



Studying Recipients of Long-Term Care Services and Supports:

A Case Study in Assembling Medicaid and Medicare Claims and
Assessment Data in California

Prepared for The SCAN Foundation

By the California Medicaid Research Institute (CaMRI),
University of California

Julie Stone, MA
Robert J. Newcomer, PhD
Arpita Chattopadhyay, PhD
Todd P. Gilmer, PhD
Phillip Chu, MA
Chi Kao, Ph.D.
Andrew B. Bindman, MD

N o v e m b e r 1 6 , 2 0 1 1

Acknowledgments

This report was supported by a grant from The SCAN Foundation. The SCAN Foundation is dedicated to creating a society in which seniors receive medical treatment and human services that are integrated in the setting most appropriate to their needs. For more information, please visit www.TheSCANFoundation.org.

Authors of this report wish to acknowledge the contributions of the dedicated staff at the California Department of Health Care Services (DHCS) who assisted CaMRI investigators in compiling information used in the development of this report and in organizing the study that is described herein.

CaMRI authors also thank Patrick Henderson for copy editing this report.

CaMRI

CaMRI is a multi-campus research program of the University of California that promotes the development and dissemination of evidence to improve policy decision-making in California's Medicaid program.

Executive Summary

A new body of research will be needed to help federal and state policy makers identify methods for improving the quality and efficiency of care for individuals with long-term services and supports (LTSS) needs enrolled in Medicaid only and in both Medicaid and Medicare for dual eligibles. Such research can benefit from Medicaid's and Medicare's administrative data to evaluate services provided to individuals by need and along the entire care continuum. To be useful to policy-makers, research must be comprehensive, scientifically valid, and timely.

To date, however, these studies are extremely limited in number and reach. This is in part because researchers face significant challenges in acquiring and assembling the large sets of Medicaid and Medicare program data needed for an analysis of this scale.

At the request of the California Department of Health Care Services (DHCS) and with co-funding from The SCAN Foundation, the California Medicaid Research Institute (CaMRI) is conducting a statewide comprehensive study of Medicaid LTSS in California between calendar years (CYs) 2005-2008. To do this, CaMRI is developing an integrated and longitudinal database containing claims and assessment data from both Medi-Cal (California's Medicaid program) and Medicare. This database will be used to describe LTSS users in California and their characteristics, including demographics, medical conditions, disabilities, costs and patterns of service use across both Medi-Cal and Medicare. This study will assist DHCS and the Centers for Medicare and Medicaid Services (CMS) in better understanding how to deliver care for this population more effectively and efficiently.

This report describes the process CaMRI underwent with its state and federal partners to acquire these data for analysis. Specifically, it contains a timeline of the various activities involved in requesting Medi-Cal data from the state and Medicare data from CMS. This report then describes the contracts and data use agreements that allowed for the sharing of Medi-Cal and Medicare data, as well as some of the methods CaMRI underwent to comply with these agreements. Examples of some of the complexities CaMRI faced in linking and cleaning these different data files are also provided. Further, this report includes a discussion of some of the major issues, challenges, and obstacles CaMRI has faced to date, including the complexity of acquiring assessment data from multiple departments with the Agency, the discrepancies in the delivery of the data that were delivered, and the significant role the state Medicaid department played in facilitating this study, among other things.

The authors of this report believe the following recommendations, if implemented, can better facilitate additional research of this kind.

1. CMS should establish more efficient, timely and routine approaches to share accurate, complete and current federal claims and assessment data with state Medicaid programs.
2. CMS can play a leadership role in standardizing the definitions of certain demographic and other variables across and within federal and state data sets; establish additional procedures for auditing and editing data to enhance their accuracy; and create and disseminate tools for the assembly and analysis of Medicaid and Medicare data.

3. To attract research partners, state Medicaid programs need to minimize financial and administrative barriers to delivering data to researchers in a timely fashion,
4. States should establish interagency agreements to support the sharing of data across departments to operate and evaluate LTSS and other services.
5. To monitor the performance of Medicaid managed care programs, Medicaid agencies should require plans to routinely submit accurate and complete information on all encounters.
6. Agencies that make assessments of Medicaid beneficiaries to determine eligibility for LTSS should establish a standardized, electronic minimum data set so comparable assessments can be made of beneficiaries' functional and cognitive status over time, regardless of care setting.

Table of Contents

Acknowledgments	i
CaMRI	i
Executive Summary.....	ii
Introduction.....	1
Background: Study Approach.....	3
Defining the Study Cohort of Recipients of LTSS	5
Summary of Data Needs	6
Medi-Cal and Other State Data	6
Medicare Data	9
Data Linking and Cleaning.....	13
Study Timeline	14
Contracts and Data Use Agreements for Data Sharing.....	17
CaMRI’s Security Measures.....	19
Issues, Challenges and Obstacles	21
Medi-Cal Data Housed Across Multiple State Departments	21
Limited Access to Managed Care Data	23
Discrepancies in the Delivery of Medicare Data.....	24
Partnering with State Government	25
Limitations to Analysis of Administrative Data.....	25
General Observations and Recommendations	28

Tables

Table 1. Users of Medi-Cal LTSS, 2005-2008	4
Table 2. State Medi-Cal Files Collected, CYs 2005-2008	8
Table 3. Medicare Data Files Collected, CYs 2005-2008	11
Table 4. Study Timeline, September 2009 to Present.....	15
Table 5. Summary of Contracts and Other Agreements	18
Table 6. California’s HHS Departments and Counties Responsible for Administering, Purchasing and Overseeing Medi-Calal Files Collected, CYs 2005-2008.....	21
Table 7. Source of Data Files by Department within the California’s HHS, CYs 2005-2008.....	23

Appendix 1. Glossary of Terms	31
--	-----------

Introduction

Users of long-term care services and supports (LTSS)¹ constituted 22.6% of Medicaid enrollees (13.3 million of a total 58.8 million enrollees) nationally in fiscal year (FY) 2008, and used 62.3% of total dollars spent on Medicaid-covered benefits (\$210.8 billion of a total \$338.6 billion on all enrollees).² As high service users, they are the most costly group of Medicaid beneficiaries. Although many users of LTSS are cared for in Medicaid-covered institutional settings, such as nursing homes, many are also served in their homes, group residential settings, and assisted living facilities. Some enter Medicaid as children, after being born with disabling conditions; others enroll as working-age adults with inherited or acquired disabling conditions, and still others join much later as aging retirees who have lost the ability to care for themselves.

Many users of Medicaid LTSS are also enrolled in Medicare. Such persons are referred to as dual eligibles.³ Dual eligibles are more likely than other Medicare beneficiaries to be in fair or poor health, cognitively and/or functionally impaired, and have more chronic conditions.⁴ In FY 2008, 9.2 million people nationally were enrolled in both Medicaid and Medicare.⁵ In FY 2005, the most recent year for which this information is available, spending on duals accounted for about 46% of total Medicaid spending on benefits and 25% of total Medicare program spending.^{6,7} Medicaid and Medicare combined spent \$196.3 billion on duals in 2005.⁸

As a means of containing Medicaid spending on high cost recipients of LTSS, many states are attempting to cover services for beneficiaries in the most cost-effective settings (i.e., community-based settings rather than in institutions). Similarly, states and the federal government have begun to recognize the value of better coordinating Medicaid and Medicare benefits for dual

¹ LTSS refers to a broad range of health and social services needed by people with a limited capacity for self-care due to a physical, cognitive, or mental disability or condition that results in functional impairment and dependence on others for an extended period of time.

² Figure 5, "Medicaid and CHIP Program Statistics: June 2011 MACStats," Medicaid and CHIP Payment and Access Commission (MACPAC), Washington, DC, June 2011.

³ Such individuals may qualify for full Medicare benefits and all of the Medicaid benefits offered in their state of residence, including Medicaid coverage of Medicare premiums and cost-sharing obligations. These dual eligibles are often referred to as full duals. Others may qualify for full Medicare benefits and only Medicaid coverage of their Medicare cost-sharing obligations, including Medicare premiums and/or Medicare deductibles and coinsurance. These dual eligibles are referred to as partial duals. For the purpose of the CaMRI study, full and partial duals are included in the analysis.

⁴ Duals often require a continuum of acute and LTSS that meet their changing health and social service needs, including LTSS delivered in home and in community settings to enable them to maximize their independence at home. Together, the continuum of acute and LTSS benefits available to duals across both Medicare and Medicaid are intended to meet these complex service needs.

⁵ Excerpt from, "Report to Congress on Medicaid and CHIP: March 2011," Medicaid and CHIP Payment and Access Commission (MACPAC), Washington, DC, March 2011. This number includes both full duals and partial duals.

⁶ Ibid.

⁷ Almost all Medicaid beneficiaries who are age 65 years and older and more than one-third of non-elderly Medicaid beneficiaries with disabilities are dual eligibles. Source: Korb, MA, Jody, McCall, MS, Nelda, "Integrated Care Program: Final Evaluation," prepared for Center for Health Care Strategies, Inc. October 2008.

⁸ Coughlin, Teresa, Waidmann, Timothy, and O'Malley Watts, Molly, "Where Does the Burden Lie? Medicaid and Medicare Spending for Dual Eligible Beneficiaries" Kaiser Commission on the Uninsured, Kaiser Family Foundation, April 2009. <http://www.kff.org/medicaid/upload/7895-2.pdf>

eligibles as a means of containing costs. This includes integrating the management of primary, acute, post-acute and LTSS across the two programs, and improving transitions between institutional and home-based LTSS. The establishment of the Federal Coordinated Health Care Office (often referred to as the “Medicare-Medicaid Coordination Office”) for dual eligibles within the Centers for Medicare and Medicaid Services (CMS), by the Patient Protection and Affordable Care Act (PPACA, P.L. 111-148⁹), demonstrates a new federal focus on these priorities.

Care improvements and cost containment cannot be accomplished effectively without a substantial body of expert research and analysis describing where we are today and identifying the most promising approaches for improvement. Such research and analysis must be comprehensive and timely. To date, however, these studies are extremely limited in number and reach, in part because researchers face significant challenges in acquiring and assembling the large sets of Medicaid and Medicare program data needed for an analysis.

At the request of the California Department of Health Care Services (DHCS),¹⁰ and with co-funding from The SCAN Foundation, the California Medicaid Research Institute (CaMRI) is conducting a statewide comprehensive study of Medicaid LTSS in California between calendar years (CYs) 2005-2008. To do this, CaMRI is developing an integrated and longitudinal database containing claims and assessment data from both Medi-Cal (California’s Medicaid program) and Medicare. This database will be used to describe users of LTSS in California and their characteristics, including demographics, medical conditions, disabilities (e.g., limitations in activities of daily living), costs and patterns of service use across both Medi-Cal and Medicare.

In general, this study aims to inform California, other states and federal entities about the extent to which home and community based services and other LTSS are cost-effective and result in improved health outcomes. Specifically, this study will assist DHCS and CMS, the federal agency responsible for administering all aspects of the Medicare program and for Medicaid administration at the national level, in better understanding how to deliver care for this population more effectively and efficiently.

To conduct this study, CaMRI obtained claims, assessment and enrollment data for both the Medi-Cal and Medicare programs. Obtaining Medi-Cal data from the state presented a unique set of logistical challenges for CaMRI. Obtaining Medicare claims and assessment data from CMS at the federal level raised a separate set of challenges.

This report describes the study, the data files used, and the activities involved in acquiring Medicaid and Medicare data from the state and federal governments. The report also describes the contracts and data use agreements that allowed for the sharing of Medi-Cal and Medicare data as well as some of the methods CaMRI underwent to comply with these agreements.

⁹ PPACA was signed into law by the President on March 23, 2010.

¹⁰ CMS is responsible for Medicaid program administration at the federal level but individual states administer their own programs on a day-to-day basis. DHCS is a department within the California Health and Human Services Agency (Agency). DHCS administers Medi-Cal on a day-to-day basis.

Examples of some of the complexities CaMRI faced in linking and cleaning these different data files are also provided. The report includes a discussion of some of the major issues, challenges and obstacles CaMRI has faced to date (i.e., within the first two years of this three-year study). Finally, the report describes suggested steps that researchers, states and the federal government might take to facilitate an easier development of linked databases for future research opportunities.

This report is intended as a case-study to assist researchers, states, and CMS plan for other research projects of this kind. The release of this report is particularly timely in that the Medicare-Medicaid Coordination Office and the Center for Medicaid, CHIP and Survey & Certification (CMCS) within CMS recently issued an informational bulletin to inform state Medicaid agencies of the availability of technical assistance, and a process for requesting timely Medicare data to support care coordination, improve quality, and control costs for dual eligible beneficiaries.¹¹

Background: Study Approach

Fundamental to the study is the creation of an integrated and longitudinal Medi-Cal and Medicare database, including claims, assessments, enrollment and certain supplemental files from both programs. For Medi-Cal, this involved the compilation and linkage of individual records from several administrative data files in California for the four-year study period (CYs 2005-2008). These files describe the experiences of 1.579 million users. Some of the study's recipients are enrolled in Medi-Cal only while others are dually enrolled in Medicare. Further, some are enrolled in fee-for-service, while others are enrolled in Medi-Cal and/or Medicare managed care plans. Finally, some are enrolled in both Medi-Cal managed care and Medicare managed care simultaneously during some point of the study period. **Table 1** shows the count of these beneficiaries by category.

¹¹ See, <https://www.cms.gov/CMCSBulletins/downloads/Coordinated-Care-Info-Bulletin.pdf>

Table 1. Users of Medi-Cal LTSS, 2005-2008

Recipient Category	Counts	Percent of Total
All Study Participants	1.579 million	100%
Medi-Cal Only	711,423	45%
Dual Eligibles	867,594	55%
All Study Participants by Service Delivery System		
Medi-Cal Fee-For-Service	1,079,674	68.4%
Medi-Cal Managed Care ^a	499,163	31.6%
Dual Eligibles by Service Delivery System		
Medicare Fee-For-Service	713,613	82%
Medicare Managed Care ^b	153,981	18%
Ever used Medicare Managed Care and Medi-Cal Managed Care ^c	36,011	4.2%

Source: Study data.

^a Participation in Medi-Cal managed care refers to fully and partially capitated managed care, including participation in the Program for All-Inclusive Care for the Elderly (PACE). Count describes individuals who have ever participated.

^b Participation in Medicare managed care refers to participation in Medicare Advantage (including Medicare Special Needs Plans) and PACE. Count describes individual who have ever participated.

^c Dual eligibles can be enrolled in more than one service delivery system. For example, it is possible to be simultaneously enrolled in Medicare Advantage and Medi-Cal managed care.

CaMRI obtained Medicare’s claims, assessment and enrollment files for the study period for dual eligibles. These Medicare data are collected by providers who contract with the Medicare program. For reimbursement and accountability purposes, these data were sent by providers to CMS. A description of the data files acquired for this study are provided under the section entitled Summary of Data Needs, found later in this report.

A description of the process CaMRI underwent to acquire Medi-Cal and Medicare data for analysis is found in the section entitled, Study Timeline. In this section, a timeline is provided with detailed descriptions of the various steps that were taken by CaMRI, the state departments, CMS and its two contractors, ResDAC and Buccaneer Computer Systems & Service, Inc.

Linking and cleaning the data required significant program knowledge and technical precision among researchers. Assistance from DHCS’ data programmers and CMS, including its two contractors, was provided to CaMRI. Throughout the file linking process, a substantial number of discrepancies across data sets were identified by researchers. CaMRI is working to interpret and clean these data discrepancies and to develop protocols for doing so. The linking process is described in greater detail later in the section of this report entitled Data Cleaning and Linking.

Personal health information (PHI) was needed to create a link between federal and state data. This introduced a layer of complexity in the data requests. A number of data sharing and other contractual agreements were written to ensure CaMRI’s compliance with federal and state rules designed to protect beneficiary privacy and to ensure the security with which data are transferred between parties. These data sharing and other contractual agreements are discussed in the section of this report entitled Contracts and Data Use Agreements for Data Sharing.

To comply with these requirements, CaMRI established certain administrative, technical and physical safeguards. These security measures are described in the section of this report entitled CaMRI's Security Measures.

The first phase of statistical analysis using this large database will begin with descriptions of service users, their demographic characteristics, utilization and spending. These analyses will look at annual counts of recipients by service, counts of service users by demographic characteristics, spending by age group, spending by service, among others. The second phase of the analysis will describe program and other care system effects on beneficiaries' health and cost outcomes both within a single year and over the four-year study period. The third phase will bring together the first two phases of the analysis to develop a dynamic microsimulation model to estimate transitional probabilities between care settings and health outcomes. A series of policy reports will be issued to describe the study's findings. Such reports will be made available to DHCS, disseminated to federal and state policy-makers, and made public by The SCAN Foundation and CaMRI's forthcoming website.

Defining the Study Cohort of Recipients of LTSS

To define the study cohort, CaMRI and DHCS joined together to develop criteria for selecting just those recipients in Medi-Cal who received LTSS at any time during the four-year period between CYs 2005 and 2008. Since Medicaid, in general, and Medi-Cal, specifically, do not have an eligibility code that identifies beneficiaries receiving LTSS, CaMRI investigators began by building a list of Medi-Cal services that could be classified as LTSS. In compiling this list, a broad definition was adopted from the Medi-Cal claims data to maximize the potential to identify all such beneficiaries even if that might result in the over selection of some individuals who might not meet a strict definition of LTSS.

An assumption was made that once the pattern of services for all selected individuals could be reviewed, it would be possible to further refine the study sample to conform to widely accepted definitions of LTSS. Further, because the study tracks individuals longitudinally and looks at precipitating health events prior to needing LTSS, CaMRI compiled claims and assessment records for the whole period for all those using targeted services at any time during CYs 2005 through 2008.

The services used to define LTSS are shown below. This list is inclusive of both home and community based services and institutional services commonly used to define LTSS, as well as therapies and other supportive services that may be associated with recipients on a pathway to LTSS use, or after exiting the LTSS system.

Medicaid Home and Community Based Service Waivers (authorized under section 1915(c) of the Social Security Act)

- Acquired Immune Deficiency Syndrome Waiver
- Assisted Living Waiver
- Home and Community-Based Services Waiver for the Developmentally Disabled
- In-Home Supportive Services Plus Waiver (IHSS Plus)

- In-Home Operations Waiver
- Nursing Facility/Acute Hospital Waiver
- Multipurpose Senior Services Program Waiver

State Plan Services

- Adult & Pediatric Day Health Care Services
- Audiology Services
- Certified Hospice Service
- Durable Medical Equipment
- Home Health Agency Services/ Private Duty Nursing
- Independent Rehabilitation Facility
- In-Home Supportive Services (IHSS)
- Intermediate Care Facilities for the Developmentally Disabled
- Nursing Facility
- Occupational/Physical/Speech Therapy
- Prosthetics and Orthotics
- Renal Dialysis
- State Developmental Centers and State Mental Hospitals
- Targeted Case Management
- Medi-Cal Managed Care plans that include LTSS including the Program for All-Inclusive Care for the Elderly (PACE)
- Individuals in the California Department of Developmental Services' Regional Center CDER data base

Summary of Data Needs

Medi-Cal and Medicare claims, assessment, and certain other data files are needed to build the longitudinal database for this study. As CaMRI's research is limited to Medi-Cal recipients of LTSS between CYs 2005 and 2008, only the data records describing these individuals are needed. Because neither DHCS nor CMS routinely assemble these data into a single database, data from both programs are stored in a series of distinct files. As a result, multiple files from both programs were requested. The following describes these data files.

Medi-Cal and Other State Data

The Medi-Cal claims data describe service costs, provider types, and beneficiary utilization, among other information. Specifically, they include Medi-Cal claims for physicians, hospitalizations, nursing home care, Adult Day Health Care, In Home Supportive Services, and all home and community based services waivers, among others. These data are collected by DHCS for reimbursement purposes. In CaMRI's analysis, claims data are used to document Medi-Cal service use and expenditures of both Medi-Cal and dual eligible beneficiaries.

Claims data are not available for beneficiaries enrolled in managed care programs.

Claims on Medi-Cal beneficiaries in managed care are not provided by managed care plans to DHCS. To get around this limitation, CaMRI is using California's statewide hospital discharge abstracts (Patient Discharge Database, PDD) that capture all Medi-Cal hospital discharges in California, including discharges for Medi-Cal's managed care participants.

Assessment files provide information about recipients' demographic characteristics, health status, and functional and cognitive limitations, among other information. A number of assessment files are collected by Medi-Cal, including those collected by (1) the Department of Social Services (DDS) for the In-Home Supportive Services (IHSS) program, (2) the departments¹² administering section 1915(c) waivers,¹³ and (3) the Department of Developmental Services for Medi-Cal beneficiaries with developmental disabilities.¹⁴ Many of the specific waiver and state plan program assessment files are managed at the county or local level and are either not available in an electronic format or are not compiled in a consistent manner across the state. As a result, assessments collected for recipients of Medi-Cal's LTSS were not available for all study participants.

After significant inquiry, analysis and discussion with DHCS and its sister departments, CaMRI decided it was feasible to make use of state-based assessment files covering two large subgroups of Medi-Cal beneficiaries receiving LTSS: the Case Management Information and Payrolling System (CMIPS) used for recipients of In-Home Supportive Services¹⁵ and the state's Client Development Evaluation Report (CDER) used for all service recipients with developmental disabilities.¹⁶ CaMRI is using these assessments to report on the health status of recipients and to examine their health trajectories within care settings and along a service continuum. They are also being used to report on the functional and cognitive status of recipients and their living arrangements. Such information complements the health condition information available from claims data.

Medi-Cal's eligibility file (sometimes referred to as enrollment files) and the state's death records are also used for this study. The eligibility file contains month-to-month enrollment information for Medi-Cal beneficiaries. The death file contains dates of death. The eligibility and

¹² Although DHCS oversees all of California's section 1915(c) waivers, other agencies and counties administer the day-to-day operations of these waivers. For examples, the California Department of Public Health operates the section 1915(c) waiver for individuals with HIV/AIDS.

¹³ To grant states additional flexibility to offer a broad range of home and community based LTSS as an alternative to institutional care, Congress authorized the use of the home and community-based waiver program under section 1915(c) of the Social Security Act in 1981. California, like other states, uses this waiver authority to make a broad range of home and community-based services available to selected populations with the level-of-need that would otherwise be offered in a Medicaid-covered institution, such as a nursing home, intermediate care facility for the mentally retarded (ICF/MR), or a hospital.

¹⁴ Assessment records for users of HCBS are maintained by the departments administering the personal care programs and waivers: the Department of Aging for the MSSP Waiver; the Department of Developmental Services for the Developmentally Disabled Waiver; the Department of Health Care Services for the Assisted Living, In-Home Operations, and Nursing Facility/ Acute Hospital Waivers and for Adult Day Health Care programs; the Department of Public Health for the AIDS Waiver; and the Department of Social Services for the IHSS program.

¹⁵ IHSS assessments are collected, in part, to determine service allotment amounts for beneficiaries.

¹⁶ The absence of assessments for the home and community based service waivers was largely mitigated because most service recipients in these waivers also receive In-Home Supportive Services.

death files combined helped CaMRI establish recipient enrollment in and disenrollment from Medi-Cal during the study period.

To enable the identification and extraction of Medicare records of dual eligibles, DHCS created a Finders File containing identifying information about the study population. CaMRI originally asked CMS to send only those Medicare claims and assessment records for the study population as identified in the Finder’s File. (A discussion of the matching process by CMS is provided later in this report.)

Table 2 describes the data files CaMRI obtained from Medi-Cal. Although some of the original files include data for a broader population, just those data pertaining to the Medi-Cal participants within this study were used by CaMRI.

Table 2. State Medi-Cal Files Collected, CYs 2005-2008

File Type	Description	CaMRI’s Purpose
Claims Data		
Paid Claims	Includes cost and utilization data for DHCS claims. Data includes all state plan and waiver services used by beneficiaries, including physician visits, hospital admissions and prescription drugs. Beneficiary demographic information, specifying gender, race, age, and diagnoses (used for provider billing), is included. This file also includes information about whether beneficiaries are dually enrolled in Medicare. Cost-sharing liability information is included as well as amounts paid by Medi-Cal.	Used to describe Medi-Cal spending and frequency of service use, among others.
Patient Discharge Database (PDD, hospital discharge abstracts)	Includes diagnostic information, procedure codes, length of stay, and discharge information for each inpatient hospital stay.	Used to identify inpatient hospital stays and diagnoses. Used primarily for managed care recipients for whom other claims data were not available.
Beneficiary Assessments		
Case Management, Information and Payrolling System (CMIPS) for In-Home Supportive Services (IHSS)	Contains personal information for recipients of IHSS, including names, addresses, Social Security numbers, phone numbers, age, ethnicity, languages spoken, income, disability levels, and reasons for terminating IHSS, if appropriate. File also includes information on beneficiaries’ limitations in activities of daily living (ADLs) and instrumental activities of daily living (IADLs). Further, information on providers, including whether a provider is an immediate family member, relative, friend, or health care organization is included, among other information.	Used to describe the functional health status of recipients and their living arrangements, among other purposes.
Client Development Evaluation Report (CDER)	Includes demographic information for Medi-Cal beneficiaries with developmental disabilities, including information about their ADLs, IADLs, cognitive impairment, diagnoses, among others. The file also includes information about claims, eligibility, and purchase of service information.	Used to describe the functional and cognitive health status of recipients with developmental disabilities and their living arrangements, among other purposes.

File Type	Description	CaMRI's Purpose
Other Files		
Eligibility Files	Contains month-to-month enrollment information for Medi-Cal beneficiaries as well as birth dates, ethnicity, gender, languages spoken, among other demographic information.	Used to establish recipient enrollment and disenrollment information from Medi-Cal during the study period. Also used to identify enrollment in Medi-Cal managed care.
Death Records	Describes vital statistics of Medi-Cal beneficiaries, including date of death. This file is used to determine when a beneficiary is no longer part of the study group.	Used to identify deaths during the study period.
Matching File		
Finder's File	Contains personal identifying information of all Medi-Cal beneficiaries during 2005-2008, including Medi-Cal IDs, Social Security numbers, birth dates, and genders.	Used by CMS to identify Medicaid beneficiaries in CMS' Medicare data.

Source: CaMRI.

Medicare Data

Claims files describe service use, costs, and provider information for each beneficiary encounter with a Medicare-covered service. They also include diagnosis and procedure codes.¹⁷ These claims files describe Medicare-covered services under Parts A and B.¹⁸ Claims data from Medicare Parts C (Medicare Advantage) and D (prescription drugs) are not included in this study.¹⁹ CMS collects different Medicare claims files from different provider groups and settings of care, including (1) inpatient hospitals, (2) outpatient hospitals, clinics, renal dialysis centers, outpatient rehabilitation centers, among others, (3) skilled nursing facilities, (4) hospices, (5) home health agencies, (6) durable medical equipment, (7) physicians, labs, and ambulances collected in a Carrier File, and (8) a consolidated file for all services rendered in skilled nursing facilities and inpatient hospital settings. CaMRI acquired all of these claims files for the study. Medicare claims data are used by CaMRI to estimate the Medicare-covered service costs of dual eligible beneficiaries.

Certain post-acute Medicare providers also conduct assessments of their service users. These files contain information about beneficiaries' health status, functional and cognitive impairments, and social and environmental factors at different stages of a beneficiary's encounter with a provider. Each provider uses a distinct assessment tool. For this study, CaMRI assembled the

¹⁷ For some providers (e.g., hospitals, skilled nursing facilities and home health agencies), these codes are used to determine payment amounts (e.g., more complicated health diagnoses can result in higher Medicare provider payments).

¹⁸ Part A (Hospital Insurance) covers inpatient hospital services, skilled nursing care, and home health and hospice care. Part B (Supplementary Medical Insurance) covers physician services, outpatient services, and some home health and preventive services.

¹⁹ Part C (Medicare Advantage) is a private managed care plan option covering all Parts A and B services, except hospice. Part D covers prescription drug benefits.

following assessment files: (1) Minimum Data Set (MDS)²⁰ collected by skilled nursing facilities, (2) Outcome and Assessment Information Set (OASIS)²¹ collected by home health agencies, (3) IRF-PAI collected by inpatient rehabilitation facilities, and (4) Swing Bed collected by acute care hospitals for patients who receive skilled nursing facility services in their hospital's inpatient settings. Assessment files are used by CaMRI primarily to identify functional and cognitive impairments as well as living arrangements of recipients when they first initiate a service. Assessment files are also used to evaluate differences across the risk factors of users by service and care setting (e.g., home and community based settings versus nursing homes). Finally, because nursing homes and home health agencies conduct MDS and OASIS assessments on all of their patients, these data can be used to evaluate the health status of individuals during Medi-Cal covered nursing home stays and home health visits.

Enrollment data in Medicare are tracked in the Denominator file. This file also contains information about beneficiaries' demographic characteristics and participation in Medicare managed care, among other information. It is linked to Medicare claims data – including payments for outpatient, home health, carrier (office visit, labs), skilled nursing facility services, hospitalizations, hospice, and durable medical equipment through a common Medicare beneficiary ID. CaMRI is using the Denominator file to determine when Medi-Cal recipients of LTSS became eligible for Medicare.

Once CMS' contractor, Buccaneer, assembled the Medicare data for delivery to CaMRI, CMS created a Cross-Walk file with Medicare ID and Social Security numbers (named, "Ben ID–SSN number cross-walk file"). This file contained personal identifying information to assist CaMRI and DHCS with the linking of the Medi-Cal and Medicare data. CMS sent this crosswalk file directly to DHCS, who linked the Medicare identifier onto Medi-Cal records behind the DHCS firewall for privacy and security purposes. (This linking process is discussed in more detail.)

Table 3 describes the Medicare files collected from CMS for the dually eligible study participants.

²⁰ The Long Term Care Minimum Data Set (MDS) is a standardized, primary screening and assessment tool of health status which forms the foundation of the comprehensive assessment for all residents of long-term care facilities certified to participate in Medicare or Medicaid.

²¹ The Home Health Outcome and Assessment Information Set (OASIS) contains data items that were developed for measuring patient outcomes for the purpose of performance improvement in home health care.

Table 3. Medicare Data Files Collected, CYs 2005-2008

File Type	Description	CaMRI's Purpose
Claims Data		
Inpatient Hospital File	Contains claims of inpatient hospital providers seeking reimbursement. Among the data elements are diagnoses, (ICD-9 diagnosis), procedures (ICD-9 procedure code), diagnosis related groups (DRGs), ^a dates of service, reimbursement amounts, hospital providers, and beneficiary demographic information.	Used to identify inpatient hospital stays and diagnoses. Used to estimate the associated Medicare expenditures for the dual eligibles in the study.
Outpatient	Contains claims of institutional outpatient providers. Among the data elements are claims from hospital outpatient departments, rural health clinics, renal dialysis facilities, outpatient rehabilitation facilities, comprehensive outpatient rehabilitation facilities, and community mental health centers. Also included are diagnosis and procedure codes (i.e., ICD-9 diagnosis, ICD-9 procedure code, CMS Common Procedure Coding System (HCPCS)), dates of service, reimbursement amounts, outpatient provider numbers, and beneficiary demographic information.	Used to identify outpatient service use and compile the associated expenditures for the dual eligibles in the study.
Skilled Nursing Facility (SNF)	Contains claims submitted by SNF providers. Among the data elements are diagnosis and procedure codes (i.e., ICD-9 diagnosis and procedure code), dates of service, reimbursement amounts, provider numbers, and beneficiary demographic information.	Used to identify SNF care use and expenditures for the dual eligibles in the study.
Hospice	Contains claims submitted by hospice providers. These data describe the level of hospice care received, (e.g., routine home care, inpatient respite care), the diagnosis code for the terminal illness (ICD-9 diagnosis), dates of service, reimbursement amounts, provider numbers, and beneficiary demographic information.	Used to identify hospice care users and the associated hospice-specific expenditure for the dual eligibles in the study.
Home Health (HH)	Contains data submitted by home health agencies (HHAs). Among the data elements are number of visits, type of visit (e.g., skilled-nursing care, home health aides, physical therapy, speech therapy, occupational therapy, and medical social services), diagnosis (ICD-9 diagnosis), reimbursement amounts, provider numbers, visit dates, and beneficiary demographic information.	Used to identify HH care use and the associated expenditures for the dual eligibles in the study.
Durable Medical Equipment (DME)	Contains claims submitted by DME suppliers. Among the data elements are diagnosis, (ICD-9 diagnosis), services provided (CMS Common Procedure Coding System (HCPCS) codes), dates of service, reimbursement amount, provider numbers, and beneficiary demographic information.	Used to identify DME users and DME expenditures for the dual eligibles in the study.

File Type	Description	CaMRI's Purpose
Carrier (physicians, lab, ambulance)	Contains claims submitted by non-institutional providers, such as physicians, physician assistants, clinical social workers, nurse practitioners, independent clinical laboratories, ambulance providers and free-standing ambulatory surgical centers. Among the data elements are diagnosis and procedure codes (i.e., ICD-9 diagnosis, and CMS HCPCS codes), service dates, reimbursement amounts, non-institutional provider numbers (e.g., UPIN, PIN, NPI), and beneficiary demographic information.	Used to identify physicians, physician assistants, clinical social workers, laboratories, and ambulance service use or encounters. Used to estimate expenditures for the dual eligibles in the study.
Medicare Provider Analysis and Review (MEDPAR)	Contains consolidated annual cost information on inpatient hospital and SNF stays. Records summarize all services rendered to a beneficiary from the time of admission to a facility through a discharge.	Used to estimate the annual inpatient hospital and SNF service use and expenditures for the dual eligibles in the study.
Beneficiary Assessments		
Minimum Data Set (MDS)	SNFs use MDS to evaluate and record the health status, functional capabilities, and cognitive limitations of residents in nursing homes. The MDS file contains items that measure physical, psychological and psychosocial functioning. The items in the MDS give a multidimensional view of the patient's functional capacities. SNFs collect MDS assessments from all nursing home residents, regardless of payer.	Used to look at the health status of SNF users at the time of admission, the health outcomes associated with SNF care, and to compare expenditures and outcomes over time among recipients, among others. MDS assessments are also used to look at the health status of Medi-Cal-covered nursing home stays for all individuals in the study.
Outcome and Assessment Information Set (OASIS)	Among other things, OASIS includes information on a patient's functional status and service needs during the beneficiary's HH episode (or episodes). OASIS data items address socio-demographic, environmental, support system, health status, functional status, cognitive limitations and health service utilization characteristics of the patient. HH agencies collect OASIS assessments from all HH users, regardless of payer.	Used to look at the health status of HH users at the time of service initiation and over an episode of care. Used to compare expenditures and outcomes among recipients, among others. OASIS assessments are also used to look at the health status of Medi-Cal-covered home health visits for all individuals in the study population.
Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI)	Contains data items that were developed primarily for payment purposes for IRFs. The data address the physical, cognitive, functional, and psychosocial status of patients. These assessments are collected on all Medicare Part A fee-for-service patients who receive services under Part A from an IRF at admission and upon discharge.	Used to look at the health status of IRF users at time of admission, the health outcomes associated with IRF care, and to compare expenditures and outcomes among recipients, among others.
Swing Bed	Swing-Bed Providers are hospitals that use their beds, as needed, to provide either acute or SNF care. These assessments are used for payment purposes.	Used to look at the health status of swing bed users at the time of admission, the health outcomes associated with swing bed care, and to compare expenditures and outcomes among recipients, among others.

File Type	Description	CaMRI's Purpose
Other Files		
Denominator File	Contains demographic and enrollment information about each beneficiary in Medicare during a calendar year. Some of the information contained in this file includes the beneficiary's unique identifier, state and county codes, zip code, date of birth, date of death, sex, race, age, and whether a beneficiary is enrolled in managed care.	Used to determine when Medi-Cal beneficiaries become dually enrolled in Medicare, among others.
Beneficiary Demographics (Walk-around)		
Cross-Walk	CMS created a "Ben ID-SSN number cross-walk file." Once the beneficiary ID is attached to state data files, DHCS provides CaMRI with a limited data set from which all direct identifiers, such as Social Security numbers and date of birth were removed.	Used by DHCS to place a CMS identifier onto state data files so that Medi-Cal and Medicare records can be linked for dually eligible study participants.

Source: ResDAC website, www.resdac.org, and a variety of letters, applications and approval documents produced by CaMRI for this study.

^a Medicare pays acute care hospitals using a prospectively determined payment for each discharge. In part, these payments are based on the relative resource use associated with the Diagnosis Related Groups (DRGs) to which patients are assigned. Starting in FY 2008, Medicare began phasing in an update to the DRG classification system, called the Medicare Severity DRG (MS-DRG) system.

Data Linking and Cleaning

To link the various datasets, a unique identifier had to be created. The process of establishing a unique identifier that could link across multiple federal and state databases required that a limited number of individuals, working on behalf of CMS and DHCS have access to individually identified data containing personal health information (PHI). Both DHCS and CMS created files with individually identifying information that were used to match Medi-Cal and Medicare records. Specifically, DHCS created a "Finder's File" containing identifying information including Social Security numbers, gender and date of birth of all Medi-Cal recipients during CYs 2005 through 2008. DHCS sent this file to CMS for use in identifying all dually eligible Medi-Cal beneficiaries. CMS' contractor, Buccaneer, attached a unique identifier to the DHCS finders file that created a cross walk between Medi-Cal and Medicare data files. This Cross-Walk file was sent to DHCS so that the unique federal identifier could be added to state records for all dual eligibles who used Medi-Cal LTSS during the study period. The attachment of the federal identifier onto the state records enables CaMRI to link federal and state records for dual eligible beneficiaries without having the need for personal identifying information about these individuals.

For all of the Medicare records, Buccaneer, on behalf of CMS, created a new or "pseudo" identification number, unique for each recipient and common for the recipients across the multiple files. The new identification number replaced the original Medicare and Social Security numbers on these files. Buccaneer also created a file that cross-walked the new identification numbers with the Medi-Cal numbers in the Finder's File which was sent to DHCS.

DHCS assigned unique pseudo-Medi-Cal identification numbers to these files. These pseudo-Medicaid identification numbers were assigned to records (such as claims) based on a match between the pseudo-Medicare ID, Medicaid IDs. All direct identifiers, such as Social Security numbers, were removed from all Medicare and Medi-Cal files before their transmission to CaMRI.

Once the data files were received at CaMRI, several procedures were implemented to assess the file linkages and to identify potentially problematic records. These procedures included:

- Identification of recipients with multiple IDs (i.e., an individual with one pseudo-Medi-Cal ID and multiple pseudo-Medicare IDs; or an individual with multiple pseudo-Medi-Cal IDs and a single pseudo-Medicare ID; or an individual with multiple Medi-Cal and Medicare pseudo-IDs) totaled 2,116 recipients, or 0.13% of the total study population. If individuals with multiple IDs consistently matched to demographic information in the eligibility file (i.e., sex, date of birth, date of death), the records were assumed to be correctly linked. If one or more of the multiple identification numbers did not match to the eligibility items, then records with these IDs were excluded from the study data file.
- A second screening test used the annual Medi-Cal and Medicare eligibility files to assess the consistency in the demographic items of sex, date of birth, and date of death, and both within and between these programs. The total number of recipients is approximately 1.58 million. Of these, 867,594 (55%) are in both Medicaid and Medicare data (dual eligibles). The balance, 711,243 (45%) are solely enrolled in Medicaid. Approximately 93.8% of cases agreed on these comparisons across all observation years. Another 3.6% agreed on sex and date of death, and 0.9% agreed on sex and date of birth. Together, these matches account for 98.2% of the cases.
- Further, screening of cases will be conducted as data analysis proceeds. This will include the examination of extreme outliers on service use and expenditures, such as for hospital stays, and on the inclusion or exclusion of recipients depending on the criteria used to define LTSS.

While the matching rate of study cases to the demographic data within the enrollment file is very high, the assignment of identification numbers was largely outside the control of both DHCS and the CaMRI team.

Study Timeline

Table 4 describes the range of activities required for this study and a timeline specifying the dates in which these activities occurred. These activities are described in the following categories: (1) initial discussions between CaMRI, DHCS, The SCAN Foundation, and CMS's contractor, ResDAC; (2) CaMRI formally requests Medicare data from CMS and Medi-Cal data from DHCS; (3) matching Medi-Cal and Medicare files; (4) data files sent to CaMRI; and (5) files are linked and cleaned.

Table 4. Study Timeline, June 2009 to Present

Initial Discussions between CaMRI, DHCS, The SCAN Foundation, and ResDAC	
CaMRI begins discussions with DHCS and The SCAN Foundation about the possibility of conducting a longitudinal study of Medi-Cal beneficiaries with LTSS needs in June 2009. The SCAN Foundation formally agrees to put up seed money for this study in the fall of 2009.	
June 2009 – March 2010	
Study Design	<p>CaMRI, DHCS, The SCAN Foundation discuss:</p> <ul style="list-style-type: none"> • Scope of the project; • Available Medi-Cal data for study purposes; • Available Medicare data for dual eligibles; • Methods for defining users of Medi-Cal LTSS; and • People who would need to be involved to ensure the project’s success. <p>CaMRI, DHCS, and ResDAC discuss:</p> <ul style="list-style-type: none"> • Methods for including Medicare data in the analysis; • Contents of the Medicare data files that would be used for the study; • Application process for CMS data; and • Cost estimates.
Data Sharing Contract	DHCS determines that no additional data use agreements are needed. Specifically, DHCS’ legal department asserts that the sharing of large data files are needed for this project would fall under the scope of the existing Business Associate Agreement between DHCS and CaMRI. The need for a separate agreement between CaMRI and the DDS for CDER data is later identified.
Data: Paid Claims, Eligibility and Death Records	CaMRI and DHCS determine the kinds of files that CaMRI will use for the study, including paid claims, eligibility and death records files.
Data: Assessments	The DHCS project manager collects assessments from home and community based service waivers and state LTSS for CaMRI to review. ^a Meetings with state department staff are organized to compare the data elements within each assessment file and to determine which files can be transmitted electronically to CaMRI via DHCS. CaMRI then develops a narrative comparison and matrix describing the distinct assessment files and distributes it to DHCS’ sister departments for feedback. With assistance from the DHCS Project Manager and DHCS’ sister departments, CaMRI makes a decision as to which assessment files it will use.
CaMRI formally requests Medicare data from CMS and Medi-Cal data from DHCS	
CaMRI, as a subcontractor to DHCS, formally requests Medicare data from CMS, via ResDAC. Medi-Cal claims data are already housed within DHCS and can be transferred to CaMRI under an existing Business Associate Agreement. DHCS requests that the California Department of Social Services (DSS), the California Department of Public Health (DPH), and the DDS send assessment data to DHCS so that it can be shared with CaMRI for this evaluation project.	
Request for CMS Data (December 2009 – September 2010)	
December 2009	CaMRI sends an application to ResDAC, the CMS vendor responsible for processing data requests for researchers, for the claims, assessment, enrollment and Cross-Walk files needed for this study for CYs 2005 through 2008.
March 2010	In response to comments from ResDAC, CaMRI modifies the application and resubmits it.
May 2010	The application for these files is approved by ResDAC. It is then forwarded to CMS’ privacy board for approval.
June 2010	CMS approves the application for data for this study. CaMRI pays CMS \$119,000 for data.
September 2010	CMS processes the payment and sends approval to Buccaneer, CMS’ data contractor, for release of the data. DHCS sends the Finder’s File to CMS to enable Buccaneer to identify which Medicare beneficiaries in California are in the study population.
Request for State files	
June 2010	The amendment to the Master Agreement between CaMRI and DHCS adds this study, using Medi-Cal claims and assessment data, to the list of evaluations that CaMRI conducts on behalf of DHCS. This brings the project under the umbrella of the CaMRI-DHCS Business Associate Agreement.
June 2010	California’s DDS determines that CaMRI’s Business Associate Agreement with DHCS does not meet the state’s higher standards for data sharing required for accessing the CDER data file. ^b To obtain these data, DHCS needs to submit an application to California’s Institutional Review Board (IRB). DHCS submits an application on behalf of CaMRI on a fast-track basis.

July 2010	CaMRI's request for death records from the California Death Statistical Master File is signed by DHCS and submitted to the California DPH.
July 2010	DHCS receives approval from the California IRB for the sharing of CDER data with CaMRI. DHCS is required to prepare an amended application for these data on behalf of CaMRI to DDS. This application has not yet been prepared by DHCS nor sent to DDS.
December 2010	As the data access agreements are not yet fully resolved, CaMRI requests limited access to some personally identifiable records for the CDER file so as to continue to make progress on the data linking and cleaning process.
October 2010	DHCS has not yet submitted the amended application to DDS for the CDER data.
Matching Medi-Cal and Medicare Files	
DHCS develops a "Finder's File" that includes personal identifying information for the Medi-Cal study population. This file is sent to CMS so that Buccaneer, CMS' contractor, can identify the Medicare records that match the Medi-Cal study population for delivery to DHCS and CaMRI. Once Buccaneer selects the data files, it transports them to DHCS to be linked with the Medi-Cal files using a Cross-Walk file. This linking process takes place behind DHCS' firewall. DHCS then sends the linked and de-identified files to CaMRI.	
November 2009	The methodology for creating a Medi-Cal "Finders File" is developed by DHCS.
May 2010	The methodology for creating the Medi-Cal "Finders File" is revised by DHCS.
October 2010	A revised file is submitted by DHCS to Buccaneer to enable them to identify whom among their Medicare beneficiaries in California are dually eligible. ^c
December 2010	Buccaneer sends the MDS and OASIS assessment files to CaMRI. These files contain statewide records for skilled nursing facility and home health users. CaMRI forwards these files to DHCS to be matched to the Finder's File and de-identified.
March 2011	Buccaneer sends the Medicare claims, denominator, IRF-PAI files to CaMRI. These files are already matched to the Finder's file and de-identified. CaMRI places these files on its secure server.
July 2011	DHCS sends MDS and OASIS assessment files to CaMRI.
Data Files Sent to CaMRI	
DHCS de-identifies files and sends them to CaMRI as they are ready. Files sent to CaMRI contain encrypted beneficiary identification numbers, among other information. Files arrive at CaMRI at different times. Data files sent to CaMRI were typically accompanied by data dictionaries.	
March 2010	DHCS begins processing the Medi-Cal claims data as they were already available within DHCS. Work began with the claims data to identify recipients of LTSS during the study period.
March 2010	Because the application for CDER data is not yet submitted by DHCS and DHCS could not move forward on identifying the study population for CaMRI among the claims records, DDS sends a list of names from its CDER data set to DHCS to facilitate the selection of the LTC recipients from the Medi-Cal paid claims file. This enables CaMRI to make progress on the data analysis while still waiting for the CDER data to arrive.
September 2010	CaMRI obtains CMIPS/IHSS assessment data from DHCS via the California Department of Social Services (CDSS). Some data are missing. Specifically, assessments for participants of the IHSS Plus waiver and missing months for one calendar year are not included. DHCS, on behalf of CaMRI, submits a supplemental request to DSS for this missing data.
September 2010	CaMRI obtains death records for 2005-2008 from DHCS via the California Department of Public Health.
September 2010	Hospital discharge data files are prepared by the California Office of Statewide Health Planning and Development (OSHPD) to be sent to DHCS for linkage to DHCS' Finders File.
December 2010	The missing months of data for the CMIPS assessments are sent to CaMRI by DHCS.
December 2010	Death records and Medi-Cal monthly eligibility records are matched by DHCS.
December 2010	The OSHPD hospital discharge abstract files are received at DHCS for linkage with DHCS' Finders File.
December 2010	Buccaneer sends the Cross-Walk file to DHCS so that certain Medicare files can be linked to the Medi-Cal files behind its firewall.
March 2011	After DHCS links the Medi-Cal claims, hospital discharge abstracts, and Medi-Cal eligibility files, it transmits these de-identified records to CaMRI.
March 2011	Death records and Medi-Cal monthly eligibility records are transmitted to CaMRI.

April 2011	DHCS sends de-identified claims files to CaMRI.
October 2011	CaMRI determines that the MDS and OASIS files are missing records for numerous months. ResDAC is contacted for assistance.
Present	CDER assessment data are not yet available to CaMRI.
Files are Linked and Cleaned	
Once CaMRI starts receiving files, the team of statisticians and data analysts begin linking the files and cleaning discrepancies. This process is complex and is ongoing.	
March 2010 - March 2011	CaMRI develops common codes and obtains data dictionaries for all files.
March 2010 – March 2011	CaMRI creates and tests common measures across data sets. Priority is given to creating condition classifications for all recipients consolidating the ICD-9 diagnostic codes, creating common service use categories, and identifying the functional, and living arrangement items available among the data sets.
June 2010 – Present	CaMRI tests and verifies the accuracy of file linkages. This involves checking and resolving inconsistencies in demographic items among the annual Medi-Cal and Medicare eligibility records, identification of cases with multiple ID numbers, and identification of extreme outliers for hospital days and expenditures.

Source: CaMRI.

^a With the exception of IHSS, some MSSP assessment data, and the CDER file, all Medi-Cal assessments are in paper format or limited to electronic records within case manager’s computers.

^b Section 4514 of the California Welfare and Institutions Code (WIC) imposes more stringent requirements on data sharing for persons with developmental disabilities than the standard interagency agreement between DHCS and its sister departments.

^c Medicare data are being used to identify the dually eligible adults in the study. Medi-Cal data are being used to identify the dually eligible children.

Contracts and Data Use Agreements for Data Sharing

The Privacy Act of 1974 (P.L. 93-579), as amended, protects data held by the federal government, including by CMS, for which individually identifiable information, such as name, Social Security number, or other identifying number or symbol, can be retrieved. In addition, the Health Insurance Portability and Accountability Act of 1996 (HIPAA, P.L. 104-191) places restrictions on the sharing of data containing protected health information for research purposes. As directed by HIPAA, the Secretary of the U.S. Department of Health and Human Services (HHS) adopted the HIPAA Privacy Rule to govern the protection of individual privacy in the electronic exchange of certain health information. Together, HIPAA and its corresponding federal regulation specify how a limited data set may be used and disclosed, provided the recipient enters into a data use agreement (DUA) promising specified safeguards for the relevant protected health information. California also has laws that place additional restrictions on entities intending to share such data, including requirements regarding transporting data securely.

To comply with federal and state requirements regarding privacy and security of data sharing, a number of contracts and DUAs were developed for this project. In one instance, a state law was amended to allow for the sharing of data containing personal health information (PHI) between DHCS and OSHPD, the California state department that manages the statewide Patient Discharge Database (PDD, hospital discharge abstracts). The statutory change, contracts, and DUAs are described in **Table 5**.

Table 5. Summary of Contracts and Other Agreements

Data Files	Interested Parties	Contract or Other Agreements	Statutory or Regulatory Federal and State Authorities Governing Data Sharing^a	Description
All Medicare files acquired for the study	Center for Medicare and Medicaid Services (CMS), DHCS and CaMRI	- ResDAC application and addendum -DUA	Federal -Privacy Act of 1974, HIPAA Privacy Rule, CMS' data release policies	-To identify CaMRI as an approved user of Medi-Cal and Medicare data on behalf of DHCS -To limit CaMRI's use of the data to those study purposes specified in the DUA -To specify an ending date for CaMRI's use of the administrative data
Finder's File	DHCS, CMS and Buccaneer Computer Systems & Service, Inc.	DUA	Federal -HIPAA, Health Information Technology for Economic and Clinical Health Act, Public Law 111-005 ("the HITECH Act"), HIPAA Privacy Rule and Security Rule, (45 CFR Parts 160 and 164) State - California Information Practices Act, State Administrative Manual and Health Administrative Manual	-To secure data and documents that reside in DHCS' Medi-Cal system - To ensure the integrity, security, and confidentiality of such data -To permit only appropriate disclosure and use as permitted by law
Medi-Cal Data: CMIPS	DHCS and the Department of Social Services (CDSS)	Interagency Agreement	Federal -HIPAA Privacy Rule State -State Administrative Manual and Health Administrative Manual	Agreement established with CDSS as a subcontractor to DHCS for the administration of Medi-Cal and included provisions requiring CDSS to share Medi-Cal data with DHCS
State Data: Death Records	DHCS and Department of Public Health (DPH)	Interagency Agreement	Federal -HIPAA Privacy Rule State -State Administrative Manual and Health Administrative Manual	Agreement established DPH as a subcontractor to DHCS for the administration of Medi-Cal and included provisions requiring DPH to share Medi-Cal data with DHCS
Medi-Cal Data: CDER	CaMRI and Department of Developmental Services (DDS)	CaMRI and DDS (after approval by the Agency's Institutional Review Board)	Federal -HIPPA Privacy Rule and Section 4514(e) of the Welfare and Institutions Code	State law requires that the sharing of data for individuals with developmental disabilities be first approved by the Agency's Institutional Review Board

Data Files	Interested Parties	Contract or Other Agreements	Statutory or Regulatory Federal and State Authorities Governing Data Sharing ^a	Description
State Patient Discharge Database (hospital discharge abstracts)	DHCS and the Office of Statewide Health Planning and Development (OSHPD)	State Statute	State -Section 128730(a)(3) of the California Health and Safety Code	-For OSHPD to share these data with DHCS, the California Health and Safety Code was amended, adding DHCS as a department to whom these data may be made available -DHCS must ensure that the patient's rights to confidentiality not be violated
Medi-Cal Data: claims, assessment, and eligibility files	CaMRI and DHCS	Business Associate Agreement (BAA)	Federal -Section 1902(a)(7) of the Social Security Act of 1965, HIPAA Privacy Rule and Security Rule, (45 CFR Parts 160 and 164), and Standards for Privacy of Individually Identifiable Health Information State -State Administrative Manual and Health Administrative Manual	This agreement establishes CaMRI as a subcontractor to DHCS for purposes of evaluating the Medi-Cal program using DHCS' Medi-Cal data

Source: CaMRI.

^a The list of authorities provided in this column is not comprehensive.

Notes: DUA refers to data use agreement.

Contracts and DUAs are legally binding and assist government entities in ensuring that researchers will meet the relevant federal and state requirements. Such agreements track the transfer of data, identify their custodians, define the purpose of the data use, and limit the amount of time for which researchers may use the data.

CaMRI's findings from the data analysis will be made available to DHCS, disseminated to federal and state policy-makers, and made publicly available through The SCAN Foundation and CaMRI's forthcoming website. The data itself, however, cannot be shared with other researchers without a modification to CaMRI's DUAs and other agreements.

CaMRI's Security Measures

Because this study requires analysis of administrative data that contains personal health information, CaMRI investigators were required to seek approval from the appropriate IRBs. Distinct IRBs reside at each University of California campus. These Institutional Review Boards operate under federal authority of the U.S. Food and Drug Administration and the Office for

Human Research Protections which govern research conducted on human subjects. IRBs are designed to ensure that such research is ethical and equitable in its treatment of human subjects. Because CaMRI's investigators reside at the University of California, San Francisco and the University of California, San Diego, investigators received approval from two university IRBs.

The agreements and DUAs specific to CaMRI bind it to establish appropriate administrative, technical, and physical safeguards to protect the confidentiality of the data and to prevent their unauthorized use or access. Further they require CaMRI to ensure that data not be used via unsecured telecommunications, such as the Internet, to transmit individually identifiable or deductible information derived from the files, among other specifications.

To comply with the federal and state requirements regarding privacy and security, CaMRI is relying on existing resources, such as secure servers, with the University of California, San Francisco. It also expanded its secure computer capacity and instituted a number of security protocols. These activities are described below.

First, access to personal health information is restricted only to the study's investigators named on DUAs and IRB applications. In addition, CaMRI is storing and using these data in a password-protected secure server at the University of California, San Francisco's Information Technology Facility. This facility is used for the data files that have been stripped of direct identifiers. The University of California, San Francisco's main Sun Microsystems server also meets a high level of security standards required for storing personal health information. It is protected by the university's firewall.

The production server at the University of California, San Francisco Information Technology Facility, running the Sun Microsystems operating system, is housed in dedicated computer machine rooms containing emergency backup power and authorization-based limited access. The computer and corresponding disk storage are locked in a computer cabinet within the computer room with keys to the server and rack distributed only to key personnel.

In addition, according to industry best practices, all software services and corresponding ports known to be security risks are disabled. Administrative access to databases and corresponding data are limited to the IT facility team using Secure Shell (SSH) and/or a Virtual Private Network (VPN). All databases reside behind industry-strength firewalls, with the direct identifiers being placed in a separate and secure location. Only completely de-identified data are placed on a computer with File Transfer Protocol (FTP) access for analysts on this project.

UCSF's security procedures included encryption of the physical data received from CMS. These are kept in a locked cabinet with keys available only to designated personnel. Password protection are also used at the server and web portal levels for all transactions that allow entry, editing and linking of data and that provide access to sensitive subject data or administrative privileges. All users are required to change their password every 90 days, following strict protocol for strong passwords.

Since the data are behind multiple firewalls, monitored regularly, and accessible only to key personnel, the risk of unlawful penetration is not a significant data safeguard concern. In

addition, all research files maintained by CaMRI contain only a pseudo identification number while the key linking the pseudo number with the direct identifier is in a separate secure location. This further reduces the risk of data breach.

Finally, individually identifiable or deducible data are not being transmitted by unsecured telecommunications. Lastly, any printed material containing individual identifiers will be shredded.

Issues, Challenges and Obstacles

The following describes some of the issues, challenges and obstacles CaMRI encountered throughout the process of acquiring Medi-Cal and Medicare data from the state and federal governments, and in linking and cleaning these data for analysis. This discussion is intended to assist researchers, states and CMS in preparing and planning for future studies of this kind.

Medi-Cal Data Housed Across Multiple State Departments

DHCS is both the administrative department overseeing the Medi-Cal program and a direct purchaser of certain services. DHCS also subcontracts administrative oversight and implementation of certain components of the Medi-Cal program to its sister departments within the California Department of Health and Human Services (Agency). Counties also play a significant role in administering and providing Medi-Cal services. **Table 6** shows how Medi-Cal is organized within the Agency and the counties.

Table 6. Departments and Counties Responsible for Administering, Purchasing, and Overseeing Medi-Cal

Department	Medi-Cal Oversight Responsibilities
Program Administration	
California Department of Health Care Services (DHCS)	Administers Medi-Cal Sets eligibility, benefit, and fee-for-service provider payments and beneficiary cost-sharing levels
County Social Services Departments	Conducts eligibility determinations Oversees enrollment and recertifications
Acute Care	
DHCS	Contracts with health plans to manage care Reimburses fee-for-service primary and acute care
County Mental Health Departments	Deliver mental health services
Long-Term Supportive Services	
DHCS	Reimburses LTSS Oversees all section 1915(c) waivers Reimburses nursing homes Operates Nursing Facility/Acute Hospital Waiver Operates In-Home Operations Waiver Operates Assisted Living Waiver Operates Pediatric Palliative Care Waiver
California Department of Aging	Oversees the Adult Day Health program Operates the Multipurpose Senior Services Program
Department of Social Services (DSS)	Oversees the In-Home Supportive Services program

Department	Medi-Cal Oversight Responsibilities
County Social Services Departments	Implements the In-Home Supportive Services program (IHSS)
Department of Developmental Services (DDS)	Oversees community-based services and regional centers for case management Operates HCBS Waiver for Persons with Developmental Disabilities Operates developmental centers (intermediate care facilities for the mentally retarded, ICF/MRs)
Department of Public Health	Operates HIV/AIDS Waiver

Source: Interviews with DHCS staff and California Health Care Almanac: Medi-Cal Facts and Figures, California HealthCare Foundation, September 2009. See <http://www.chcf.org/publications/2009/09/medical-facts-and-figures>

Because DHCS is responsible for reimbursement for most Medi-Cal services, DHCS’ sister departments and counties send their claims data on a regular basis to DHCS. With the exception of the Medi-Cal managed care plans, DHCS houses claims data for all of Medi-Cal. Because the Patient Discharge Database (PDD)/hospital discharge abstracts data contain hospital information from all payers in California and are not limited to Medi-Cal beneficiaries, a statutory change had to be made to allow OSHPD to share these data with DHCS. (An explanation of these agreements and the statutory change can be found in the section of this report entitled Contracts and Data Use Agreements.)

Unlike claims data, assessment data are held by multiple departments within the Agency that subcontract to DHCS. With the exception of the DDS,²² the interagency agreements establishing these sister departments as subcontractors required them to send their assessment data to DHCS upon its request. DHCS could then send these data to CaMRI, as authorized under CaMRI’s Business Associate Agreement with DHCS. **Table 7** shows the state departments responsible for housing each of the data files used for this study.

²² According to California state law, the sharing of data containing information about individuals with developmental disabilities for research purposes is subject to a higher standard. This resulted in the establishment of a separate DUA between CaMRI and DDS for the CDER data. See the section of this report entitled Contracts and Data Use Agreements for Data Sharing.

Table 7. Source of Data Files by Department within the California Health and Human Services Agency (Agency), CYs 2005-2008

File Type	State Departments within the Agency
State Plan Claims Data	
Paid Claims File	Department of Health and Human Services (DHCS)
Patient Discharge Database (PDD, hospital discharge abstracts)	Office of Statewide Health Planning and Development (OSHPD)
Beneficiary Assessments	
CMIPS: In-Home Supportive Services (IHSS) File	California Department of Social Services (CDSS)
Client Development Evaluation Report (CDER)	Department of Developmental Services (DDS)
Beneficiary Enrollment	
Eligibility Files	DHCS
Death Records	California Department of Public Health (DPH)
Matching File for Medicare Records	
Finder's File	Produced by DHCS and sent to Center for Medicare and Medicaid Services (CMS)

Source: CaMRI.

Limited Access to Managed Care Data

Some of the study's participants are mandatorily enrolled in county-operated Medi-Cal managed care plans; others are voluntarily enrolled. Such plans arrange for the delivery of acute care services, such as physician, hospital and clinic services for Medi-Cal-only recipients. During the study period of interest, certain Medi-Cal-covered services were carved out of managed care and delivered through FFS (fee-for-service) providers. Such services included LTSS, dental, mental health and children's services for the seriously ill and disabled.

Contracts between the state and managed care plans require plans to provide all those medically necessary state plan services to participants for which they are entitled. For this assurance, California makes risk adjusted capitated payments to managed care plans (on a per member per month basis). These rates are set on an annual basis. Unlike FFS providers, plans are not paid per service delivered, or episode of care. As a result, plans do not need to submit claims data for each service delivered to DHCS for reimbursement purposes. Without a requirement to submit such data to DHCS, any claims and assessment information that these plans collect are used for internal purposes only.

As a result, DHCS has limited claims data describing Medi-Cal managed care participation. Specifically, it has no claims data from the managed care plans for acute care service use, such as hospital, physician and clinic services. It also does not have assessment data from providers for managed care participants using these acute care services. DHCS does, however, have claims data for individuals receiving carved out services, such as the LTSS.

CaMRI tried to work around this data gap for its study participants in managed care by collecting hospital discharge abstract data that are submitted by California's hospitals to the Office of Statewide Health Planning and Development within the Agency. Although these data describe

diagnoses and hospitalizations for all payers, CaMRI needed just those records that describe the hospital inpatient stays for the Medi-Cal's beneficiaries with LTSS use during CYs 2005 through 2008. Claims data for non-hospital acute and post-acute care services for Medi-Cal managed care participants are not available.

Because certain state plan and waiver services are carved out of Medi-Cal managed care and paid by DHCS on a fee-for-service basis (e.g., section 1915(c) waivers, Adult Day Health services, and In-Home Supportive Services), some claims and assessment data for these services are available for Medi-Cal managed care participants during CYs 2005 through 2008. Medicare claims and assessment data are available for those Medi-Cal managed care participants who are dually enrolled in Medicare fee for service programs. For dually enrolled beneficiaries in Medicare Advantage Plans, Medicare claims data are not available.

Discrepancies in the Delivery of Medicare Data

Linking and cleaning Medi-Cal and Medicare data files to prepare them for reliable analysis requires an experienced and well-organized staff of statisticians and data analysts. These individuals must also have extraordinary technical expertise and deep knowledge of how Medicaid and Medicare operate. Even with such a team, the process for linking and cleaning data is slow and extraordinarily complex. For CaMRI, this process is further complicated by some factors that are out of the control of the CaMRI team. One example of this is that the data sent to CaMRI by CMS did not match the written data request. Although these discrepancies are being resolved with the help of ResDAC and its subcontractor, Buccaneer, this challenge slows the linking and cleaning process and imposes some significant challenges to the data team. The following describe some of these discrepancies.

CaMRI requested the MDS and OASIS files for CYs 2005 through 2008. The MDS data originally sent to CaMRI by CMS contains only assessments from CYs 2005 through part of CY 2007. In addition, the OASIS data originally sent to CaMRI are missing assessments from October, November and December of CY 2008.

In addition, CaMRI originally requested Medicare files that matched the Social Security numbers of just those study participants with LTSS claims. However, because CMS didn't match the MDS and OASIS data to these Social Security numbers, CMS sent MDS and OASIS assessments for all nursing home and home health users in the state of California between CYs 2005 and 2008. These data contain PHI and include a much larger population than is needed by CaMRI. As a result, the files had to be sent to DHCS to be appropriately de-identified and linked to the study population behind its firewall. This process considerably delayed CaMRI in getting started with incorporating these files into the larger database for analytic purposes.

Further, the original Swing Bed file sent to CaMRI from CMS did not contain Social Security numbers for beneficiaries and thus could not be matched to the corresponding Medi-Cal records. CMS had to resend this file containing the Social Security numbers.

Finally, Buccaneer (CMS' data contractor) sent records with multiple Medicare identification numbers for individual beneficiaries. This problem could not be resolved via CaMRI's linking

and cleaning process because CaMRI did not have access to the PHI that would be needed to correctly link these individuals to the Medi-Cal records. As a consequence CaMRI may have to exclude about two thousand dual eligible beneficiaries from the study cohort.

Partnering with State Government

The success of this study depends upon a strong partnership between CaMRI and the state department operating the Medi-Cal program, DHCS. Because the study requires Medi-Cal data containing PHI, DHCS first had to agree to hire CaMRI as its contractor to conduct an evaluation of Medi-Cal on its behalf. This arrangement is facilitated with a standing renewable Master Agreement between CaMRI and DHCS. Under this Agreement, CaMRI investigators negotiated numerous evaluation projects of Medi-Cal, including this one, on DHCS' behalf.

In addition to a formal agreement with the state department, dedicated staff within DHCS had to be made available to support this project. The DHCS Director assigned a single staff person within its long-term care division to be the coordinator for this project. This individual has invested hundreds of hours to date in assisting CaMRI in (1) defining the scope of the project, (2) coordinating with DHCS' legal department and those legal departments within DHCS' sister departments, (3) recruiting and coordinating with DHCS' data analytics team, and (4) further representing CaMRI's project to other interested parties within DHCS and its sister departments, among other activities.

In addition to the project manager, members of DHCS' legal department contributed to this study by overseeing the legal and regulatory agreements between (1) DHCS and its sister departments, (2) DHCS and CaMRI, and (3) DHCS and CMS to ensure that they contained the necessary provisions to comply with federal and state requirements. These legal experts also ensured that various individuals responsible for transferring data understood the legal protocol for avoiding a breach in security of the PHI. DHCS' legal staff also worked with their sister departments – DDS, CDSS, DPH, and OSHPD – to demonstrate that their interagency agreements allowed for the transfer of Medi-Cal assessment data to DHCS. In response to these requests, members of the legal teams within DHCS' sister departments also reviewed their contracts with DHCS to authorize, when appropriate, the transfer of assessment data from their departments to DHCS.

Finally, a team of data analysts within DHCS are dedicating substantial amounts of their time to linking, cleaning, and securely transporting data to CaMRI. After CaMRI receives these data, these DHCS staff will continue to assist CaMRI's data team in sorting through some of the complex discrepancies within the Medi-Cal data and with the linked federal data.

Limitations to Analysis of Administrative Data

The administrative data assembled for this study provide a unique opportunity to understand the amount, type, and costs of all services provided to individuals with LTSS needs in both Medicaid and Medicare. However, even after the expense of money and staff time to create this linked dataset, there are several important limitations to using it to draw conclusions about the effectiveness of LTSS. The following describes some of these limitations.

Diagnoses in Claims Data

Claims data from hospitals, physicians and certain other providers contain ICD-9 diagnoses codes (International Statistical Classification of Diseases and Related Health Problems). These ICD-9 codes are one of the main measures of illness and diagnoses in the Medicaid and Medicare claims data. Although these codes provide a rich source of diagnostic data, they do not provide detailed information on clinical severity, such as the degree of sugar control and end-organ damage among beneficiaries with diabetes. When looking at how health affects the use of certain services on health status, information on the differences in severity of a patient's illness is important. Without this information, our study results cannot consider how differences in the effectiveness of certain services might be based on the degree of severity of a beneficiary's diagnosis. CaMRI does, however, make adjustments for hierarchical condition groupings and for co-morbidities.

In addition, health status measures that rely on diagnostic information in administrative data may be affected by utilization. That is, individuals with higher rates of utilization may receive more diagnoses independent of their actual health status. As a result, study results may be affected by service use. These effects are mitigated by including multiple sources (e.g., Medicaid as well as Medicare claims, assessments, and hospital discharge abstracts) of information, as well as by increasing the number of years from which the data are obtained.

Finally, differences also exist between Medicare and Medi-Cal in the recording of diagnoses. Because there is less incentive to record diagnoses in Medi-Cal claims,²³ Medi-Cal claims contain fewer diagnoses than Medicare claims (e.g., there are two diagnoses variables in Medi-Cal claims data and seven in Medicare claims data). This makes it more difficult to draw comparisons on health status and the effectiveness of service use across Medi-Cal and Medicare.

Assessment Data

Assessment files provide information about recipients' demographic characteristics, health status, and functional and cognitive limitations, among other information. The availability of these data should help CaMRI gain a more robust understanding of beneficiaries' health status and how it is associated with the use of LTSS. However, assessment data are not collected on all Medi-Cal beneficiaries who use LTSS. Some of the beneficiaries who fall into this group are those whose only use of LTSS is the receipt of durable medical equipment. Further, some of the assessment information that was collected by Medi-Cal providers was unusable in this study because they are not available electronically. For example, beneficiaries whose only use of LTSS is adult day health care had assessments that were not available to the CaMRI team. For this study, CaMRI was limited to using Medi-Cal assessment data for users of the IHHS benefit and for individuals with developmental disabilities. Fortunately, IHHS is the most commonly

²³ Medicare payments to certain providers, such as acute care hospitals, are based on the relative resource use associated with a patient's diagnoses. More complex diagnoses result in higher payments. Under this system, providers have incentives to report diagnoses on their claims records. Medi-Cal payments are calculated based on service use and not on diagnoses. As a result, Medi-Cal providers do not have the same incentive to report diagnoses on Medi-Cal claims records.

received LTSS benefit so that the majority of beneficiaries have some assessment data. Medicare MDS and OASIS data also include assessments for Medi-Cal nursing home and home health users. For the dual eligible population, CaMRI relied on assessment data collected by Medicare’s post-acute providers (i.e., SNFs, HHAs, IRFs, and swing bed hospitals). However, without assessment data from all Medi-Cal and Medicare providers, some users of LTSS in our claims data will have no information on their health status for analytic purposes (other than diagnoses on claims files).

In addition, assessment data collected by Medi-Cal and Medicare provider entities and DDS differ by provider. Each provider collects different information. As a result, this information is often non-comparable. For example, IHSS providers collect detailed information about limitations in activities of daily living and instrumental activities of daily living. Skilled nursing facilities, however, collect information on beneficiaries’ limitations in ADLs but no information about IADLs. Regarding limitations in mental and behavioral function, IHSS providers collect information on three variables: memory, orientation, and judgment. SNFs collect information on many more variables, including memory, decision making, consciousness, behavioral measures of delirium, and resident’s interpersonal relationships, among others. Further, IHSS collects information about beneficiaries’ access to informal caregivers. SNFs do not. These differences in data collected complicate the ability to conduct longitudinal analyses of beneficiaries as they transition across care settings throughout the study period.

Further, the frequencies with which assessment data are collected vary by provider type. For example, IHSS providers generally collect assessment data on a bi-annual basis. More frequent assessments may be collected when individuals’ demonstrate significant changes in health status or condition. While the number of IHSS hours authorized for Medi-Cal payment is linked to the IHSS assessments, changes in hours may be authorized without a reassessment. This limits the precision with which assessments are updated based on changes in status. HHAs, on the other hand, are required by federal regulation to conduct assessments (1) within 24 hours of an initial episode of care, (2) after 5 days, (3) whenever a patient’s condition warrants a new assessment based on a major decline or improvement in status, (4) at least every 60 days, (5) within 48 hours of the patient’s return to home after a 24 hour hospitalization, and (6) at discharge from an HHA episode of care.²⁴ Because Medicare payments to HHAs are linked to these OASIS assessments, the frequency with which assessments are made may be more reliable. The different frequencies of data collection across provider settings may steer certain analyses toward questions related to provider settings in which there are stronger, more reliable data, and away from other service settings in which less reliable data are available.

Non-Covered Health Services and LTSS

Although Medicaid and Medicare cover a broad range of preventive, acute, post-acute and LTSS, they do not cover all health services and LTSS that a beneficiary may use. For example, most personal care received by people with disabilities is delivered by informal providers—family and

²⁴ See 42 CFR 484.55.

friends—who give care without compensation. Private long-term care insurance also pays for some long-term care services. People also pay out-of-pocket for care. Information on the services used by our study population that are not covered by Medi-Cal or Medicare is not available for this study. The absence of this utilization and assessment information will limit CaMRI's ability to assess whether some individuals who are eligible and have a need for LTSS are not receiving them.

Study Period

The fixed time window of the study (January 2005 through December 2008) limits the ability to compare beneficiaries who were already receiving LTSS when the study period began with those who obtained these services after January 2005. The issue is that those who were receiving services as of January 2005 have an unknown duration of exposure to LTSS prior to January 2005. The left censoring of the data could introduce bias in drawing conclusions about the effectiveness of LTSS. This bias can be explored by performing a sensitivity analysis among a subset of beneficiaries who started using LTSS after the beginning of the study period.

Finally, although CaMRI received the most current data available at the time the study began, much has happened in health care and LTSS since CY 2008.²⁵ Because the data are time limited, CaMRI's findings cannot reflect how the Medi-Cal and Medicare programs have changed since January 2009.

General Observations and Recommendations

As federal and state policy-makers, Medicaid and Medicare administrators, and beneficiary advocates work to identify methods for improving systems of care for individuals with LTSS needs and dual eligibles, a new body of research will be needed. Such research can benefit from Medicaid and Medicare's administrative data to evaluate the services provided to individuals by need and along the entire care continuum. To be useful, research must be comprehensive, scientifically valid, and timely.

With knowledge of the challenges and obstacles CaMRI faced while assembling a database of Medicaid and Medicare administrative data for precisely these research purposes, the authors of this report are putting forth a number of recommendations. We believe these recommendations can reduce the challenges and obstacles future researchers and states might face when attempting to undertake a project of this magnitude. If addressed, such recommendations would better enable federal and state decision-makers to make policies based on authoritative research so as to ensure effective change.

CaMRI recommends the following:

²⁵ For example, many study participants are enrolled in adult day health care during the CYs 2005 through 2008. As of December 1, 2011, however, the adult day health Medi-Cal benefit will be eliminated.

CMS

1. CMS should establish more efficient, timely and routine approaches to sharing accurate and complete Medicare claims and assessment data with state Medicaid programs and researchers. Making these data more broadly available on a real-time basis has the potential to improve the performance and efficiency of public programs for dually eligible populations. However, a balance needs to be found between the speed of delivering such data and the risk that such speed might compromise the data's quality.
2. CMS can play a leadership role in standardizing the definitions of certain demographic and other variables across and within federal and state data sets. Further, the value of these data can be enhanced by establishing additional procedures for auditing and editing data to enhance their accuracy. Finally, creating and disseminating tools that states can use to assemble and analyze Medicaid and Medicare data could reduce the barriers to research of this kind.

States

3. States often lack adequate capacity to fully explore linked administrative databases or to apply sophisticated statistical techniques to draw valid inferences from administrative data. Partnerships between state Medicaid programs and researchers offer an opportunity to learn from their claims and assessment data. To attract research partners, state Medicaid programs need to minimize financial and administrative barriers for obtaining data in a timely fashion. This might be done by dedicating selected staff within the state Medicaid department, or via contract (such as CMS's arrangement with ResDAC whose primary function is to facilitate the sharing of Medicare data with researchers). It can also be done by developing strategies and protocols for researchers to link Medicaid with Medicare data.²⁶
4. Given the complex interplay in many states between their Medicaid departments and other state departments operating parts of the Medicaid program, states should ensure that interagency agreements include provisions clearly describing the responsibility of these departments to share all Medicaid data with the lead Medicaid department. (The lack of clarity within these agreements at the outset of CaMRI's study was one of the greatest sources of delay in CaMRI's receipt of the study data.)
5. In many states, including California, Medicaid services are increasingly being delivered through managed care. When managed care plans are paid on a capitated basis they do not submit claims to the Medicaid agency. To support a state's needs to monitor the

²⁶ Locally, this project was a catalyst for the Medi-Cal program in California to develop a gateway for data sharing in partnership with the University of California. Processes that have been established as a part of this project should contribute to more rapid availability of data on LTSS and other health care services in California going forward.

performance of Medicaid managed care, Medicaid agencies should require plans to routinely submit accurate and complete information on all encounters.²⁷

6. Agencies that make assessments of Medicaid beneficiaries to determine eligibility for LTSS should establish a standardized, electronic data survey of key measures so that comparable assessments can be made of beneficiaries' functional and cognitive status over time, regardless of care setting. New measures should be established to evaluate innovative care strategies that attempt to make improvements in systems of care for recipients of LTSS and dual eligibles. Such measures can evaluate care transitions, care coordination, managed care, the cost-effectiveness of care setting placements for LTSS, among other measures.²⁸

²⁷ In the absence of these data, CaMRI is working with Medi-Cal to pioneer a strategy for using the statewide hospital discharge abstracts to capture hospitalizations among beneficiaries in Medi-Cal managed care.

²⁸ In California, many of these assessments are performed by county-based agencies that employ methods that are inconsistent across programs and counties. In many cases, information is not recorded in electronic format, limiting the state's ability to perform program evaluation on a statewide basis.

Appendix 1. GLOSSARY OF TERMS

ADL:	Activities of daily living
CaMRI:	California Medicaid Research Institute
CDDS:	California Department of Social Services
CDER:	Client Development Evaluation Report
CHIP:	Children's Health Insurance Program
CMCS:	CHIP and Survey & Certification
CMIPS	Case Management Information and Payrolling System
CMS:	Centers for Medicare and Medicaid Services
DDS:	Department of Social Services
DHCS:	California Department of Health Care Services
DME:	Durable Medical Equipment
DPH:	Department of Public Health
DRG:	Diagnosis related groups
DUA:	Data use agreement
FFS:	Fee-for-service
FTP:	File Transfer Protocol
HCBS:	Home and community based services
HCPCS:	CMS Common Procedure Coding System
HH:	Home health
HHA:	Home health agencies
HIPAA:	Health Insurance Portability and Accountability Act
HITECH:	Health Information Technology for Economic and Clinical Health Act
IADL:	Instrumental activities of daily living
ICD-9	International Statistical Classification of Diseases and Related Health Problems
ICF/MR:	Institutional care facility for the mentally retarded (ICF/MR)
IHSS:	In-Home Supportive Services
IRB:	Institutional Review Board
IRF-PAI:	Inpatient rehabilitation facilities Patient Assessment Instrument
LTC:	Long-Term Care
LTCSS:	Long-Term Care Services and Supports
MACPAC:	Medicaid and CHIP Payment and Access Commission
Medi-Cal:	California's Medicaid program
MEDPAR:	Medicare Provider Analysis and Review
MDS:	Minimum Data Set
MS-DRG:	Medicare Severity DRG (diagnosis related groups)
MSSP:	Multipurpose Senior Service Program
OASIS:	Outcome and Assessment Information Set
OSHPD:	Office of Statewide Health Planning and Development
PACE:	Program for All-Inclusive Care for the Elderly
PDD:	Patient Discharge Database
PHI:	Personal health information
PPACA:	Patient Protection and Affordable Care Act
ResDAC:	Research Data Assistance Center

SNF: Skilled Nursing Facility
SSH: Secure Shell
VPN: Virtual Private Network
WIC: Welfare and Institutions Code