MAX and MCBS Survey Linkage, 2007-2008

Final Report

May 7, 2012

Rosalie Malsberger

MATHEMATICA
Policy Research

This page has been left blank for double-sided copying.

MAX and MCBS Survey Linkage, 2007-2008

Final Report

May 7, 2012

Rosalie Malsberger

MATHEMATICA
Policy Research

This page has been left blank for double-sided copying.

CONTENTS

This page has been left blank for double-sided copying.

TABLES

This page has been left blank for double-sided copying.

ACRONYMS

ADL	Activities of daily living
ATC	Access to Care
BENE_ID	Beneficiary identification number
C&U	Cost and Use
CER	Comparative effectiveness research
CMS	Centers for Medicare & Medicaid Services
CY	Calendar year
DOB	Date of birth
HIC	Health insurance claim number
IADL	Instrumental activities of daily living
IP	MAX inpatient claims file
LT	MAX institutional long-term care claims file
MAX	Medicaid Analytic eXtract
MCBS	Medicare Current Beneficiary Survey
MSIS	Medicaid Statistical Information System
OT	MAX other services claims file
PS	MAX person summary file
RIC	Record identification code
RX	MAX prescription drug claims file

This page has been left blank for double-sided copying.

I. INTRODUCTION

The Centers for Medicare & Medicaid Services (CMS) contracted with Mathematica Policy Research to link 2007 and 2008 survey data from the Medicare Current Beneficiary Survey (MCBS) with existing research-oriented CMS Medicaid data files known as the Medicaid Analytic eXtract (MAX). In 2007, dually eligible beneficiaries ("duals")—low-income seniors and persons with disabilities who qualify for both Medicaid and Medicare—accounted for 15 percent of Medicaid enrollees and 39 percent of total Medicaid spending (Kaiser Family Foundation 2011). Duals are among the sickest and poorest enrollees in either program (Kaiser Family Foundation 2011). A combined MCBS and MAX data set will provide a rich database of service use, expenditures, and health outcomes for duals that would not be available with either data set alone. This report summarizes how we merged the two data sets and assesses the quality of the linked MCBS-MAX records.

A. Background on the Medicare Current Beneficiary Survey (MCBS)

The MCBS, begun in 1991, is an annual, continuous, longitudinal, multipurpose survey of a representative national sample of the Medicare population. The primary goals of the MCBS are to determine expenditures and sources of payment for all services used by Medicare beneficiaries; ascertain all types of health insurance coverage, and link coverage to payment sources; and trace processes over time, such as changes in health status or spending down to Medicaid eligibility (CMS 2012). Between 12,000 and 16,000 Medicare beneficiaries are interviewed each year, and each beneficiary in the sample is interviewed a total of 12 times over four years. One-third of the sample is retired each year, with a new sample introduced in the fall survey round. MCBS data are linked to Medicare claims and administrative data, which results in a more complete utilization data set for the MCBS respondents than would be available based on interviews alone.

B. Background on the Medicaid Analytic eXtract (MAX)

MAX data are derived from CMS's Medicaid Statistical Information System (MSIS) and contain comprehensive information about Medicaid enrollees and their Medicaid-financed service utilization and expenditures in a calendar year. MSIS data contain enrollment information and Medicaid claims paid on a quarterly basis; however, because claims are typically paid several months after service use, claims can appear in MSIS for periods after the service occurred. Seven quarters of MSIS data are processed together to create the MAX data for a calendar year. The data contain individual-level demographic information, monthly enrollment status and eligibility group, managed care and waiver enrollment information, and use and costs of services (Borck et al. 2012). Data are available for every Medicaid enrollee in all 50 states and the District of Columbia.

C. Benefits from Linking MCBS and MAX Files

Although the MCBS contains information on all services covered by Medicare, it does not contain any information on benefits—particularly long-term care services—that are covered only by Medicaid. Given that more than two-thirds of Medicaid expenditures on duals in 2007 were for long-term care services (Kaiser Family Foundation 2011), this lack of information in the MCBS prevents attainment of a complete picture of expenditures for all duals. In addition, Medicare is the first payer for many services used by duals but MAX captures these services only if additional Medicaid payments are made for the enrollee (such as for coinsurance and deductibles) (Borck et al. 2012). For this reason, expenditures in MAX for Medicare-covered services for duals are substantially understated. By linking the MCBS and MAX data, researchers and policymakers can capture a more complete picture of total Medicare and Medicaid expenditures for and service utilization by duals.

In addition, the only health status information available in MAX is based on the diagnosis codes for Medicaid-financed services. MCBS has information on the health conditions, health-related behaviors, and health status of enrollees, such as activities of daily living (ADLs) and instrumental activities of daily living (IADLs). By linking MCBS and MAX data, researchers and policymakers can provide valuable insight into the health status of dual eligibles, a crucial element for comparative effectiveness research (CER).

D. Overview of Report

In Chapter II, we describe the linkage steps. In Chapter III, we describe the linkage results. In Chapter IV, we assess the quality of the linkage. Finally, in Chapter V, we summarize the report and offer advice to researchers interested in using the linked MCBS-MAX data.

This page has been left blank for double-sided copying.

II. MCBS AND MAX LINKAGE STEPS

We merged the MCBS and MAX files using data corresponding to calendar years (CY) 2007 and 2008. We chose CY 2007 and 2008 because they were the most recent MCBS years available. To merge the files, we followed these seven steps:

1. Create a crosswalk to link beneficiaries in the MCBS and MAX files (done by Buccaneer)
2. Extract data from the MCBS files
3. Merge the MCBS files into one analysis file
4. Extract data from the MAX files
5. Merge the MAX files into one analysis file
6. Merge the MCBS-MAX crosswalk to the MCBS file
7. Merge the crosswalked MCBS file to the MAX file

The steps are described in detail below.

A. Create the MCBS-MAX Crosswalk

The first step in linking the two sets of files was to create a crosswalk to link dual eligibles included in the MCBS survey to the same set of beneficiaries in the MAX files. The two datasets use different identification numbers: the MCBS files use BASE_ID while the MAX files use a combination of MSIS_ID and STATE_CD (jointly referred to as "MAXID"). A CMS contractor, Buccaneer, created a crosswalk between the BASE_ID and the MAXID. To create it, Buccaneer used the CMS "enterprise cross-reference file," which contains the beneficiary identification number (BENE_ID) assigned to every new Medicare or Medicaid beneficiary record stored in the CMS data warehouse.

Buccaneer first focused on all Medicare beneficiaries ever sampled for the MCBS survey. They extracted two "finder" files from the enterprise cross-reference file. The first one

contained all unique combinations of BENE_ID and BASE_ID. The second one contained all unique combinations of health insurance claim number (HIC)[1] and BASE_ID. Buccaneer created the latter finder file because some beneficiaries are not assigned a non-missing BENE_ID. By using the HIC instead of the BENE_ID, we're assured that all Medicare beneficiaries ever sampled for the MCBS survey are selected from the cross-reference file.

Next, Buccaneer focused on the Medicaid beneficiaries. They used the enterprise cross-reference file and the MAX Person Summary (PS) files. First, they extracted all unique combinations of MAXID and BENE_ID from the enterprise cross-reference file. Then, they extracted two files from the MAX PS files: the first one contained the MSIS HIC and MAXID and the second one contained the Medicare HIC and MAXID. Buccaneer merged and de-duplicated these two HIC extracts to create one HIC-MAXID file.

Finally, Buccaneer merged the two Medicare crosswalk files to the two Medicaid crosswalk files using the BENE_ID and HIC respectively. These two crosswalked files were concatenated and de-duplicated to create a final BASE_ID-to-MAXID crosswalk. The final crosswalk contains 48,304 records with a total of 42,656 unique BASE_IDs.[2] It should be noted that the record count of this MCBS-MAX crosswalk is much higher than the number of records Mathematica linked between the two data sources, because the crosswalk contains all duals who were identified in any year of MCBS, while Mathematica's linked files (described in this report) were only for CY 2007 and 2008.

---

[1] HIC is an identification number that CMS assigns to an individual when he or she first enrolls in Medicare.

[2] Buccaneer reported that over 99.5 percent of the BASE_ID-MAXID crosswalk matches came from the BENE_ID–BENE_ID linkage; less than 0.5 percent of the matches came from the HIC–HIC linkage.

B. Extract Data from the MCBS Files

MCBS survey data are released in two annual files: Access to Care (ATC) and Cost and Use (C&U) files. The ATC files provide information on access to, satisfaction with, and usual source of care, as well as beneficiary health indicators, such as ADLs and IADLs. The ATC files are released sooner than the C&U files, but they only include persons who were enrolled in Medicare for the full year. Persons who entered Medicare during the year (new eligibles) and persons who died during the year are not included. In contrast, the C&U files include persons who were enrolled in Medicare at any point in the CY. The C&U files take longer to release because they link detailed Medicare claims to survey-reported cost and utilization data (CMS 2012). The C&U files present a comprehensive view of Medicare-covered health services received and the amount and sources of payment for those services (for example, beneficiary copayments and deductibles or third-party payer coverage) (CMS 2012).

The ATC file for one year of data consists of 20 content-specific data files. The C&U file for one year of data consists of 25 content-specific files. Each file is called a "record identification code" (RIC) file. Because the content in the ATC and C&U overlap, the filenames are often the same. For example, the RIC 1 file contains demographic information, the RIC 4 file contains information about Medicaid coverage, and both of these RIC files are available as an ATC and C&U file.

Because our objective was simply to link MCBS to MAX, we extracted a very small set of variables from a small set of RIC files. Table II.1 lists the variables extracted from each of the MCBS files, and Table II.2 shows the record counts for each of the MCBS files.

Table II.1. Variables Extracted from MCBS and MAX Files

Source File	Extracted Variable	Extracted Variable Description
MCBS Access to Care (ATC) Files		
RIC 1	BASEID	Unique Identifier
	D_DOB	Date of birth
	ROSTSEX	Sex
	D_RACE2	Race
	HISPORIG	Of Hispanic origin
RIC 4	BASEID	Unique Identifier
	D_MCARE	Medicare coverage
	D_MCAID	Medicaid eligibility
	MCAIDHMO	Enrolled in a Medicaid HMO
RIC A	BASEID	Unique Identifier
	H_DOD	Date of death
	H_MCSW	Some Medicaid eligibility for the year
	H_MCDE01 - H_MCDE12	Type of Medicaid eligibility, by month
	H_RESST	State of residence
MCBS Cost & Use (C&U) Files		
RIC 1	BASEID	Unique Identifier
	D_DOB	Date of birth
	ROSTSEX	Sex
	D_RACE	Race
	D_ETHNIC	Ethnicity
RIC 4	BASEID	Unique Identifier
	D_CARE	Medicare coverage
	D_CAID	Medicaid eligibility
	D_HMO	HMO coverage
RIC A	BASEID	Unique Identifier
	H_DOD	Date of death
	H_MCSW	Some Medicaid eligibility for the year
	H_MCDE01 - H_MCDE12	Type of Medicaid eligibility, by month
	H_RESST	State of residence
MAX Person Summary (PS) File		
	MSIS_ID	MSIS Identification Code
	STATE_CD	State
	EL_SSN	SSN
	EL_HIC_NUM	HIC number from MSIS
	EDB_HIC_NUM	HIC number from Medicare Enrollment Database
	EL_DOB	Date of birth
	EL_SEX_CD	Sex
	EL_RACE_ETHNCY_CD	Race/ethnicity code
	MDCR_DOD	Date of death
	EL_MDCR_DUAL_ANN	Annual Medicare dual code

Table II.2. Record Counts from MCBS and MAX Files, 2007 and 2008

Source File	2007	2008
MCBS ATC Files (RIC 1, RIC 4, RIC A)	15,806	14,547
MCBS C&U Files (RIC 1, RIC 4, RIC A)	11,995	11,723
MAX PS File	61,673,088	63,842,647

Source: MCBS and MAX files, 2007 and 2008.

C. Merge the MCBS Files into One Analysis File

We next merged each set of ATC and C&U RIC files together by BASE_ID (still keeping the years and file types separate). We confirmed the RIC file merges were all one-to-one merges; that is, within each year and file type, there was only one record per BASE_ID in each of the merged RIC files. We then merged together both years of MCBS data for a given file type to create two 2007/2008 files: one for ATC and one for C&U. Among 4,821 BASE_IDs identified as Medicaid beneficiaries in at least one year of the ATC files, 23 percent were beneficiaries in 2007 only, 30 percent were beneficiaries in 2008 only, and 47 percent were beneficiaries in both years. Among the 4,299 BASE_IDs identified as Medicaid beneficiaries in at least one year of the C&U files, 32 percent were beneficiaries in 2007 only, 29 percent were beneficiaries in 2008 only, and 39 percent were beneficiaries in both years. We then performed a final merge to combine the ATC and C&U files, so that we only had to merge one MCBS file to the MAX data.

D. Extract Data from the MAX Files

For each CY for each state, there are five MAX data files: (1) the inpatient claims file (IP); (2) the institutional long-term care claims file (LT); (3) the prescription drug claims file (RX); (4) the other services claims file (OT); and (5) the person summary (PS) file, which contains demographic, insurance, Medicaid/Medicare program information, and utilization summaries for inpatient hospital, institutional long-term care, prescription drug, and other services.

Because our objective was simply to link MCBS to MAX, we extracted a very small set of variables from the state-level MAX PS files for 2007 and 2008 (Table II.1). We then concatenated all 51 state-level extract files into a national file (still keeping the years separate). Table II.2 shows the record counts for each year.

E. Merge the MAX Files into One Analysis File

We then merged both years of MAX PS data together by MAXID to create one MAX file. Out of a total of 74,361,922 unique values of MAXID, 17 percent were in the 2007 file only, 14 percent were in the 2008 file only, and 69 percent were in both files.

F. Merge MCBS-MAX Crosswalk to the MCBS File

Next, we merged the Buccaneer BASE_ID-MAXID crosswalk to the MCBS file (created in Step 3) by BASE_ID. Out of a total of 22,606 unique BASE_IDs in the merged MCBS file, 29 percent (6,653 records) merged to the crosswalk. This means that 29 percent of those in the 2007 or 2008 MCBS ATC or C&U samples were identified as Medicaid eligible through MAX data at some point in time, which is comparable to the approximately 21 percent of the Medicare population that an external data source identified as dually eligible in 2009 (Kaiser Family Foundation 2011). Our percentage is somewhat higher than the external estimate because the Buccaneer crosswalk includes persons who were ever enrolled in Medicaid in 1999-2008. This means that the 6,653 records can include a Medicare beneficiary in the 2008 MCBS file who was enrolled in Medicaid in 2006 but was not enrolled in 2008, whereas the external source statistic applies to Medicare beneficiaries enrolled in Medicaid in the single year 2009. We only kept the MCBS records that linked to the crosswalk.

G. Merge of the Crosswalked MCBS File to the MAX File

As the final step in the linkage, we merged the crosswalked MCBS file to the MAX file by MAXID. Out of the 6,653 crosswalked MCBS records, 77 percent merged to the 2007/2008

MAX file. This final set of 5,161 duals is the one that we next assessed for the accuracy of the

linkage between MCBS and MAX.

This page has been left blank for double-sided copying.

III. LINKAGE RESULTS

Even though we combined the MCBS files into one file before the merge, we used indicator variables to keep track of whether the record came from the 2007 and/or 2008 ATC files and whether the record came from the 2007 and/or 2008 C&U files. Table III.1 shows the linkage of the four MCBS files (2007 and 2008 ATC, 2007 and 2008 C&U) to the Buccaneer crosswalk and to the 2007/2008 MAX file, for those Medicare beneficiaries identified as Medicaid eligible in the MCBS (D_MCAID = 1, 2, or 3 in the ATC file or D_CAID = 1, 2, or 3 in the C&U file). The table presents two sets of counts: duplicated and unduplicated counts of Medicaid-Medicare beneficiaries in each of the four files, before and after linkage to the crosswalk, and after linkage to the MAX file. The final row shows the duplicated count for MCBS records that link to the MAX file in the same year. There are two sets of counts because some BASE_IDs (records in the MCBS) link to more than one MAXID (records in the MAX file). (Possible explanations for these duplicate records are discussed in the subsequent section.) The duplicated counts include all of these records, while the unduplicated counts include only one MAX record per BASE_ID.

Table III.1 is important because it shows the decrease in the sample size of the duals when we linked to the Buccaneer crosswalk and then to the MAX file. In the ATC 2007 file, for example, there are 3,700 duals. After linking to the Buccaneer crosswalk, there are 3,502 duals. After linking to the MAX file, there are 3,453 duals. Thus, through the linkage process, the number of MCBS duals declines by 7 to 13 percent, depending on the MCBS file type and year. The unduplicated counts in row 3 of Table III.1 are the sample sizes likely to be applicable to most studies using the linked MCBS-MAX files.

Because we anticipate that analyses with merged MCBS-MAX files may want to also use only records that are in both years of data in one or both sets of MCBS files, and in one or both years of MAX data, Table III.2 lists sample sizes after enforcing these restrictions. Depending

on the restrictions put on the sample, the sample size of duals can decline by roughly two-thirds

from the numbers shown in Table III.1.

Table III.1. Number of Medicaid Beneficiaries Linked in MCBS and MAX, by MCBS File Type

	ATC		C&U	
	2007	2008	2007	2008
Before linkage	3,700	3,397	3,029	2,929
After linkage to crosswalk (unduplicated count—one record per BASEID)	3,502	3,168	2,720	2,600
After linkage to PS file (either year) (unduplicated count—one record per BASEID)	3,453	3,127	2,666	2,556
After linkage to crosswalk (duplicated count—multiple records per BASEID)	4,093	3,667	3,159	2,992
After linkage to PS file (either year) (duplicated count—multiple records per BASEID)	3,543	3,195	2,727	2,606
After linkage to PS file (same year) (duplicated count—multiple records per BASEID)	3,485	3,156	2,683	2,564

Source: MCBS-MAX file, 2007-2008.

Note: Medicaid beneficiaries are identified through the variables D_MCAID (ATC file) or D_CAID (C&U file).

Table III.2. MCBS and MAX Linkage Results

	Number of Medicaid Beneficiaries
All four MCBS files and either 2007 or 2008 MAX PS file	1,339
All four MCBS files and both MAX PS files	1,323
Both MCBS ATC files and either 2007 or 2008 MAX PS file	2,120
Both MCBS ATC files and both MAX PS files	2,097
Both MCBS C&U files and either 2007 or 2008 MAX PS file	1,535
Both MCBS C&U files and both MAX PS files	1,508

Source: MCBS-MAX file, 2007-2008.

Note: Medicaid beneficiaries are identified through the variables D_MCAID (ATC file) or D_CAID (C&U file).

IV. QUALITY OF LINKAGE

We performed multiple analyses to assess the quality of the linkage. First, we compared the date of birth (DOB) and sex from the MCBS to the MAX files for those identified as Medicaid beneficiaries in the MCBS who linked to the MAX file in the same year. We also matched on whether the date of death for the beneficiary was either present or missing in the MCBS versus MAX files. For the two annual ATC files, the date of death variables matched on all records. For the two annual C&U files, less than 0.2 percent failed to match in either year. Additionally, we matched race/ethnicity from the MCBS and MAX files and, as anticipated, we found a much lower matching rate due to the subjective nature of race/ethnicity reporting and to differences in race/ethnicity categories reported in the two data sources. Consequently, we focus our discussion of the quality of the linkages primarily on the comparison of the DOB and sex. We also present the comparison separately for MCBS records that linked to one MAX record from those that linked to more than one MAX record. We considered a record to "match" if the MCBS and MAX values were the same for at least two of three data elements: sex, month of birth, and year of birth.

Table IV.1 shows that across the four files for the single-linked records, roughly 98 percent of records had matching values of DOB and sex, which suggests these records are correctly linked. About 0.5 percent of records have the same DOB but nonmatching sex, which suggests these records are also correctly linked, with sex likely being an entry error on one of the records. The majority of the remaining records have a matching value of sex but a nonmatching value of either month *or* year of birth only, so it is likely these are also correctly linked, with the month or year being an entry error on one of the records. A very small number of records remain, which have either a matching value of sex and nonmatching month *and* year of birth, or

Table IV.1. Comparison of Linked MCBS and MAX Demographic Variables for Single Records

	ATC 2007		ATC 2008		C&U 2007		C&U 2008	
	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Same DOB and same sex	3,281	98.06	2,992	98.19	2,521	97.64	2,431	97.83
Same DOB and different sex	14	0.42	9	0.30	12	0.46	10	0.40
MAX sex ≠ 'U'	13	0.39	9	0.30	0	0.00	0	0.00
MAX sex = 'U'	1	0.03	0	0.00	12	0.46	10	0.40
Different DOB and same sex	48	1.43	42	1.38	41	1.59	40	1.61
Same year of birth, different month	15	0.45	16	0.53	11	0.43	13	0.52
Same month of birth, different year	33	0.99	26	0.85	28	1.08	27	1.09
Different year and month of birth	0	0.00	0	0.00	2	0.08	0	0.00
Different DOB and different sex (MAX sex ≠ 'U')	0	0.00	1	0.03	8	0.31	4	0.16
Same year of birth, different month	0	0.00	0	0.00	0	0.00	0	0.00
Same month of birth, different year	0	0.00	1	0.03	0	0.00	0	0.00
Different year and month of birth	0	0.00	0	0.00	8	0.31	4	0.16
Different DOB and different sex (MAX sex = 'U')	3	0.09	3	0.10	0	0.00	0	0.00
Same year of birth, different month	0	0.00	0	0.00	0	0.00	0	0.00
Same month of birth, different year	0	0.00	0	0.00	0	0.00	0	0.00
Different year and month of birth	3	0.09	3	0.10	0	0.00	0	0.00
Total Number of Single MAX Records	**3,346**		**3,047**		**2,582**		**2,485**	

Source: MCBS-MAX file, 2007-2008.

Notes: This table shows Medicaid beneficiaries who linked to MCBS and MAX in the same year.

MAX sex = 'U' represents unknown sex.

nonmatching sex and nonmatching month *and/or* year of birth. This suggests these records are not correctly linked. Overall, there are very few records that appear to be incorrectly linked. These results indicate that the MCBS-MAX crosswalk produced very plausible linkages for a very high percentage of individuals and is of high quality.

Table IV.2 shows that there is a similar distribution among the MCBS records that linked to multiple MAX records, although the percentage of linked records with matching values of DOB and sex declines to roughly 90 percent. About half of the remaining 10 percent of matched records have the same sex but nonmatching DOB. More than half of the nonmatching DOB records have nonmatching month *and* year of birth, which indicates these records are not correctly linked. We assume also that the remaining linked records with both nonmatching sex and DOB are incorrectly linked. Among MAX records that link to more than one MCBS record, therefore, the crosswalk appears to be fairly good, although somewhat less accurate than among the single-linked records.

We did not drop any records in the linked MCBS and MAX research file because of nonmatching demographic information or multiple linkages. We attempted to reconcile nonmatching information if possible, and if not possible, to create a categorical variable (DROP_REC) to indicate why a researcher might want to drop the record (Table IV.3). In cases where we reconciled information, we chose to use the MCBS value, because MCBS interviews are in person and may be assumed to be more accurate than MAX records.

As shown in Table IV.4, we took a second look at the MCBS records that linked to multiple MAX records. We examined MCBS records that linked to two MAX records (the large majority for each MCBS file type and year) and examined those that linked to more than two MAX records. The table shows that among those that linked to two MAX records, roughly 20 percent

Table IV.2. Comparison of Linked MCBS and MAX Demographic Variables for Multiple Records

	ATC 2007		ATC 2008		C&U 2007		C&U 2008	
	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Same DOB and same sex	128	92.09	99	90.83	88	87.13	72	91.14
Same DOB and different sex	0	0.00	0	0.00	0	0.00	0	0.00
MAX sex ≠ 'U'	0	0.00	0	0.00	0	0.00	0	0.00
MAX sex = 'U'	0	0.00	0	0.00	0	0.00	0	0.00
Different DOB and same sex	6	4.32	7	6.42	7	6.93	5	6.33
Same year of birth, different month	0	0.00	1	0.92	2	1.98	1	1.27
Same month of birth, different year	0	0.00	2	1.83	0	0.00	0	0.00
Different year and month of birth	6	4.32	4	3.67	5	4.95	4	5.06
Different DOB and different sex (MAX sex ≠ 'U')	4	2.88	2	1.83	6	5.94	2	2.53
Same year of birth, different month	0	0.00	0	0.00	0	0.00	0	0.00
Same month of birth, different year	0	0.00	0	0.00	0	0.00	0	0.00
Different year and month of birth	4	2.88	2	1.83	6	5.94	2	2.53
Different DOB and different sex (MAX sex = 'U')	1	0.72	1	0.92	0	0.00	0	0.00
Same year of birth, different month	0	0.00	0	0.00	0	0.00	0	0.00
Same month of birth, different year	0	0.00	0	0.00	0	0.00	0	0.00
Different year and month of birth	1	0.72	1	0.92	0	0.00	0	0.00
Total Number of Multiple MAX Records	**139**		**109**		**101**		**79**	

Source: MCBS-MAX file, 2007-2008.

Notes: This table shows Medicaid beneficiaries who linked to MCBS and MAX in the same year.

MAX sex = 'U' represents unknown sex.

Table IV.3. Rules to Reconcile Single- and Multiple-Linked Records with Nonmatching Demographic Information

Demographic Information	Single-Linked		Multiple-Linked	
	Rule	Number of Records Affected	Rule	Number of Records Affected
MCBS Record Linked to One or More MAX Records				
Same DOB, same sex	None needed	N/A	Assume same person: Set variable DUP_SAME[a] = 1	387
Same DOB, different sex	Assume same person: Set sex to MCBS sex if MCBS sex is male or female. Set sex to MAX sex if MCBS sex is unknown.	45	N/A	0
Different month *or* year of DOB, same sex	Assume same person: Set DOB to MCBS DOB.	169	Assume same person: Set DOB to MCBS DOB. Set variable DUP_SAME[a] = 2	6
Different month *and* year of DOB, same sex	Assume not same person: Set variable DROP_REC[b] = 1	2	Assume not same person: Set variable DROP_REC[b] = 3	19
Different month *and/or* year of DOB, different sex	Assume not same person: Set variable DROP_REC[b] = 2	19	Assume not same person: Set variable DROP_REC[b] = 4	16
MAX Record Linked to More than One MCBS Record				
Same DOB, same sex	N/A	N/A	Assume same person: Set variable DUP_SAME[a] = 1	2
Different month *and/or* year of DOB, different sex	N/A	N/A	Assume not same person: Set variable DROP_REC[b] = 5	2

Source: MCBS-MAX file, 2007-2008.

[a] The categorical variable DUP_SAME takes on the two values listed above to indicate why we assume the duplicate record is the same person. Otherwise, it is set to zero.
[b] The categorical variable DROP_REC takes on the five values listed above to indicate why a researcher may want to drop the record. Otherwise, it is set to zero.

Table IV.4. Comparison of Year and State Variables for Multiple MAX Records That Link to a Single MCBS Records

	ATC 2007		ATC 2008		C&U 2007		C&U 2008	
	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Two MAX Records Linked to a Single MCBS Record								
Same state, same year	15	18.52	12	19.35	12	21.05	11	22.92
Same state, both in one year, only one in the other year	3	3.70	1	1.61	3	5.26	1	2.08
Different state, same year	13	16.05	9	14.52	7	12.28	8	16.67
Different state, both in one year, only one in the other year	50	61.73	40	64.52	33	57.89	28	58.33
Different state, different year	0	0.00	0	0.00	2	3.51	0	0.00
Total Number of MCBS Records with Two MAX Records	**81**		**62**		**57**		**48**	
More than Two MAX Records Linked to a Single MCBS Record								
All different states, one in both years, two or more in one year	4	100.00	2	66.67	2	100.00	1	100.00
All different states, two in both years, one in only one year	0	0.00	1	33.33	0	0.00	0	0.00
Total Number of MCBS Records with More than Two MAX Records	**4**		**3**		**2**		**1**	

Source: MCBS-MAX file, 2007-2008.

Notes: This table shows year and state comparisons across the full crosswalked file (only duals who linked to PS, multiple records per BASEID).

The counts in this table do not match the counts of duplicated records in Table IV.2 because this table has one count for each MCBS record, and Table IV.2 has one count for each MAX record to which an MCBS record linked.

of these pairs of MAX records were in the same state in the same year, between 2 percent and 5 percent were in the same state with both records in one MAX year and in only one state in the other MAX year, and the remaining records were in different states with some combination of MAX years. Among the MCBS records that linked to more than two MAX records, all but one set of these records were in different states, with one record in both years and two or more in only one year.

Previous studies confirm that the same people appear in MAX with different values of MSIS_ID and STATE_CD (Czajka and Verghese, 2011). There are two reasons an MCBS record could link to multiple MAX records. First, people can be enrolled in Medicaid in more than one state (and would be included in MAX with a new MSIS_ID and STATE_CD); for example, this situation might include individuals who moved into a nursing home or moved closer to their children. Second, the state can assign more than one MSIS_ID to the same person (the person would be included in MAX with a new MSIS_ID but the same STATE_CD); this should not happen, but it sometimes does. If the multiple MAX records that matched to the same MCBS record match on at least two of the match criteria (sex, month of birth, year of birth), we created a categorical variable (DUP_SAME) to indicate that we think the multiple records are the same person (Table IV.3).

In addition to MCBS records linking to multiple MAX records, there were two instances where one MAX record linked to two MCBS records. We examined DOB and sex for the MAX record and the MCBS records, and in both cases it was clear which MCBS record correctly linked to the MAX record: DOB and sex matched with the MAX record for one of the two MCBS records and did not match for the other. Mathematica recommends that the two nonmatching records be dropped from the analysis file. We used a categorical variable (DROP_REC) to indicate which records to drop (Table IV.3).

This page has been left blank for double-sided copying.

V. CONCLUSION

For future research purposes, Mathematica linked Medicare survey data from the MCBS with Medicaid administrative data from the MAX files, using a crosswalk to link beneficiaries from each data set. This report details the methods used to create the linked data set. A comparison of available demographic information across the MCBS and MAX files for MCBS records that linked to MAX yielded a high percentage of matching information, suggesting that the vast majority of linkages are accurate and that the MCBS-MAX crosswalk is of high quality.

Overall sample sizes for the linked MCBS-MAX files are around 2,500 to 3,000 beneficiaries. In selecting research questions to address, researchers will need to evaluate whether the sample sizes for particular variables, even when combining two years of data, are large enough to identify statistically significant results.

This page has been left blank for double-sided copying.

REFERENCES

Borck, Rosemary, Allison Hedley Dodd, Ashley Zlatinov, Shinu Verghese, Rosalie Malsberger, and Cara Petroski. "The Medicaid Analytic eXtract 2008 Chartbook." Draft report submitted to the Centers for Medicare & Medicaid Services. Washington, DC: Mathematica Policy Research, 2012.

Centers for Medicare & Medicaid Services. *Overview Medicare Current Beneficiary Survey (MCBS).* Washington, DC: Centers for Medicare and Medicaid Services, 2012. https://www.cms.gov/MCBS/. Accessed February 2012.

Centers for Medicare & Medicaid Services. *Overview Medicare Current Beneficiary Survey (MCBS), Release Overview.* Washington, DC: Centers for Medicare and Medicaid Services, 2012. https://www.cms.gov/MCBS/downloads/Segment_Overview.pdf. Accessed February 2012.

Czajka, John, and Shinu Verghese. "Continued Development of the Medicaid Analytic Extract Enrollee Master (MAXEM) File." Final report submitted to the Centers for Medicare & Medicaid Services. Washington, DC: Mathematica Policy Research, 2011.

Kaiser Family Foundation. *Dual Eligibles: Medicaid's Role for Low-Income Medicare Beneficiaries.* Washington, DC: Henry J. Kaiser Family Foundation, 2011. http://www.kff.org/medicaid/upload/4091-08.pdf. Accessed February 2012

This page has been left blank for double-sided copying.