid programs, Medicaid data from different states may describe different populations and services. There is large variation in Medicaid coverage by state with regard to program eligibility and covered services. Medicaid-supported health care services are delivered through FFS and managed care programs, although trends indicate that Medicaid programs are relying more on managed care programs to deliver services.

### **Medicaid Data Sources**

As with Medicare data, new users of Medicaid data are referred to the ResDAC and the CCW for support in using these data for research. Researchers have used the following Medicaid data files:

- Medicaid Statistical Information System (MSIS) data contain enrollee eligibility information and Medicaid claims paid in each quarter of the federal fiscal year. Each state MSIS file contains one record for each person covered by Medicaid.
- The Medicaid Analytic eXtract (MAX) Files are derived from the MSIS and linked to the Medicare Enrollment Database to help identify people dually enrolled in Medicare and Medicaid. MAX data are available from 1999 to present.

- Mini-MAX is a 5 percent cross-sectional national sample of the MAX files, stratified by state and by basis of eligibility (aged, disabled, children, and adults) among full beneficiaries and separately for restricted beneficiaries.
- Alpha MAX files are in development and are a more recent version of the MAX files.
- The Medicare-Medicaid Linked Enrollee Analytic Data Source (MMLEADS) allows for the examination of all Medicare and Medicaid enrollment and claims data for people who are dually enrolled. The file includes records for Medicare-only, Medicare/Medicaid dually enrolled, and Medicaid-only with blind/disabled eligibility codes.
- A web-based application system, the Medicaid Budget and Expenditure System/State Children's Health Insurance Budget and Expenditure System (MBES/CBES) has been implemented nationwide. The system allows states to report budgeted and actual expenditures for Medicaid and the Children's Health Insurance Program (CHIP).

### **The Challenges of Using Medicaid Data for Research**

Some of the challenges involved in using Medicaid data for research are as follows:

- Not all payments are based on claims and are not always processed through a state's MMIS. An example is retrospective settlement amounts for providers who are paid on the basis of costs, rather than a fee schedule. Supplemental payments to providers made under various statutory authorities are also not processed through the state's MMIS. Finally, certain public providers receive funding through state or local budget processes, sometimes in lieu of direct payments by the state Medicaid agency.
- Eligibility data come from different systems. Although federal law requires states to operate their programs under the authority of a single state agency, multiple state and local government entities may have responsibility for different program functions.
- Managed care payment data are underreported by many states, and the quality and completeness of these data at the federal level are poor.
- Linking the records of individual beneficiaries is challenging within states and nonexistent across states.

Specific challenges related to the quality of MSIS data include the following:

- There are translation errors from the raw MMIS to MSIS data and inconsistency of the inclusion of claims data across states.
- Unique billing codes are used by some states for certain services such as long-term care, and the MSIS data are never used for program integrity purposes because of their unreliability. Other problems with the MAX data are the errors and ambiguity in adjudicating MSIS inconsistencies as well as the long delays in data availability.
- MAX data are not typically updated after they are first produced. There is limited cross-state linkage of Medicaid records. The successful use of Medicaid data requires state-specific program knowledge.

Many Medicaid services are bundled and directly paid. These services sometimes appear as gross adjustment claims and are not included in MAX. Waivers define special services and populations that are specific to individual states and time periods. Under managed care

arrangements, certain services may be excluded from the contract or “carved-out” and covered under FFS or other arrangements. Carve-out arrangements in capitated Medicaid managed care programs are fairly common for behavioral health, dental services, transportation services, and pharmacy benefits. Knowledge of state-specific eligibility codes is essential and difficult to acquire. These codes are fundamental to understanding the services available to individual Medicaid beneficiaries.

### **The Linkage of Medicaid to Other CMS Sources and Survey Data**

Medicaid data have been linked to other CMS data including Medicare claims and enrollment data, Outcome and Assessment Information Set (OASIS; a CMS partnership with the home care industry to foster and monitor improved home health care outcomes), MDS, Provider Enrollment, Chain, and Ownership System (PECOS; the electronic portal through which physicians enroll in Medicare), and National Provider Identifier (NPI) records. Medicaid data have been linked to survey data for research purposes, including the Health and Retirement Study, Panel Study of Income Dynamics, National Health Interview Survey, National Health and Nutrition Survey, and National Nursing Home Study.

### **USING PRIVATE HEALTH INSURANCE CLAIMS DATA FOR RESEARCH**

Martin Gaynor, Carnegie Mellon University

Although Medicare provides high-quality information, and is the basis for most of our knowledge about health care costs in the United States, it reflects the experience of only elderly and disabled enrollees. Medicare beneficiaries comprise only 16 percent of the population and 20 percent of health care spending.

Private health insurance claims data are an important source of information for presenting a more complete picture of health care costs and factors driving health care costs. Half of the U.S. population (and two-thirds of individuals with health insurance) has private health insurance. Not as much is known in detail about the privately insured, but private health insurance companies have recently offered access to their claims databases through third party entities as potential resources for research and analysis. Examples of companies that hold such data include the Health Care Cost Institute (HCCI), Blue Health Intelligence (BHI), FAIR Health, and Truven Health Analytics.

### **Potential Data Sources of Private Health Insurance Claims**

#### ***Health Care Cost Institute (HCCI)***

Established in 2011 as a not-for-profit organization, the HCCI has information on 88.7 million individuals from Aetna, Humana, and UnitedHealthcare. Data include individual, outpatient, employee-sponsored, fully funded, self-funded, and Medicare Advantage. In addition, there is information on all covered expenses including inpatient, outpatient, pharmaceutical, facility, and physician services. The almost 3 billion claims cover about 27 percent of Americans with

employer-sponsored insurance. Gaining access to these data is costly and currently limited to certain academic partnerships.<sup>2</sup>

***Blue Health Intelligence (BHI)***

BHI is a for-profit firm with data on 165 million individuals from every county in the United States contributed by 23 of the Blue Cross Blue Shield plans. Data include enrollment, benefits, and claims information (medical, pharmaceutical, lab, and dental). There is some ambiguity about the membership information. A plan ID exists, but it is not linkable to the benefit structure. Both the plan and insurance companies are unknown. Access to data is costly.<sup>3</sup>

***FAIR Health***

FAIR Health is a not-for-profit corporation. They have data from more than 60 nonidentifiable insurers covering 150 million individuals. One challenge is variation in available fields depending on which insurer has supplied the data. Negotiated transaction prices, for example, are not populated for all plans. Physician pricing data are good but hospital pricing is not. The BHI and HCCI databases are more complete and consistent in this regard. Patient identifiers are available but cannot be linked in a meaningful way.<sup>4</sup>

***MarketScan Research by Truven Health Analytics***

The MarketScan Research database has data on 170 million patients since 1995 and is unique in that complete payment information is captured, including payments from the benefit plan and the patient. Data on specialty pharmacy and mail-order prescriptions are included. The core databases include Commercial, Medicare Supplemental, and Medicaid Specialty. The longitudinal integrity is strong, but it is unclear whether data can be analyzed at the hospital or geographic region levels. For example, HRRs are not reported in the data. Medicare Advantage data are available. Truven Health Analytics data are easy to work with, and they existed for the longest time period; however, these data are not appropriate for understanding variation at the system level.

***All-Payer Claims Databases (APCDs)***

APCDs are created and managed by the states and include health insurance claims from all payers including private insurers, Medicare, and Medicaid. Private, self-funded plans are not required to contribute information to these databases. Some private plans regard their information as proprietary and are not likely to voluntarily participate in an APCD. Others might volunteer to participate in APCDs under certain conditions. In Massachusetts, for example, Blue Cross Blue Shield has announced that it will not submit its data to the APCD unless required by the employers.<sup>5</sup>

---

<sup>2</sup> See <http://www.healthcostinstitute.org> for more information.

<sup>3</sup> See <http://www.bluehealthintelligence.com> for more information.

<sup>4</sup> See <http://www.fairhealth.org/Research> for more information.

<sup>5</sup> See <https://www.apcdouncil.org> for more information.

### **Other Data Sources**

Data from other sources such as web data and independent searches are a largely untapped area. UnitedHealthcare, for example, has tools for its enrollees to use. This information could be captured and analyzed. There is also a large amount of detailed data on individuals held by third party data brokers. Data are obtained via web searches and county courthouses. These data are available for commercial use and may prove valuable.

### **Linking Private Insurance Claims Databases with External Datasets**

All of the private insurance claims databases can be linked with external datasets (e.g., Medicare, Medicaid, the American Hospital Association Annual Survey, county characteristics). Linking can be done at both the geographic and provider levels. Linking at the patient level is more problematic because of Health Insurance Portability and Accountability Act (HIPAA) restrictions. Linking the APCDs is state specific and less clear. There are several examples of research questions that can be addressed with private data. Investigators have analyzed the differences between charges, negotiated prices, and Medicare prices. Medicare spending per beneficiary and overall private spending per beneficiary can be compared across HRRs.

Investigators often informally share their experiences using these databases with other researchers. Much of this information is undocumented. Some investigators have developed valuable working relationships with representatives of the companies holding the data, which assists the investigators in navigating the process of obtaining these data. Investigators' unique research questions result in unique data needs.

The existing databases on the privately insured cover different parts of this population (e.g., commercial versus Blue Cross). The organizations holding these data have not as yet made them jointly available. Using the private data sources along with public data sources would be tremendously valuable as a way to allow a complete picture of health care utilization and spending in the U.S. Using the data in this way does not require merging of datasets. The databases can continue to be separate if so desired. All that would be required is agreement to allow use of the databases together to generate knowledge in the public's interest. This is feasible and should be a national priority—it is an investment in our country's knowledge infrastructure.

### **THE CMS PERSPECTIVE ON POTENTIAL OPPORTUNITIES FOR RESEARCH**

Allison Oelschlaeger, Office of Enterprise and Analytics

The CMS must balance multiple competing interests and sensitivities regarding data release practices to protect beneficiary privacy. Too much flexibility could result in security breaches that may compromise individual privacy. Too little flexibility could inhibit health system insights that may improve the care that beneficiaries receive. Historically, the CMS has mainly collected data for operational purposes. Over the past several years, there has been an increase in demand for CMS data for various quality improvement and research initiatives. The CMS has several ongoing initiatives intended to facilitate the sharing of data to a range of entities. Maintaining protections for beneficiary privacy as required by the Privacy Act of 1974 and HIPAA must remain a priority. There are three areas of ongoing development.

### **Release Aggregated Data**

During the past 6 years, the CMS developed a number of public datasets that now include data files segmented by geography (HRR and county), as well as provider, across the majority of Medicare services. The provider datasets cover more than 85 percent of Medicare payments (not including those in Medicare Advantage) and have grown over time to include hospital inpatient and outpatient files, physician and other supplier data, and skilled nursing facility files. An example of research conducted using these data is analysis of the distribution of Medicare office visits by level of service for top specialties.

### **Share Beneficiary-level Data with External Data Users**

For researchers interested in analyzing claims level Medicare and Medicaid data as well as linking survey data to Medicare or Medicaid data, the CMS is working on facilitating the access and use of its data using multiple mechanisms, while maintaining beneficiary privacy and ensuring appropriate use of the data. Examples of such mechanisms include the following:

- ResDAC provides assistance for researchers interested in using Medicare and/or Medicaid data.
- The CCW is designed to support external researchers and internal CMS research and analytic functions.
- Traditionally, the CMS has shipped encrypted files to researchers who must ensure the security of the data. The CMS has recently developed the VRDC, which allows researchers a secure and efficient way to access CMS data virtually from their independent workstations. Researchers can now choose between physical and virtual data access.
- The CMS developed the Data Navigator tool as a resource to find different aggregated and beneficiary-level data files because it is not always easy to find specific data on the CMS website.
- The Transformed Medicaid Statistical Information System (T-MSIS) is an expansion of the existing CMS MSIS extract process. The T-MSIS extract format is intended to improve the reliability and timeliness of the data through monthly updates. A few states have begun the transition to T-MSIS and CMS is starting to evaluate the quality of the data states are submitting.
- The CMS has improved the timeliness of Medicare fee-for-service data availability, allowing researchers to request data of Medicare FFS claims on a quarterly rather than annual basis. These files are 90 to 93 percent mature depending on the type of service. This is a good option for clinical trials that require active monitoring.
- The beneficiary ID is a unique beneficiary identifier (unlike the HIC number) encrypted specifically to the researcher's DUA. This ID allows data linkages across files.

The CMS makes two types of files available to researchers:

- Limited Data Set (LDS) files, which exclude specific direct identifiers, including name, address, health insurance claim, Social Security Number, date of birth, zip code, and medical record number. These data are much easier to request than restricted data; the approval process is less rigorous and less expensive, which is optimal for new investigators with limited funding.
- Research Identifiable Files are custom CMS data extracts that may contain restricted beneficiary identifiers.

### **Enhance Ability for Beneficiaries to Use and Share their Data**

Medicare Blue Button is a way for individual beneficiaries to obtain their personal health records electronically at [www.mymedicare.gov](http://www.mymedicare.gov). Health information is available for download in a text file or PDF format. Once downloaded, beneficiaries can send their health information to others, such as a researcher, family member, medical provider, or caregiver. The Medicare Blue Button interface allows beneficiaries to access data directly at no charge without interacting with the CMS. The CMS anticipates that beneficiary data sharing through Medicare Blue Button to be fully functional as early as next year.

### **IMAGING DEMENTIA—EVIDENCE FOR AMYLOID SCANNING**

Constantine Gatsonis, Brown University

One example of innovative research that links registry data to CMS claims data is in the beginning stages of data collection and patient recruitment. The Imaging Dementia-Evidence for Amyloid Scanning (IDEAS) study is a collaboration of several public and private entities. Stakeholders from both industry and professional organizations have come together to assess the impact of amyloid positron emission tomography (PET) scanning on the management of patients meeting Appropriateness Use Criteria. The study will compare hospital admissions and emergency room visits within 12 months in patients enrolled in the registry to matched controls (no amyloid PET) identified from CMS claims data. The aims are to determine whether having an amyloid PET changes the management of the patient in terms of Alzheimer's disease drug therapy, other drug therapy, or counseling about safety and future planning. Other endpoints that will be investigated and compared between the registry and the CMS control groups include inpatient hospital admissions and emergency room visits over the next 12 months. Amyloid plaque scans, cognitive function tests, data on caregiver burden, and functioning are some examples of the data collected on IDEAS participants. The IDEAS study has attracted several other add-on studies using the cognitive function tests and caregiver burden and functioning tests, for example.

### **USING MEDICARE CLAIMS WITH ADD-INS**

Alan M. Zaslavsky, Harvard Medical School

Medicare claims data provide a wealth of information allowing for investigation of geographical variation in costs and utilization, variation across providers, costs associated with specific treatments, conditions and specialties, and risk estimation and adjustment (costs and

mortality). There are also challenges in using Medicare claims data for research, several of which were presented earlier. These include:

- limited availability of claims data from Medicare Advantage;
- limited amount of clinical information reflected in billing data;
- lack of information on exposures, clinical experience or health before gaining Medicare eligibility; and
- limited availability of social characteristics, either contemporaneously or over the life course.

Because of these limitations, it is often critical to supplement claims with other data sources, including:

- enrollment files, which track coverage as beneficiaries move in and out of different Medicare programs;
- Part D (drug plan) data;
- survey data;
- other quality measures, such as the HEDIS quality and utilization measures collected for samples of enrollees in Medicare Advantage contracts; and
- geographically-based linkages for social characteristics and health resources.

Several large surveys sponsored by CMS are conducted regularly with Medicare beneficiaries. These include the Consumer Assessments of Healthcare Providers and Systems (CAHPS) surveys of patient experiences with care, administered to Medicare Advantage and Fee-for-Service beneficiaries, the Health of Seniors surveys of health and functional status administered in Medicare Advantage, and the Medicare Current Beneficiary Survey, a more intensive study of health and health care which follows sampled individuals over a year. Although these are sample surveys, they are quite large, with responses cumulatively in the millions. Other important data linkages involving Medicare claims include those associated with the Health and Retirement Study, a longitudinal cohort study of health, functional, economic, and social status of retirement-age individuals, and the SEER-Medicare database, which links claims to data on cancer epidemiology and treatment from state cancer registries.

Some of the structural and access problems affecting enhancement of claims by data linkage include the following:

- The technical challenges of each linkage require a heavy initial investment, which is best justified if it can serve a program of research, but linkages are performed at the project level and are not (and under current rules cannot be) shared broadly with the research community; even maintaining continuity of access to a linkage within a research center can be challenging.
- Medicare Advantage and FFS data content are inconsistent.
- The administrative “owners” of the data in the various program areas of the CMS appear to have varying degrees of commitment to facilitating research by making data accessible and useful to the research community. While CMS staff concerned with

monitoring and improving quality and efficiency are potentially users of and advocates for research, they are located in separate program areas.

## Panel Discussion

Michael Chernew summarized the access and content issues that challenge both new and experienced investigators and led the panel discussion.

The linkage process in and of itself is a challenge. There is no mechanism or resource in place whereby researchers can learn from each other on the process for linking datasets. Experienced researchers in the field have accumulated much of their knowledge through personal interactions and the development of relationships with representatives at the CMS and the data holders. Some participants recommended creating a publicly available resource for documenting and sharing experiences. Useful tips for the preparation and cleaning of data before the linkage should be documented and made accessible to new investigators.

A publicly available resource for sharing experiences could also be used to share code. For example, ResDAC defines emergency room visits, but does not include the actual code. Sharing code would be invaluable to ensure that variables are measured in the same way across studies. A crosswalk of the codes from hospital IDs, HCCI data, and Medicare data is another example of a useful resource that investigators might be willing to share. There would need to be incentives in place to encourage sharing and to address concerns about sharing code that took 2 years to develop or code that is the basis of as yet unpublished work.

New investigators often do not understand the gaps in the CMS data. Medicare claims data, for example, does not contain long-term care data. Investigators erroneously believe that home care is the same as long-term care support. Investigators may struggle to write an effective application for funding if they do not fully understand the potential research possibilities of the data.

Another challenge is the accuracy of the diagnoses in the claims data. Dementia is particularly problematic. In order to be captured in claims data, a beneficiary must present to the health care system and the diagnosis must be recorded. Clinicians are often not effective at making dementia diagnoses, and the diagnostic codes used are inconsistent.

Chernew noted that CMS data-use agreements (including those for access to the VRDC) are too narrowly focused on specific projects. If a question outside of the original project protocol is raised during use of a particular database, it cannot be addressed. The investigator already has the data, yet has to go through another VRDC process to be able to ask other questions. The current way in which data access is granted is very project specific as opposed to group, person, or database specific. This results in a lot of lag in projects. The added security risk in this scenario is difficult to understand.

Researchers view the CMS data linkage process as being detached and difficult to understand. Problems can arise when the Principal Investigator's institution holds a data-use agreement but the co-investigator's institution does not. The process does not encourage collaboration, and a broader approach would be beneficial to the field. Multiple researchers should be looking for answers to similar questions from different angles. New research should focus on specific health events or health conditions instead of mining all the claims data.

## **Appendix 1 Workshop Agenda**

### **Expert Meeting on the Potential Value of Centers for Medicare and Medicaid Services Data as a Research Resource for National Institute on Aging Studies**

The National Academies of Sciences, Engineering, and Medicine Building  
2101 Constitution Ave. NW, Washington, DC, Room 118  
May 4, 2016  
1 P.M. – 5 P.M.

---

The goal of this expert meeting is to discuss the potential value of data available from the Centers for Medicare and Medicaid Services (CMS) as a research resource for National Institute on Aging (NIA) studies. The meeting will examine the role CMS data can play in enhancing NIA behavioral and social research, the existing research opportunities, and the limitations associated with the data.

---

- |                  |                                                                                                                                                                                           |
|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1:00-1:10</b> | <b>Introductions and Welcome on Behalf of the National Academies</b><br>Connie Citro, Committee on National Statistics                                                                    |
| <b>1:10-1:30</b> | <b>Welcome on Behalf of the National Institute on Aging and Rationale for the Meeting</b><br>John Haaga, National Institute on Aging<br>Partha Bhattacharyya, National Institute on Aging |
| <b>1:30-2:00</b> | <b>Overview of Medicare Claims Data</b><br>Julie Bynum, Dartmouth College, Geisel School of Medicine                                                                                      |
| <b>2:00-2:30</b> | <b>Medicaid Data as a Research Resource</b><br>Thomas MaCurdy, Stanford University                                                                                                        |
| <b>2:30-3:00</b> | <b>The Feasibility of Linkages with All Payer Claims Databases</b><br>Martin Gaynor, Carnegie Mellon University                                                                           |
| <b>3:00-3:15</b> | <b>Break</b>                                                                                                                                                                              |

- 3:15-3:45**      **CMS Perspective on Potential Opportunities for Research**  
Allison Oelschlaeger, Centers for Medicare and Medicaid Services
- 3:45-4:30**      **Panel Discussion of Potential Uses of CMS data for NIA Research**  
Michael Chernew, Harvard Medical School  
Constantine Gatsonis, Brown University  
Alan Zaslavsky, Harvard Medical School
- 4:30-5:00**      **Floor Discussion and Wrap-Up**
- 5:00**            **Adjourn**

**Appendix 2**  
**List of Meeting Participants**  
**May 4, 2016**

**Invited Participants**

**Julie P.W. Bynum, MD, MPH**, Dartmouth Institute for Health Policy & Clinical Practice  
**Michael Chernew, PhD**, Harvard Medical School  
**Constantine Gatsonis, PhD**, Brown University  
**Martin Gaynor, PhD**, Carnegie Mellon University  
**Thomas MaCurdy, PhD**, Stanford University and Acumen, LLC  
**Allison Oelschlaeger, BSFS**, Office of Enterprise Data and Analytics, Centers for Medicare and Medicaid Services  
**Alan Zaslavsky, PhD**, Harvard Medical School

**National Institute on Aging, Division of Behavioral and Social Research**

**Partha Bhattacharyya, PhD**, Program Director  
**Prisca Ndella Fall**, Research Program Analyst, BSR  
**John Haaga, PhD**, Director  
**Evelyn Neil**, Program Analyst  
**Georgeanne Patmios**, Acting Deputy Director  
**Anneliese Hahn Ebersole, PhD** (Contractor, Rose Li and Associates, Inc.)

**National Academies Committee on National Statistics**

**Constance F. Citro, PhD**, Director  
**Krisztina Marton, PhD**, Senior Program Officer  
**Eileen LeFurgy**, Program Coordinator