

The Centers for Medicare & Medicaid Services' Office of Research, Development, and Information (ORDI) strives to make information available to all. Nevertheless, portions of our files including charts, tables, and graphics may be difficult to read using assistive technology.

Persons with disabilities experiencing problems accessing portions of any file should contact ORDI through e-mail at [ORDI\\_508\\_Compliance@cms.hhs.gov](mailto:ORDI_508_Compliance@cms.hhs.gov).

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

**Expanding Access to Medicaid  
Data for Research: Feasibility of  
a MAX Sample**

Final Report

December 30, 2010

Audra Wenzlow

Ellen Bouchery

John Czajka



**MATHEMATICA**  
Policy Research, Inc.

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

Contract Number:  
HHSM-500-2005-00025I

Mathematica Reference Number:  
06759.470

Submitted to:  
Centers for Medicare & Medicaid Services  
Office of Research, Development, and  
Information  
Mail Stop C3-23-05  
7500 Security Blvd.  
Baltimore, MD 21244-1850  
Project Officer: Dave Baugh, William Clark

Submitted by:  
Mathematica Policy Research  
600 Maryland Ave., SW  
Suite 550  
Washington, DC 20024-2512  
Telephone: (202) 484-9220  
Facsimile: (202) 863-1763  
Project Director: Julie Sykes

**Expanding Access to Medicaid  
Data for Research: Feasibility of  
a MAX Sample**

Final Report

December 30, 2010

Audra Wenzlow  
Ellen Bouchery  
John Czajka

**MATHEMATICA**  
Policy Research, Inc.

This report was prepared by Mathematica Policy Research under contract to the Office of Research, Development, and Information (ORDI) within the Centers for Medicare & Medicaid Services (CMS). The findings and conclusions presented in this report are those of the authors and do not necessarily represent the views of ORDI or CMS.

**ACKNOWLEDGMENTS**

The authors gratefully acknowledge the contributions of the project's Technical Expert Panel and CMS representatives in guiding this work. We also wish to express our thanks to Gerri Barosso at ResDAC for advising us on the format of Medicaid and Medicare files and the potential audience for Mini-MAX. In addition, we are grateful to Shinu Verghese for providing expert statistical and programming assistance and for drafting the glossary of sampling terms. Marilyn Ellwood, Frank Potter, and Bob Schmitz provided helpful comments on earlier drafts of this report, Susan Moore provided excellent secretarial support, and Jim and Molly Cameron provided editorial assistance. Finally, we wish to thank Susan Radke and Dave Baugh at the Office of Research, Development, and Information at CMS for their commitment to the development and continual improvement of MAX for Medicaid research.

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

## CONTENTS

EXECUTIVE SUMMARY .....		ix
I	INTRODUCTION.....	1
	A. Challenges to Increasing Medicaid Research via a Sample File .....	2
	1. Ensuring Sufficient Sample Size to Support a Wide Range of Medicaid Research.....	2
	2. Minimizing the Volume and Complexity of Mini-MAX Research Files .....	5
	3. Minimizing Potential for User Error .....	8
	4. Ensuring the Security of Identifiable Data.....	9
	B. Overview of Feasibility Study Components .....	10
	C. Roadmap to the Report.....	11
II	POTENTIAL PRECISION OF SAMPLE ESTIMATES .....	13
	A. Overview.....	13
	B. Simple Random Sample .....	14
	C. Stratified Sample with Differential Selection by State .....	27
	D. Expansion of the Sample .....	31
III	OPTIONS FOR REDUCING FILE SIZE WITHOUT SIGNIFICANT LOSS OF RESEARCH CAPABILITIES .....	37
	A. Options for Reducing File Size by Excluding MAX Variables .....	37
	B. Options for Reducing File Size by Separating Files into Two or More Components .....	40
IV	TEP RECOMMENDATIONS .....	43
	A. Potential Users and Uses.....	44
	B. Sampling Approach.....	46
	C. File Structure .....	48
	D. Documentation.....	50
	E. Summary .....	51

V CONCLUSIONS.....55

    A. Options for Mini-MAX.....55

    B. Mini-MAX Design Recommendations.....56

    C. Next Steps .....58

REFERENCES .....61

GLOSSARY OF SAMPLING TERMS.....63

APPENDIX A: MINI-MAX TEP MEETING PARTICIPANTS.....67

APPENDIX B: MAX 2007 RECORD LAYOUT AND POTENTIAL VARIABLES TO BE  
EXCLUDED FROM MINI MAX.....71

APPENDIX C: TEP SURVEY INSTRUMENT AND SUMMARY OF RESULTS.....85

## TABLES

I.1	Approximate Size of Existing Medicare and Medicaid Research Files and a Potential Mini-MAX Five Percent Sample.....	7
I.2	Approximate Size of Selected Sample Files.....	7
II.1	Number of Medicaid Enrollees by State and Enrollee Subgroup, 2006 .....	15
II.2	Expected Sample Counts for a Five Percent Simple Random Sample of Medicaid Enrollees by State and Enrollee Subgroup, 2006 .....	16
II.3	Percentage of All Records with No Expenditures by State and Enrollee Subgroup, 2006 .....	18
II.4	Mean Expenditures Among Records with Nonzero Expenditures by State and Enrollee Subgroup, 2006 .....	20
II.5	Mean Inpatient Expenditures Among Records with Nonzero FFS Expenditures by State and Enrollee Subgroup, 2006 .....	21
II.6	CVs of Percentage of Records with No Expenditures by State and Enrollee Subgroup, 2006 for a Five Percent Random Sample .....	23
II.7	CVs of Mean Expenditures Among Records with Nonzero Expenditures by State and Enrollee Subgroup, 2006, for a Five Percent Random Sample.....	25
II.8	CVs of Mean Inpatient Expenditures Among Records with Nonzero FFS Expenditures by State and Enrollee Subgroup, 2006, for a Five Percent Random Sample .....	26
II.9	State Sampling Rates for Alternative Medicaid Sample Design.....	29
II.10	CVs of Percentage with No Expenditures Among all Medicaid Enrollees by State and Enrollee Subgroup, 2006, with Differential Sampling by State .....	32
II.11	CVs of Mean Expenditures Among Records with Nonzero Expenditures by State and Enrollee Subgroup, 2006, with Differential Sampling by State .....	33
II.12	CVs of Mean Inpatient Expenditures Among Records with Nonzero FFS Expenditures by State and Enrollee Subgroup, 2006, with Differential Sampling by State.....	34
III.1	Estimated File Size of a Five Percent Sample of MAX and Levels 1 and 2 of MAX Skinny Files.....	39

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

## **EXECUTIVE SUMMARY**

As the primary source of person-level data on the Medicaid program, the Medicaid Analytic Extract (MAX) is a critical Medicaid research tool that is being used widely by the research community. However, the size and complexity of the MAX files have limited their use for some members of the research community, especially those with limited computing capacity. To facilitate wider use of MAX, the Centers for Medicare & Medicaid Services (CMS) is considering developing a sample of MAX data, referred to as Mini-MAX. A Medicaid sample has the potential to make it easier for current users to design and conduct MAX studies and could make MAX data accessible to a broader community of Medicaid researchers, policymakers, program administrators, and advocacy groups—including those without mainframe access or who are otherwise unable to process MAX due to its size or complexity.

Although the usefulness of a Medicaid sample has long been evident, challenges to developing one that meets general research needs have to date precluded its creation. Specifically, the heterogeneity across states of enrolled populations and optional eligibility and service coverage, the sheer volume of data in MAX, the uneven quality of the data, and the restrictions on user access make it challenging to construct a sample that can be made available to a broader audience without compromising the security of identifiable data and the quality of Medicaid research. Another challenge is educating new data users about the complexities of the data.

To determine whether it is possible to address these challenges in a sample file, CMS contracted with Mathematica Policy Research to assess the feasibility of developing a broadly useful Medicaid sample based on MAX that could be available to users as a standard MAX product. It is important to note that a MAX sample, if developed, will not be able to satisfy all research needs and interests. Some research activities will be better served by creating

customized data extracts from the entire MAX database (for example, research for small populations such as foster children or analyses that focus on specific illnesses or medical conditions). For this study, CMS requested that we focus on assessing the feasibility of creating sample research identifiable files (RIFs) rather than public use or less restrictive files. Currently, MAX files are available to users only as RIFs. In this report, we describe the findings of our feasibility study and resulting recommendations for Mini-MAX.

### **Precision of Sample Estimates**

One of the most critical factors in determining the utility of a Medicaid sample is whether researchers can use it to estimate Medicaid outcomes with sufficient precision. Accordingly, in the first component of our study, we evaluated whether a sample could produce sufficiently precise estimates for the Medicaid subpopulations most commonly studied by Medicaid researchers. At a minimum, variation in state Medicaid programs and reporting implies that precise estimates for Medicaid populations within each state would be required.

We examined three measures: the percentage of enrollees with no expenditures in 2006, mean expenditures for enrollees with nonzero expenditures, and mean inpatient expenditures for enrollees with nonzero fee-for-service (FFS) expenditures. Using 2006 MAX Person Summary (PS) data, we calculated the precision of these measures by state and for select subpopulations under two alternative sampling assumptions: (1) simple random sampling and (2) stratified sampling by state with sampling rates ranging from 3.5 percent for the state with the most Medicaid enrollees to 25 percent for the two states with fewest enrollees. The subpopulations we examined were the four major eligibility groups of children, adults, aged, and disabled, plus two additional subgroups—foster children, who accounted for nearly a million enrollees nationally, and infants, who accounted for almost 2.4 million in 2006. We expressed precision as a

coefficient of variation (CV), which shows the standard error of an estimate as a percentage of the estimate and thus allowed us to compare precision across outcomes.

Based on our analysis, we concluded that a five percent sample could support state-level estimates of use and expenditures for larger Medicaid groups as long as smaller states are sampled at higher rates than larger states. Oversampling of the aged and disabled could improve precision for the most costly subpopulations. In contrast, neither of the tested sampling approaches would be useful for studies of smaller subpopulations, such as foster children or infants. Due to their small size, and because the distribution of these small groups often varied substantially from state to state, analyzing them would likely require full MAX data.

### **Potential Size and Complexity of Sample Files**

In the second component of our feasibility study, we focused on developing options for reducing the size and complexity of Mini-MAX without a significant loss in research capabilities. We aimed to identify approaches for ensuring file sizes were no more than 3 to 6 GB, the typical maximum size for individual files to be processed easily on desktop computers. Furthermore, we attempted to identify solutions that would require analysis of only one file per file type compared to at least 51 (one for each state) needed with MAX.

We found that a five percent sample that excluded less commonly used, duplicated, and unreliable measures would meet our goals for file size and structure as long as the MAX Other (OT) claims file was split into multiple files by type. If all MAX variables were retained in Mini-MAX, a five percent sample of the MAX PS, OT claims, and Prescription Drug (RX) claims files would still be too large (9 GB, 23 GB, and 7 GB, respectively) to process on standard desktop computers. However, if less commonly used, duplicated, and unreliable measures were excluded, all file types except OT claims (19 GB) would be under the 6 GB threshold. Excluding identifier variables would further reduce the OT file to 13 GB, still greater

than the desired size. However, breaking the OT file into four claims types, to be determined after sampling, would satisfy the 6 GB limit. Possible claims types include capitated claims, crossover claims, physician and other ambulatory services (physician, other practitioner, outpatient hospital, clinic, and dental services), and all other services. All of these categories would be about 5 GB in size in a five percent MAX 2007 random sample.

A 10 percent sample would enable a wider range of analyses. However, it would contain files too large to be analyzed on desktop computers or would require multiple files per file type or files that exclude important research variables.

### **Recommendations for the Mini-MAX Design**

Given these findings, the project's Technical Expert Panel (TEP) and staff recommended producing a five percent Medicaid RIF sample to increase the volume of research conducted with MAX data, as long as the sample has the following features:

- A single design (for example, only one five percent sample) to minimize any confusion about data products or their associated documentation and ensure that research is replicable
- Timely release of the files soon after the release of Beta-MAX or MAX
- Stratification by state and major eligibility group (child, adult, aged, disabled) to ensure adequate sample sizes for state-level estimates
- A simple file structure that allows users to process one file per MAX file type (with the very large OT file split into three or more files by detailed claim type)
- A user's guide that includes reporting anomalies by state, links to sources of information on the differences in Medicaid programs across states, a guide for applying file weights and calculating standard errors, sample programs, and sufficient information for prospective users to determine whether the sample will be sufficient to support their research.

In addition, the TEP recommended that CMS consider other options for increasing access to MAX and suggested that CMS develop a plan for soliciting input about the needs of potential Mini-MAX users that could inform the design of its future editions. CMS should also consider

how the privacy board requirements for the RIF file may affect the potential of using Mini-MAX for exploratory and study-design purposes.

## **Conclusions**

Overall, we found that a stratified five percent sample, with the removal of less commonly used, duplicated, and unreliable variables, would substantially reduce the computing resources required to analyze MAX data while still allowing users to address many common research questions with sufficient precision. Since one goal of producing Mini-MAX is to expand the community of MAX users, such a sample should only be produced if it will be accompanied by user-friendly documentation that educates new users about the complexities of the Medicaid program, MAX data, and the Mini-MAX sample. Before Mini-MAX could be produced, CMS would need a plan for developing such documentation and preparing specifications for Mini-MAX (including sampling rates; source year and data; a sample selection algorithm; and the design of a longitudinal component, if applicable).

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

## I. INTRODUCTION

The U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services (CMS), contracted with Mathematica Policy Research to conduct a study assessing the feasibility of developing a Medicaid sample based on the Medicaid Analytic Extract (MAX). We call this sample Mini-MAX. Mini-MAX—a new standard product in the MAX product line—is being considered by CMS as a method to facilitate the use of MAX and thereby to increase the volume of high-quality research conducted with it. Currently, individuals in the research community without access to sufficient computer resources to process MAX data have no opportunity to use person-level Medicaid administrative files. Even for those researchers with sufficient computing capacity, there are substantial costs to acquiring and processing the full MAX files. A sample file would make it easier for current users to design and conduct studies and would make MAX data accessible to a broader community of researchers and analysts. Mini-MAX would allow researchers to conduct Medicaid comparative effectiveness research (CER) and other types of analyses on a much smaller database, at far lower cost. It might also improve the ability of researchers and policymakers to address ad hoc questions and identify issues without needing to expend a high level of resources on data processing.

Although the usefulness of a Medicaid sample long has been evident, the challenge of developing a sample that meets general research needs has to date precluded creating one. Specifically, the complexity of the state-administered Medicaid program, the sheer volume of data in MAX, its quality, and its restricted access make it difficult to construct a sample that improves its access to a broader audience without compromising the security of identifiable data and the quality of Medicaid research. As a result, the question addressed in this study is not whether it is possible to design a Medicaid sample, but to determine whether a design can be

identified that extends the audience of MAX while ensuring the integrity of the data and minimizing its inappropriate uses.

This feasibility study began in July 2010 and included an assessment of various Medicaid sample options; a Technical Expert Panel (TEP) advised the project (see Appendix A for list of participants). This report is the culmination of the study. In the remainder of this chapter, we summarize the challenges to designing a Medicaid sample and summarize our approach to assessing the feasibility of Mini-MAX.

## **A. Challenges to Increasing Medicaid Research via A Sample File**

The objective of creating a Mini-MAX sample is to increase the volume of high-quality Medicaid research. There are four key obstacles to developing a broadly useful Medicaid sample: (1) ensuring that sample sizes are sufficiently large to support a variety of Medicaid research studies, (2) reducing costs and other barriers to MAX by minimizing the size and complexity of Mini-MAX files, (3) identifying and addressing the potential for user error, and (4) ensuring the security of identifiable data.

### **1. Ensuring Sufficient Sample Size to Support a Wide Range of Medicaid Research**

Unlike the Medicare program, which has national coverage rules and centralized administration, Medicaid is a state-administered program that functions differently in each state and for different populations. The federal government has established broad guidelines for eligibility options, service coverage, payment approaches, and program administration and reporting. However, within these guidelines, states have substantial flexibility. The resulting differences in program features highlight the need to conduct Medicaid analysis at the state and often the subgroup level and, as a result, require that Mini-MAX have sufficient sample sizes for each state and distinct populations to support state- and subgroup-level analyses. Because a

sample cannot support all Medicaid research, identifying the priority research areas for Mini-MAX and ensuring that it can support them is one of the key challenges to developing its design.

**Cross-State Differences.** To receive federal Medicaid funds, all states must cover certain mandatory eligibility groups. These include low-income children; parents and caretaker relatives in low-income families with children; pregnant women; infants born to Medicaid-eligible pregnant women; Supplemental Security Income (SSI) recipients, with the exception of individuals living in 209(b) states;<sup>1</sup> low-income Medicare beneficiaries; and select other groups. Beyond the mandatory eligibility groups, optional eligibility groups include children, pregnant women, individuals with disabilities, and aged persons who have higher income levels or are medically needy (have sufficiently high medical costs to bring their net income below a state-determined level). Institutionalized individuals with somewhat higher income or those who would be eligible for institutional care and receive home care services also can be optionally covered by a state.

To receive federal Medicaid funds, state programs also must cover a mandatory set of services for certain subgroups. These include physician care; hospital inpatient and outpatient care; laboratory and x-ray services; early and periodic screening, diagnostic, and treatment (EPSDT) for individuals under 21; federally qualified health centers and rural health clinics; and home health for persons eligible for nursing facility care. In addition, states can cover an optional set of services, including prescription drugs, dental, personal care, hospice services, and home and community-based care services through waivers. States also can define the scope of particular services they cover and eligible subgroups. For example, they can limit the number of

---

<sup>1</sup> Section 209(b) states are states that elected to use more restrictive eligibility requirements than those of the SSI program, but these requirements cannot be more restrictive than those in place in the state's Medicaid plan as of January 1, 1972.

services or prescription drugs and can vary in how they define medical necessity. As a result, most Medicaid research aimed at understanding service utilization, Medicaid spending, or service access must analyze each state's data independently.<sup>2</sup>

**Subgroup Differences.** The Medicaid program serves numerous distinct subpopulations. Because they differ in their demographic and socioeconomic characteristics and health care needs, research often is targeted to specific subgroups. Thus, sufficient sample size for any given subpopulation of research interest is important to the utility of Medicaid data. Medicaid enrollees often are studied by eligibility criteria that indicate the population subgroup—for example, child, adult, aged (65+), or disabled—or the financial eligibility criteria by which a person became eligible. Other important Medicaid subgroups identifiable in MAX include Children's Health Insurance Program enrollees (M-CHIP), foster care and adoptive children, racial and ethnic groups, people dually enrolled in the Medicare and Medicaid programs, and people enrolled in Section 1915(c) waivers to receive home-based long-term care. Some of these subgroups are of keen interest to policymakers and researchers because they are particularly vulnerable and high cost.

It is also important to keep in mind that some Medicaid populations are eligible for a restricted set of Medicaid services. For example, more than 30 percent of California's enrollees are eligible only for family-planning services. Other important restricted-benefit groups identifiable in MAX include dual Medicare and Medicaid enrollees eligible only for Medicare cost-sharing and aliens eligible only for emergency services. MAX also identifies persons enrolled in non-Medicaid programs that will cover some services typically paid for by Medicaid, including dual Medicare and Medicaid enrollees whose acute care services are covered by

---

<sup>2</sup> For further information on the Medicaid program, see Perez et al. (2008) or Schneider et al. (2002).

Medicare and Money Follows the Person (MFP) enrollees whose home-based long-term care is covered by MFP grant funds. Researchers often may need to examine these groups separately from those eligible for full Medicaid services to differentiate differences due to use and spending from those associated with Medicaid policies and access.

Finally, beneficiaries vary substantially in how they receive services. The majority of child and adult Medicaid enrollees receive services through managed care arrangements. These arrangements vary from full capitation, in which the managed care organization assumes full risk for providing a specified set of services when medically necessary, to primary care case management models, in which primary care providers are paid a small monthly fee to provide basic services only and coordinate more complex service needs. In many states, more than half of the Medicaid population is enrolled in a health maintenance organization (HMO) or health insuring organization (HIO). According to the 2006 MAX PS file, however, there was no HMO or HIO enrollment in 14 states in 2006. Because service provision and, as described further below, reported data, differ substantially for managed and fee-for-service care recipients, the two groups typically are studied separately.

Given these numerous subpopulations and their varying characteristics, the design of Mini-MAX must assess which subgroups will be supported by the sample and ensure that Mini-MAX users can easily identify them.

## **2. Minimizing the Volume and Complexity of Mini-MAX Research Files**

The current size of the full MAX research files is prohibitively large for many potential users in the research community. In addition, MAX currently is composed of 5 files for each state—person summary (PS), inpatient claims (IP), institutional long-term care claims (LT), prescription drug claims (RX), and all other claims (OT) files. This means that researchers addressing a specific question at the national level and needing to use all four claims types may

need to process 255 separate files. To reach a wider audience of potential researchers, Mini-MAX must be designed to offer a substantial reduction in file size and a streamlined file structure relative to MAX.

Table I.1 displays the size of MAX component files, along with those of Medicare claims files distributed throughout the Chronic Condition Data Warehouse (CCW), illustrating some of the challenges to reducing the size and simplifying the file structure of MAX in Mini-MAX. A five percent sample of the Medicare files is produced annually to increase accessibility of Medicare data. The five percent Medicare beneficiary summary file is substantially smaller (250 MB) than a five percent sample of the MAX PS file (8,891 MB), primarily due to the record length within the Medicare beneficiary summary file (80) relative to that of the MAX PS file (2,895). In contrast, the five percent sample of the Medicare claims file is substantially larger (87,040 MB) than a potential five percent sample of MAX claims files (30,878 MB). However, Medicare claims files are organized into seven different file types, each of relatively small size, compared to only four Medicaid claims files, one of which (the OT file) is three times the size of the others combined.

Table I.2 displays the file size of some commonly used research files, including the Healthcare Cost and Utilization Project-National Inpatient Sample (HCUP-NIS), the American Community Survey, and the Current Population Survey; these also are substantially smaller than a MAX five percent sample.

In general, the maximum size of individual files needed for processing Mini-MAX files on desktop computers is 3 to 6 GB. To meet this goal, the design of the sample must implement creative solutions for minimizing the size and complexity of Mini-MAX.

**Table I.1. Approximate Size of Existing Medicare and Medicaid Research Files and a Potential Mini-MAX Five Percent Sample**

File	Approximate Number of Records		Approximate File Size (MB)	
	Person/Beneficiary Summary File	All Claims Files	Person/Beneficiary Summary File	All Claims Files
MAX 2006	61,661,641	2,192,702,079	177,821 (record length=2,895)	617,554 (variable record length)
CCW Medicare 100%	46,500,000	4,655,200,000	5,000 (record length=80)	1,740,800 (record length=439)
CCW Medicare 5%	2,329,457	232,760,000	250 (record length=80)	87,040 (record length=439)
Potential Mini-MAX 5% MAX sample	3,083,082	109,635,104	8,891 (record length=2,895)	30,878 (variable record length)

CCW = Chronic Condition Data Warehouse.

**Table I.2. Approximate Size of Selected Sample Files**

File	Number of Records	Approximate Record Length	Approximate File Size (MB)
Healthcare Cost and Utilization Project–National Inpatient Sample (HCUP-NIS)	8,158,381	516	13,312
American Community Survey (ACS)	4,338,000	NA	2,358
Current Population Survey (CPS)	392,550	1,000	372

NA = not available.

### 3. Minimizing Potential for User Error

As Mini-MAX begins enabling a wider audience to access person-level Medicaid files, it will become more critical to ensure availability and easy accessibility of documentation directing users in appropriate uses of the files. Lack of knowledge about state Medicaid policy contexts, MAX data reporting problems, and appropriate sample weighting techniques all can result in user errors and subsequent incorrectly interpreted results and misguided policy conclusions.

In addition to the aforementioned variations in Medicaid policies, the quality, timeliness, and completeness of data reported in MAX also vary substantially across states and for various populations and outcomes. MAX is based on Medicaid Statistical Information System (MSIS) data, which have been reported by states since 1999. Because the quality and completeness of MSIS data vary by state and year, researchers have been cautious when using MAX to ensure that they understand any data anomalies and Medicaid program differences and their implications for research. Significantly, a substantial portion of managed care encounters are not reported to MSIS. Even among those states that report encounter data, the MAX team does not recommend using them for research because of inconsistencies and potential incompleteness. In addition, cost settlements, gross adjustments, and bundled services not reported at the person level are excluded from MAX. Other variations and reporting problems unique to states are also issues. For example, one state may be unable to identify enrollees of a specific waiver, whereas claims for all behavioral health services may be missing in another. The MAX team provides anomaly tables to highlight these problems for researchers. However, researchers new to Medicaid research may require additional information or need to receive this information in a more accessible format. Furthermore, for Mini-MAX, documentation would need to be enhanced to direct users in the appropriate use of weights associated with a mini-MAX sample.

#### 4. Ensuring the Security of Identifiable Data

Privacy requirements are intended to limit access to MAX data but not preclude appropriate use of the data for research. However, because of the time lag required for data request approval and processing, privacy restrictions are an obstacle to using the data for certain purposes. While Mini-MAX is intended to facilitate research and be less costly to obtain and process than full MAX, at this time, privacy requirements for Mini-MAX are expected to be the same as those for the full files.

MAX data currently are available to users only as Research Identifiable Files (RIFs). That is, individual identifiers are included that could permit the identity of a beneficiary or physician (e.g., date of birth, age, race, sex, residence information). The Privacy Act, Freedom of Information Act, and other federal government rules and regulations restrict the use of these data to analyses compatible with the purpose(s) for which the data are collected. Access to the data requires submission of a data use request to CMS's Privacy Board. Currently, the Privacy Board meets only once a month and may not be able to review all submitted requests for that month, so a review may take two months or more. Once a request has been approved, there is additional lag time for the file to be prepared and sent to the requestor. In all, obtaining MAX in a RIF format takes about four to six months from the time a request is first initiated.<sup>3</sup>

Given the delay in obtaining data from the privacy review process, addressing ad hoc questions within a limited timeframe using MAX is not feasible. Similarly, since a privacy board review requires submission of a project proposal, including a detailed study design, researchers generally would not be able to use Mini-MAX to support study design.

---

<sup>3</sup> Currently, requests for use of MAX data for studies not funded by CMS are submitted to the Research Data Assistance Center (ResDAC). For additional information on accessing RIF files from CMS, see <http://www.resdac.umn.edu/Medicaid> or [http://www.cms.hhs.gov/PrivProtectedData/02\\_Criteria.asp](http://www.cms.hhs.gov/PrivProtectedData/02_Criteria.asp).

For the purpose of this feasibility study, CMS requested that we focus on assessing the feasibility of developing a sample RIF file only. For this reason, we generally do not discuss design options for limited data sets (LDS), non-identifiable data sets, or public use file (PUF) options in this report. CMS may decide to pursue these options in the future, if and when a five percent RIF file becomes available.

## **B. Overview of Feasibility Study Components**

The above discussion highlights some of the challenges inherent in increasing the volume of MAX research through creation of a MAX sample. Given the complexity of the Medicaid program, it is essential to ensure that a sample can adequately represent differences across states and by eligibility subgroup (and potentially other smaller subgroups). Mini-MAX also must address the primary obstacle to the use of MAX—file size. In addition, users of any sample file will need to understand complexities of both the Medicaid program and its reporting, as well as any implemented sampling approach, to appropriately use and interpret findings based on Mini-MAX.

The feasibility study presented in this report includes three primary components designed to address these challenges. The first component includes an analysis of MAX data to determine the precision of state and subgroup estimates achievable in a five percent sample. For each state, the analysis develops sample size and coefficient of variation estimates by four larger eligibility groups—child, adult, aged, and disabled—plus two of the many small subpopulations likely to generate research interest: foster children, numbering under a million nationally, and infants, numbering close to 2.4 million. Using two alternative sample designs, we present results for three variables: the percentage of enrollees with no expenditures, mean expenditures for enrollees with nonzero expenditures, and mean inpatient expenditures for enrollees with nonzero inpatient expenditures. The first design is a uniform five percent sample, whereas the second is a

five percent sample with varying sampling rates by state. The analysis indicates the degree to which a standard five percent sample would be sufficient to appropriately estimate common outcomes of interest to Medicaid researchers nationally, by state, and for subgroups.

The second component of this feasibility study is a detailed analysis of the content of MAX files to determine the contributions of each variable and each file to the overall size of MAX, and an assessment of the potential for excluding variables or otherwise reducing the size of individual files in Mini-MAX. This analysis details options for and implications of excluding subsets of variables—for example, duplicated information, unreliable measures, and identifier variables—from each of the five MAX file types in Mini-MAX. We also summarize options for further reducing individual Mini-MAX files by separating the OT file by service type.

The third component of the study is a TEP, which consisted of Medicaid researchers and sampling experts who met three times over the course of the project. The TEP informed all components of this study. Most importantly, combined with input from CMS and project staff, the TEP made recommendations regarding the most critical components of Mini-MAX and how best to achieve them in a sample's design.

### **C. Roadmap to the Report**

In subsequent chapters of this report, we describe each study component and our conclusions regarding the feasibility of developing Mini-MAX. In the following three chapters, we present the findings on the extent to which a Mini-MAX sample can address the challenges posed in this Introduction. In Chapter II, we report on our ability to create a sample that can produce sufficiently precise estimates for those Medicaid subpopulations commonly studied by Medicaid researchers. In Chapter III, we discuss our ability to reduce file size without significant loss of research capabilities. We review in Chapter IV the input received from the TEP on the analyses presented in Chapters II and III, and discuss likely uses of the file and

methods for supporting high-quality research. Finally, in Chapter V, we present the project's conclusions and discuss next steps.

## II. POTENTIAL PRECISION OF SAMPLE ESTIMATES

In this chapter, we report findings from an analysis conducted to assess the feasibility of meeting the first of the four key challenges to creating a useful MAX sample. Specifically, we discuss estimates of the statistical precision that can be achieved nationally and for subgroups of the Medicaid population with a five percent sample of MAX data.

### A. Overview

One of the most critical factors that will determine the utility of a Medicaid sample is whether it can be used to estimate important Medicaid outcomes with sufficient precision to be useful for research. Working with the 2006 MAX PS file, which contains not only enrollment and eligibility data but summary expenditure data, we examined a wide range of variables, by state, for subpopulations defined by eligibility criteria and age. Our purpose was to determine the precision with which these variables could be estimated using a five percent MAX sample. We selected eight variables and six subpopulations to assess two alternative sample designs with respect to statistical precision for the nation, the 51 “states” (including the District of Columbia), and six subpopulations (for both the nation and the individual states). The subpopulations were the four eligibility groups of non-disabled children, non-disabled adults, aged, and disabled under 65, plus two additional subpopulations—foster children, numbering under a million nationally, and infants, numbering close to 2.4 million.

In this chapter, we present results for three variables:

- The percentage of enrollees with no expenditures during the 2006 calendar year
- Mean expenditures for enrollees with nonzero expenditures
- Mean inpatient expenditures for enrollees with nonzero fee-for-service (FFS) expenditures

The results are based on two alternative sample designs:

- A uniform 5 percent sample of enrollee records

- An alternative 5 percent sample with stratification by state, with state sampling rates varying from a low of 3.5 percent for the state with the largest Medicaid enrollment to a high of 25 percent for the two states with the smallest Medicaid enrollment

These two designs illustrate the strengths and limitations associated with alternative approaches to designing a Mini-MAX. In addition, we consider strategies for creating an expanded Mini-MAX that might include up to 20 percent of Medicaid enrollees in a given year.

Table II.1 provides population counts for all Medicaid enrollees and the six subpopulations by state.<sup>4</sup> Total state enrollment varied from 75,000 in North Dakota to 10.6 million in California, which indicates the challenge faced in developing a sample design that will support state as well as national estimates. For aged enrollees, who totaled 5.3 million nationally, the state counts varied from 5,500 in Wyoming to 765,000 in California. For foster children, who numbered less than a million nationally, the smallest count—579 in Massachusetts—was so far below the counts in the next smallest states as to suggest a problem with the data. North Dakota and Delaware had just over 2,000 enrollees identified as foster children, while California had 156,000, reflecting a proportionately smaller range than either total enrollees or aged enrollees but indicating a challenge for sampling owing to the very small numbers for nearly half of the states. In all, 22 states had fewer than 10,000 foster children, implying sample counts below 500 for a 5 percent sample.

## **B. Simple Random Sample**

The first sample design that we examined was a simple random sample of five percent of Medicaid enrollees. With a simple random sample, every enrollee has the same chance of being selected. Table II.2 shows expected sample counts from a five percent simple random sample of

---

<sup>4</sup> Records with no enrollment data or with only S-CHIP enrollment have been removed.

**Table II.1. Number of Medicaid Enrollees by State and Enrollee Subgroup, 2006**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	59,770,821	29,828,370	15,456,610	5,324,879	9,160,962	998,555	2,374,139
Alabama	967,777	439,256	212,748	104,417	211,356	8,472	35,618
Alaska	130,740	79,601	28,269	7,373	15,497	3,041	6,069
Arizona	1,464,622	672,404	578,239	75,084	138,895	13,494	62,498
Arkansas	758,906	426,732	150,856	64,391	116,927	6,840	25,881
California	10,611,919	4,280,264	4,441,835	764,574	1,125,246	155,692	333,863
Colorado	571,704	336,889	105,844	52,174	76,797	19,824	29,710
Connecticut	533,568	284,332	118,861	64,392	65,983	6,264	16,978
Delaware	184,542	79,128	70,561	13,226	21,627	2,076	6,507
District of Columbia	167,174	80,275	41,192	10,111	35,596	4,283	5,741
Florida	2,997,801	1,555,568	558,004	353,678	530,551	49,595	139,763
Georgia	1,758,788	1,031,574	300,880	137,078	289,256	34,272	100,955
Hawaii	231,642	109,235	74,268	22,312	25,827	7,104	7,319
Idaho	226,209	142,625	32,230	16,274	35,080	3,114	11,179
Illinois	2,404,599	1,367,081	498,078	197,732	341,708	69,102	91,461
Indiana	1,049,760	622,409	193,313	81,241	152,797	16,406	48,432
Iowa	467,347	230,481	124,663	41,813	70,390	11,538	19,353
Kansas	358,751	202,682	59,871	34,458	61,740	15,304	19,128
Kentucky	871,842	427,910	135,417	71,479	237,036	13,835	36,157
Louisiana	1,213,077	749,866	154,505	111,104	197,602	10,947	61,382
Maine	332,411	128,668	109,478	40,547	53,718	3,795	7,288
Maryland	855,745	478,217	175,650	62,751	139,127	17,978	36,211
Massachusetts	1,256,946	477,264	379,434	152,356	247,892	579	34,954
Michigan	1,938,399	1,025,731	475,968	134,837	301,863	41,050	68,874
Minnesota	778,126	390,325	182,294	92,908	112,599	10,463	31,451
Mississippi	775,999	388,035	136,057	83,517	168,390	5,035	38,936
Missouri	1,111,398	619,852	211,448	95,551	184,547	28,170	44,247
Montana	112,766	61,361	22,298	9,670	19,437	4,263	5,520
Nebraska	262,014	157,188	46,103	23,941	34,782	13,859	10,624
Nevada	254,747	143,193	50,108	23,023	38,423	9,909	16,083
New Hampshire	142,225	86,632	20,036	14,354	21,203	2,676	5,525
New Jersey	1,081,498	560,671	204,003	126,998	189,826	27,257	47,211
New Mexico	516,862	303,629	121,343	26,226	65,664	5,052	20,944
New York	5,092,937	2,029,293	1,884,735	438,055	740,854	57,761	142,250
North Carolina	1,673,629	884,020	318,005	182,851	288,753	20,651	76,914
North Dakota	74,877	37,873	16,221	9,928	10,855	2,027	3,338
Ohio	2,157,048	1,141,935	491,457	177,036	346,620	39,289	78,422
Oklahoma	761,068	466,826	120,578	65,012	108,652	13,596	36,823
Oregon	524,127	272,235	121,523	49,207	81,162	18,309	24,481
Pennsylvania	2,094,047	977,430	387,036	232,145	497,436	55,006	66,751
Rhode Island	221,050	99,916	56,122	20,946	44,066	6,298	6,572
South Carolina	949,561	482,432	232,758	77,059	157,312	12,733	40,027
South Dakota	130,651	80,365	21,150	10,337	18,799	5,308	6,183
Tennessee	1,479,366	726,110	291,520	115,228	346,508	18,041	48,296
Texas	4,151,664	2,650,688	538,278	426,251	536,447	52,951	262,355
Utah	300,166	167,009	82,278	14,277	36,602	8,549	20,940
Vermont	159,470	66,651	52,114	18,661	22,044	2,680	3,745
Virginia	907,839	507,165	140,787	99,814	160,073	16,021	41,364
Washington	1,191,471	622,822	311,977	85,955	170,717	20,255	38,751
West Virginia	392,745	187,208	58,372	36,846	110,319	7,311	11,786
Wisconsin	1,038,804	436,533	305,471	150,181	146,619	17,352	34,806
Wyoming	80,397	52,781	12,374	5,500	9,742	3,128	4,473

Source: MAX Person Summary files, 2006.

Note: Foster children and infants appear in multiple columns.

**Table II.2. Expected Sample Counts for a Five Percent Simple Random Sample of Medicaid Enrollees by State and Enrollee Subgroup, 2006**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	2,988,541	1,491,419	772,831	266,244	458,048	49,928	118,707
Alabama	48,389	21,963	10,637	5,221	10,568	424	1,781
Alaska	6,537	3,980	1,413	369	775	152	303
Arizona	73,231	33,620	28,912	3,754	6,945	675	3,125
Arkansas	37,945	21,337	7,543	3,220	5,846	342	1,294
California	530,596	214,013	222,092	38,229	56,262	7,785	16,693
Colorado	28,585	16,844	5,292	2,609	3,840	991	1,486
Connecticut	26,678	14,217	5,943	3,220	3,299	313	849
Delaware	9,227	3,956	3,528	661	1,081	104	325
District of Columbia	8,359	4,014	2,060	506	1,780	214	287
Florida	149,890	77,778	27,900	17,684	26,528	2,480	6,988
Georgia	87,939	51,579	15,044	6,854	14,463	1,714	5,048
Hawaii	11,582	5,462	3,713	1,116	1,291	355	366
Idaho	11,310	7,131	1,612	814	1,754	156	559
Illinois	120,230	68,354	24,904	9,887	17,085	3,455	4,573
Indiana	52,488	31,120	9,666	4,062	7,640	820	2,422
Iowa	23,367	11,524	6,233	2,091	3,520	577	968
Kansas	17,938	10,134	2,994	1,723	3,087	765	956
Kentucky	43,592	21,396	6,771	3,574	11,852	692	1,808
Louisiana	60,654	37,493	7,725	5,555	9,880	547	3,069
Maine	16,621	6,433	5,474	2,027	2,686	190	364
Maryland	42,787	23,911	8,783	3,138	6,956	899	1,811
Massachusetts	62,847	23,863	18,972	7,618	12,395	29	1,748
Michigan	96,920	51,287	23,798	6,742	15,093	2,053	3,444
Minnesota	38,906	19,516	9,115	4,645	5,630	523	1,573
Mississippi	38,800	19,402	6,803	4,176	8,420	252	1,947
Missouri	55,570	30,993	10,572	4,778	9,227	1,409	2,212
Montana	5,638	3,068	1,115	484	972	213	276
Nebraska	13,101	7,859	2,305	1,197	1,739	693	531
Nevada	12,737	7,160	2,505	1,151	1,921	495	804
New Hampshire	7,111	4,332	1,002	718	1,060	134	276
New Jersey	54,075	28,034	10,200	6,350	9,491	1,363	2,361
New Mexico	25,843	15,181	6,067	1,311	3,283	253	1,047
New York	254,647	101,465	94,237	21,903	37,043	2,888	7,113
North Carolina	83,681	44,201	15,900	9,143	14,438	1,033	3,846
North Dakota	3,744	1,894	811	496	543	101	167
Ohio	107,852	57,097	24,573	8,852	17,331	1,964	3,921
Oklahoma	38,053	23,341	6,029	3,251	5,433	680	1,841
Oregon	26,206	13,612	6,076	2,460	4,058	915	1,224
Pennsylvania	104,702	48,872	19,352	11,607	24,872	2,750	3,338
Rhode Island	11,053	4,996	2,806	1,047	2,203	315	329
South Carolina	47,478	24,122	11,638	3,853	7,866	637	2,001
South Dakota	6,533	4,018	1,058	517	940	265	309
Tennessee	73,968	36,306	14,576	5,761	17,325	902	2,415
Texas	207,583	132,534	26,914	21,313	26,822	2,648	13,118
Utah	15,008	8,350	4,114	714	1,830	427	1,047
Vermont	7,974	3,333	2,606	933	1,102	134	187
Virginia	45,392	25,358	7,039	4,991	8,004	801	2,068
Washington	59,574	31,141	15,599	4,298	8,536	1,013	1,938
West Virginia	19,637	9,360	2,919	1,842	5,516	366	589
Wisconsin	51,940	21,827	15,274	7,509	7,331	868	1,740
Wyoming	4,020	2,639	619	275	487	156	224

Source: MAX Person Summary files, 2006.

Note: Foster children and infants appear in multiple columns.

enrollees by state and the same subgroups shown in Table II.1. Nationally, a five percent sample yielded just under 3 million records; by state, however, the counts varied from a low of 3,700 in North Dakota to a high of 531,000 in California. For aged beneficiaries, the smallest of the four eligibility subgroups, the sample counts ranged from 275 in Wyoming to 38,000 in California. For foster children—ignoring Massachusetts—the sample counts varied from a low of 101 in North Dakota to 7,800 in California, with 22 states expecting fewer than 500 sample records. The counts given in this table, we should note, reflect numbers prior to any reductions that will occur for researchers that are studying characteristics that apply only to a subset of the enrollees in each of the subgroups.

Tables II.3 through II.5 present means for the three Medicaid variables examined in this study. Table II.3 provides estimates of the percentage of enrollees with no expenditures during 2006, by state and enrollee subgroup. Nationally, 11.3 percent of persons ever enrolled in Medicaid during the year had no reported expenditures. For children, foster children, and infants, the fraction with no expenditures was below 7 percent but rose to nearly 9 percent for disabled beneficiaries, 19 percent for adults, and 22 percent for aged enrollees. We also found substantial variation by state—not just in the overall proportion of beneficiaries with no associated expenditures but also in how these proportions differed across enrollee subgroups. The variation makes this table not only interesting but particularly well suited for use in evaluating alternative sample designs. In Alabama, for instance, less than half a percent of children, foster children, and infants had no expenditures, but nearly 10 percent of the disabled, 38 percent of the aged, and 42 percent of adults had no expenditures. At the opposite extreme, South Dakota had no enrollees of any type without expenditures. Finally, we note that estimates for Maine are not shown because the state submitted very limited FFS expenditure data, so the estimates we obtained from the MAX data would not reflect its actual Medicaid expenditures.

**Table II.3. Percentage of All Records with No Expenditures by State and Enrollee Subgroup, 2006**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	11.26	6.06	19.12	22.00	8.65	6.80	5.38
Alabama	15.69	0.35	42.38	38.11	9.63	0.22	0.39
Alaska	13.79	14.99	15.37	9.06	7.02	10.75	3.36
Arizona	30.98	6.33	55.94	64.80	28.11	0.01	0.96
Arkansas	8.43	2.16	27.60	12.89	4.14	1.15	0.17
California	15.42	8.16	27.97	4.62	0.88	2.05	9.66
Colorado	2.78	0.36	4.04	11.62	5.60	0.02	0.65
Connecticut	5.28	2.41	3.31	18.92	7.91	1.63	2.30
Delaware	8.85	3.38	10.14	31.90	10.54	1.06	0.98
District of Columbia	7.10	3.40	4.52	19.90	14.76	19.38	4.42
Florida	9.85	3.88	8.00	34.85	12.64	7.36	3.87
Georgia	4.90	1.54	3.08	24.82	9.36	1.27	1.66
Hawaii	4.04	1.65	3.45	12.68	8.39	2.29	8.54
Idaho	5.82	3.58	6.88	18.48	8.06	2.79	3.48
Illinois	18.32	14.93	22.32	42.59	12.05	14.86	7.16
Indiana	8.24	5.43	10.05	20.63	10.78	5.93	2.33
Iowa	10.13	5.00	21.93	14.83	3.24	0.90	1.87
Kansas	7.90	5.57	8.33	16.07	10.56	9.29	3.16
Kentucky	4.70	1.84	1.21	22.85	6.38	5.88	13.55
Louisiana	10.68	7.72	9.42	28.47	12.92	12.60	36.34
Maryland	13.30	2.62	43.52	22.05	7.90	2.68	2.61
Massachusetts	10.74	7.00	15.65	18.50	5.66	6.74	5.79
Michigan	9.42	3.34	25.07	11.96	4.27	0.75	2.53
Minnesota	9.42	5.71	9.04	29.63	6.22	6.59	3.16
Mississippi	20.27	17.89	29.04	25.03	16.30	18.03	4.86
Missouri	6.26	6.44	10.85	4.59	1.25	6.59	2.41
Montana	3.77	1.93	2.80	13.83	5.71	0.80	1.74
Nebraska	7.76	6.50	11.20	11.88	6.01	7.63	2.16
Nevada	7.42	3.92	5.27	25.37	12.50	2.86	3.30
New Hampshire	13.59	12.65	15.88	18.00	12.26	5.04	4.31
New Jersey	8.63	4.60	7.07	22.96	12.64	13.27	14.83
New Mexico	6.86	2.05	19.75	10.06	4.06	9.30	4.70
New York	11.46	10.56	11.66	19.92	8.42	10.42	4.27
North Carolina	6.83	2.32	8.75	20.25	10.00	5.05	1.49
North Dakota	8.89	6.47	8.87	17.53	9.44	7.30	2.88
Ohio	8.00	6.16	8.33	16.66	9.17	24.02	2.89
Oklahoma	5.54	1.82	12.32	16.20	7.63	1.89	2.45
Oregon	8.15	5.40	10.81	17.28	7.82	2.28	1.82
Pennsylvania	4.32	1.52	2.40	18.36	4.74	1.83	2.01
Rhode Island	6.00	2.60	3.22	22.06	9.61	7.54	2.01
South Carolina	14.66	8.91	25.08	20.73	13.91	11.46	2.32
South Dakota	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tennessee	2.83	0.10	0.07	26.99	2.85	0.05	0.77
Texas	10.38	4.10	8.54	40.32	19.46	5.61	1.67
Utah	19.27	21.26	18.11	21.21	12.01	11.69	11.85
Vermont	10.89	9.42	17.10	5.59	5.14	5.41	5.95
Virginia	10.54	8.36	10.88	20.85	10.70	12.87	3.76
Washington	10.77	4.30	23.24	14.45	9.75	15.42	4.75
West Virginia	7.83	2.98	5.59	23.94	11.88	9.26	8.53
Wisconsin	8.35	5.94	10.31	13.87	5.80	12.41	7.23
Wyoming	17.37	17.27	18.16	22.47	14.00	21.71	12.52

Source: MAX Person Summary files, 2006.

Notes: Data for Maine were not available. Foster children and infants appear in multiple columns.

Table II.4 reports the mean expenditures for the year for enrollees with nonzero expenditures, including both FFS and managed care, and Table II.5 shows mean inpatient expenditures among enrollees with nonzero FFS expenditures.<sup>5</sup> Both tables again illustrate substantial variation in means by enrollee subgroup and state. Nationally, the mean nonzero expenditure was \$4,789 and ranged from \$1,831 for children to just above \$13,500 for both the disabled and the aged. Mean inpatient expenditures were \$819; these ranged from \$382 for children to \$2,502 for infants.

For each of these population means, we calculated the variability of the sample estimates in a 5 percent simple random sample. We summarized this variability in the form of a coefficient of variation (CV). A CV expresses the standard error of an estimate as a percentage of the estimate. For example, a CV of one percent on an estimated mean of \$1,000 implies a standard error of just \$10. Since the 95 percent confidence interval of an estimate is plus or minus approximately two standard deviations, a CV of one percent in this case implies a confidence interval of plus or minus two percent of the estimate, or plus or minus \$20. CVs are useful in comparing precision across estimates with different means, which is why we use them here. Furthermore, the absolute value of the CV is informative about the degree of precision. A CV below one percent is indicative of a high level of precision, whereas a CV above 10 percent begins to raise concern about the precision of an estimate; a CV of 20 percent or more (indicating a confidence interval of plus or minus 40 percent) indicates a low level of precision for most purposes.

The biggest drawback of the CV as a measure of precision is that it becomes arbitrarily large as the statistic being estimated approaches zero—as it does for some of the percentage and mean

---

<sup>5</sup> Expenditures by type of service are available in MAX only for FFS enrollees.

**Table II.4. Mean Expenditures Among Records with Nonzero Expenditures by State and Enrollee Subgroup, 2006**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	4,789	1,831	2,657	13,513	13,547	6,181	3,953
Alabama	3,733	1,836	1,932	10,717	6,871	8,969	2,514
Alaska	8,260	4,556	5,871	21,245	23,579	13,467	9,731
Arizona	3,249	2,204	3,722	4,597	8,279	4,165	3,771
Arkansas	4,023	1,768	1,524	12,401	10,663	7,887	5,389
California	3,081	1,353	1,193	7,968	11,394	4,074	1,752
Colorado	4,505	1,671	2,649	14,058	14,149	7,285	2,655
Connecticut	7,788	2,530	2,765	27,073	24,732	4,579	5,093
Delaware	5,724	2,399	4,221	19,457	17,396	8,311	4,450
District of Columbia	8,214	2,937	4,816	22,230	22,366	19,911	8,374
Florida	4,153	1,446	2,688	11,436	10,886	4,480	4,191
Georgia	3,553	1,602	3,218	9,408	9,180	7,150	4,232
Hawaii	4,199	1,758	2,917	11,564	13,104	3,015	5,045
Idaho	5,121	1,790	4,043	14,809	16,343	5,049	4,404
Illinois	4,486	1,719	2,521	10,631	15,404	4,979	5,041
Indiana	4,878	1,970	3,248	15,527	14,472	4,888	2,985
Iowa	5,656	1,765	2,687	15,645	17,183	5,273	4,165
Kansas	5,954	2,212	3,284	15,437	16,612	8,486	4,129
Kentucky	4,904	2,225	3,687	11,956	8,955	10,826	3,450
Louisiana	3,532	1,133	3,005	9,275	10,955	4,090	3,653
Maryland	6,833	2,369	5,473	19,076	19,431	8,307	3,813
Massachusetts	7,382	3,633	3,263	17,760	14,624	4,704	6,372
Michigan	3,332	1,203	2,350	11,159	8,635	2,550	2,853
Minnesota	7,677	2,545	3,067	20,390	24,928	9,326	3,944
Mississippi	4,580	1,852	3,150	11,581	8,618	6,974	3,896
Missouri	4,550	2,114	2,950	10,831	10,815	7,439	4,472
Montana	5,642	2,463	3,659	18,405	12,624	7,287	4,693
Nebraska	5,857	2,483	3,441	15,774	17,648	8,158	7,759
Nevada	4,237	1,751	2,156	10,763	14,012	6,980	3,733
New Hampshire	6,902	3,045	3,499	19,798	17,517	15,561	3,170
New Jersey	7,110	2,031	3,090	20,629	20,113	10,220	2,472
New Mexico	4,684	2,280	3,331	11,115	15,718	11,639	6,104
New York	8,575	2,391	4,051	26,823	26,781	7,267	6,964
North Carolina	5,199	2,041	3,578	11,202	14,134	9,676	3,792
North Dakota	7,417	2,156	2,911	21,723	21,236	7,943	4,015
Ohio	5,775	1,713	3,093	20,234	16,665	4,044	4,426
Oklahoma	4,103	1,888	2,839	10,415	12,123	7,647	4,344
Oregon	4,553	1,907	3,775	11,781	10,856	6,556	3,549
Pennsylvania	5,753	2,318	3,132	16,724	10,433	4,286	4,283
Rhode Island	7,681	2,878	2,994	22,176	19,867	17,413	6,464
South Carolina	3,776	1,818	2,369	8,807	9,672	8,534	4,229
South Dakota	4,621	2,020	3,124	11,840	13,455	8,215	4,980
Tennessee	3,851	1,612	2,967	9,935	7,919	3,958	3,444
Texas	3,761	1,709	2,784	10,911	12,740	7,839	4,314
Utah	4,167	1,874	2,477	11,733	14,424	7,726	3,888
Vermont	5,670	2,925	3,508	9,252	15,047	17,783	3,503
Virginia	5,111	2,051	3,351	11,100	13,296	10,193	4,169
Washington	4,071	1,606	2,633	12,765	11,692	3,436	3,427
West Virginia	5,270	2,082	2,543	13,992	10,259	10,276	2,764
Wisconsin	4,553	1,314	2,255	9,840	13,785	5,300	3,316
Wyoming	6,399	2,601	4,380	20,066	21,675	9,215	5,909

Source: MAX Person Summary files, 2006.

Notes: Data for Maine were not available. Foster children and infants appear in multiple columns.

**Table II.5. Mean Inpatient Expenditures Among Records with Nonzero FFS Expenditures by State and Enrollee Subgroup, 2006**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	819	382	807	497	2,123	650	2,502
Alabama	243	41	1,008	263	168	39	67
Alaska	1,158	846	1,462	309	2,514	859	6,666
Arizona	1,471	744	2,087	1,090	2,477	1,599	4,148
Arkansas	437	287	411	338	1,052	453	3,870
California	645	325	484	596	1,786	578	1,104
Colorado	691	344	1,315	439	1,616	378	1,298
Connecticut	1,035	734	648	457	2,070	1,009	3,798
Delaware	426	226	450	438	1,072	361	1,964
District of Columbia	4,867	2,522	2,893	1,576	7,389	1,801	9,542
Florida	1,119	523	1,416	536	2,619	1,116	2,705
Georgia	826	368	959	324	2,306	467	2,916
Hawaii	582	84	424	420	2,003	28	4,248
Idaho	791	411	1,652	376	1,685	522	3,099
Illinois	1,113	542	878	826	3,701	1,011	3,871
Indiana	513	173	245	207	1,798	328	2,104
Iowa	598	319	713	366	1,418	175	2,635
Kansas	951	409	1,131	514	2,508	615	2,560
Kentucky	737	319	996	172	1,434	1,195	2,403
Louisiana	693	262	1,213	361	2,025	865	2,416
Maryland	839	274	878	967	2,716	915	1,925
Massachusetts	684	533	583	316	1,211	82	5,002
Michigan	459	271	484	81	1,034	274	2,106
Minnesota	805	430	482	205	1,874	674	2,588
Mississippi	877	510	1,227	392	1,683	1,172	2,510
Missouri	817	531	498	257	1,804	1,263	2,508
Montana	760	424	724	283	2,124	444	3,145
Nebraska	768	438	857	389	2,342	838	3,930
Nevada	1,254	639	998	437	3,039	689	2,914
New Hampshire	446	272	570	343	1,110	491	1,858
New Jersey	1,154	637	982	632	1,962	1,090	2,210
New Mexico	634	447	544	338	1,337	1,170	3,693
New York	1,674	616	1,677	1,327	4,086	578	4,318
North Carolina	709	312	992	118	1,988	470	2,164
North Dakota	714	449	820	232	1,829	609	2,755
Ohio	1,005	371	513	492	2,926	372	3,048
Oklahoma	826	424	1,217	531	2,292	812	2,659
Oregon	530	452	527	107	906	159	2,286
Pennsylvania	618	361	816	374	951	286	2,525
Rhode Island	994	256	126	589	2,769	284	5,682
South Carolina	701	364	726	316	1,901	491	2,720
South Dakota	801	463	1,055	209	2,247	409	3,487
Tennessee	564	252	704	218	1,357	428	2,344
Texas	689	380	1,009	308	2,228	1,002	2,208
Utah	832	553	787	553	2,163	401	2,987
Vermont	382	239	440	138	881	1,004	2,143
Virginia	791	342	770	687	1,988	593	2,852
Washington	566	127	343	388	2,106	478	1,839
West Virginia	442	67	306	116	1,223	362	732
Wisconsin	492	225	252	142	1,625	287	2,962
Wyoming	895	634	1,172	340	2,208	775	4,298

Source: MAX Person Summary files, 2006.

Notes: Data for Maine were not available. Foster children and infants appear in multiple columns.

expenditure items for specific subpopulations in specific states. The fact that we are using population rather than sample estimates for the percentages and mean expenditures in our calculations reduces the likelihood of observing very small values purely by chance. Nevertheless, as noted above, South Dakota had zeroes across the board for the percentage of enrollees with no expenditures, which prevented the calculation of CVs for that measure and state. The impact of small, nonzero estimates of percentages and mean expenditures will be evident when we examine the CVs.

Turning now to our results, Table II.6 reports the CVs of the percentage of enrollees with no expenditures during the year based on a 5 percent simple random sample of Medicaid enrollees. For the nation as a whole, the CV of the estimated percentage of enrollees with no expenditures was only 0.16 percent. The CVs rose appreciably when we examine individual states. While the CV for California, 0.32 percent, was just twice the national CV, only five other states had CVs below one percent: Arizona, Florida, Illinois, New York, and Texas. Nine other states had CVs that fell between one and 1.25 percent, and a dozen more had CVs below 2 percent. At the upper end, 13 states had CVs in excess of 3 percent, led by Montana, with a CV of nearly 7 percent.

Generally, the CVs for the four larger eligibility subgroups ran two to three times the size of the CVs for each state's total Medicaid enrollees. The larger CVs tended to be those associated with very low means. This is true, for example, of the CVs for the estimates for children in Alabama, Colorado, and Tennessee. In each of these states, less than 0.5 percent of children had no Medicaid expenditures. In Tennessee, only 0.07 percent of adult beneficiaries had no expenditures during the year, and the CV for that estimate was 32 percent. On the whole, though, very few of the CVs for these four subpopulations exceeded 10 percent, the point at which imprecision would begin to become a concern to users.

**Table II.6. CVs of Percentage of Records with No Expenditures by State and Enrollee Subgroup, 2006, for a Five Percent Random Sample**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	0.16	0.32	0.23	0.36	0.48	1.66	1.22
Alabama	1.05	11.46	1.13	1.76	2.98	102.43	37.72
Alaska	3.09	3.78	6.24	16.49	13.07	23.37	30.80
Arizona	0.55	2.10	0.52	1.20	1.92	447.10	18.20
Arkansas	1.69	4.60	1.86	4.58	6.30	50.02	67.36
California	0.32	0.73	0.34	2.32	4.49	7.83	2.37
Colorado	3.50	12.75	6.70	5.40	6.62	258.21	32.00
Connecticut	2.59	5.34	7.02	3.65	5.94	43.93	22.38
Delaware	3.34	8.50	5.01	5.68	8.86	94.75	55.66
District of Columbia	3.96	8.41	10.12	8.92	5.70	13.94	27.44
Florida	0.78	1.79	2.03	1.03	1.61	7.12	5.96
Georgia	1.49	3.52	4.58	2.10	2.59	21.30	10.82
Hawaii	4.53	10.46	8.69	7.85	9.20	34.63	17.11
Idaho	3.78	6.14	9.16	7.36	8.07	47.23	22.28
Illinois	0.61	0.91	1.18	1.17	2.07	4.07	5.33
Indiana	1.46	2.37	3.04	3.08	3.29	13.91	13.16
Iowa	1.95	4.06	2.39	5.24	9.22	43.65	23.31
Kansas	2.55	4.09	6.06	5.50	5.24	11.30	17.90
Kentucky	2.16	4.99	10.97	3.07	3.52	15.21	5.94
Louisiana	1.17	1.79	3.53	2.13	2.61	11.26	2.39
Maryland	1.23	3.94	1.22	3.36	4.09	20.12	14.36
Massachusetts	1.15	2.36	1.69	2.40	3.67	69.10	9.65
Michigan	1.00	2.38	1.12	3.30	3.85	25.38	10.58
Minnesota	1.57	2.91	3.32	2.26	5.18	16.47	13.96
Mississippi	1.01	1.54	1.90	2.68	2.47	13.43	10.03
Missouri	1.64	2.17	2.79	6.59	9.27	10.03	13.53
Montana	6.73	12.88	17.64	11.35	13.03	76.42	45.24
Nebraska	3.01	4.28	5.86	7.87	9.48	13.21	29.24
Nevada	3.13	5.85	8.47	5.05	6.04	26.21	19.09
New Hampshire	2.99	3.99	7.27	7.96	8.22	37.48	28.37
New Jersey	1.40	2.72	3.59	2.30	2.70	6.92	4.93
New Mexico	2.29	5.61	2.59	8.26	8.49	19.63	13.91
New York	0.55	0.91	0.90	1.35	1.71	5.46	5.62
North Carolina	1.28	3.08	2.56	2.08	2.50	13.50	13.11
North Dakota	5.23	8.74	11.26	9.74	13.29	35.45	44.97
Ohio	1.03	1.63	2.12	2.38	2.39	4.01	9.26
Oklahoma	2.12	4.80	3.44	3.99	4.72	27.63	14.71
Oregon	2.07	3.59	3.68	4.41	5.39	21.63	21.01
Pennsylvania	1.46	3.64	4.59	1.96	2.84	13.96	12.07
Rhode Island	3.77	8.66	10.35	5.81	6.54	19.73	38.51
South Carolina	1.11	2.06	1.60	3.15	2.81	11.01	14.50
Tennessee	2.15	16.75	32.10	2.17	4.44	149.04	23.16
Texas	0.64	1.33	1.99	0.83	1.24	7.97	6.71
Utah	1.67	2.11	3.32	7.21	6.33	13.30	8.43
Vermont	3.20	5.37	4.31	13.45	12.94	36.12	29.06
Virginia	1.37	2.08	3.41	2.76	3.23	9.19	11.13
Washington	1.18	2.67	1.46	3.71	3.29	7.36	10.17
West Virginia	2.45	5.90	7.61	4.15	3.67	16.36	13.50
Wisconsin	1.45	2.69	2.39	2.88	4.71	9.02	8.59
Wyoming	3.44	4.26	8.53	11.20	11.23	15.21	17.66

Source: MAX Person Summary files, 2006.

Notes: Data for Maine and South Dakota were not available. Foster children and infants appear in multiple columns.

Foster children and infants are another story, however. The exceedingly large CV estimates for foster children in Alabama, Arizona, Colorado, and Tennessee—all over 100 percent and some as high as 447 percent—are all due to very small percentages with nonzero expenditures (as low as 0.01 percent in Arizona). Setting these aside, we still find that nearly all of the CVs for both foster children and infants exceeded 10 percent, and a third exceeded 20 percent (implying a 95 percent confidence interval of plus or minus 40 percent or more). Nevertheless, most of the estimates to which these CVs apply (as reported in Table II.3) were under 10 percent for foster children and under 5 percent for infants, so even a 20 percent CV may not be excessive when translated into the width of a confidence interval.

Table II.7 reports the CVs of mean expenditures among enrollees with nonzero expenditures. For all Medicaid enrollees, the CV was 0.21 percent, or about one-third higher than the CV of 0.16 percent reported in Table II.6. On the whole, however, the CVs for mean expenditures were lower than the CVs for the percentage of enrollees with no expenditures. While this was true for just over half of the 50 state CVs for all enrollees, it held true for about two-thirds of the state subgroup CVs. Consistent with this, there was only one subgroup in which the CV for mean expenditures across all states was lower than the CV for the percentage with no expenditures. For foster children, only five states had CVs as high as 20 percent, and for infants, only six states had CVs of this magnitude.

Table II.8 reports CVs for mean inpatient expenditures among enrollees with nonzero FFS expenditures. For this expenditure item, we see a substantial increase in CVs at all levels. The CV for the national estimate for all enrollees was 0.70 percent, and only one of the subgroup CVs was under one percent. For foster children, the national CV was 7.34 percent. For all Medicaid enrollees, there were no state CVs under one percent. The smallest CV, for California, was 1.63 percent, and only 19 other states had CVs under 5 percent, while 11 states had CVs

**Table II.7. CVs of Mean Expenditures Among Records with Nonzero Expenditures by State and Enrollee Subgroup, 2006, for a Five Percent Random Sample**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	0.21	0.32	0.27	0.35	0.39	1.33	1.13
Alabama	1.31	1.06	1.69	2.86	2.38	6.59	4.61
Alaska	3.56	5.49	5.20	8.57	6.57	19.13	13.08
Arizona	1.99	2.80	3.15	11.13	4.86	4.66	3.43
Arkansas	1.58	2.54	2.67	2.38	2.73	12.22	10.53
California	0.59	0.90	0.82	1.01	1.04	3.49	4.99
Colorado	1.71	2.58	2.59	2.92	2.97	5.70	9.70
Connecticut	1.76	2.56	1.41	2.21	3.34	19.65	6.35
Delaware	3.04	4.21	2.18	7.31	6.59	22.09	13.04
District of Columbia	3.07	5.54	4.49	7.42	4.47	9.50	18.34
Florida	0.86	1.48	1.05	1.42	1.50	6.11	4.53
Georgia	1.18	1.58	1.29	2.14	2.42	5.31	4.90
Hawaii	2.82	2.57	1.81	6.15	6.35	11.83	17.83
Idaho	2.84	4.77	4.63	6.41	4.24	21.11	18.58
Illinois	1.24	2.70	2.17	1.98	1.80	8.08	7.71
Indiana	1.41	1.43	0.83	2.12	2.83	9.61	6.04
Iowa	1.95	3.05	2.61	2.56	3.23	7.85	11.41
Kansas	2.02	3.23	3.10	2.66	3.42	7.14	10.35
Kentucky	1.32	1.90	1.62	2.89	2.22	6.84	9.16
Louisiana	1.52	2.59	2.13	2.55	2.50	33.36	10.50
Maryland	1.46	2.73	2.28	2.89	2.18	8.54	10.58
Massachusetts	1.14	1.01	1.15	1.83	2.28	51.38	4.50
Michigan	0.94	1.50	1.14	1.97	1.55	6.48	7.05
Minnesota	1.33	2.20	1.43	1.61	2.13	10.46	8.08
Mississippi	1.61	2.79	2.20	2.91	2.82	18.66	8.86
Missouri	1.09	1.69	1.13	1.85	2.00	5.77	4.26
Montana	3.64	6.77	5.76	4.92	6.98	15.93	23.39
Nebraska	2.91	6.76	5.37	3.67	4.61	9.66	11.82
Nevada	3.40	4.87	4.00	5.85	5.94	12.18	16.80
New Hampshire	3.09	5.36	5.78	4.16	5.52	12.18	24.91
New Jersey	1.27	2.32	1.25	1.79	2.06	8.21	9.79
New Mexico	1.79	2.74	1.64	4.19	3.08	19.23	13.65
New York	0.80	1.20	0.76	1.11	1.45	6.64	3.35
North Carolina	1.03	1.70	1.50	1.82	1.76	7.90	5.56
North Dakota	4.40	9.14	7.03	5.17	7.67	21.08	27.13
Ohio	0.90	1.32	0.98	1.29	1.56	6.43	5.13
Oklahoma	1.74	2.44	2.52	2.41	3.42	8.45	6.91
Oregon	1.49	2.00	1.53	3.05	2.95	3.87	5.47
Pennsylvania	0.71	0.89	0.88	1.42	1.14	4.80	3.90
Rhode Island	3.02	4.65	1.98	4.88	5.01	9.89	24.07
South Carolina	1.48	2.36	2.17	3.10	2.63	8.94	7.86
South Dakota	3.65	6.00	5.90	5.60	6.57	13.25	19.17
Tennessee	1.43	1.75	1.83	3.03	2.57	10.94	11.17
Texas	0.77	1.20	0.84	1.04	1.50	6.08	3.30
Utah	3.32	7.07	4.14	6.52	5.28	10.26	14.19
Vermont	3.04	6.16	4.08	6.08	5.44	15.88	37.00
Virginia	1.37	2.31	1.35	2.56	2.26	9.42	8.22
Washington	1.56	1.74	1.85	2.07	3.23	15.22	12.16
West Virginia	1.91	3.39	2.76	4.05	2.76	16.06	10.25
Wisconsin	1.36	2.36	1.33	1.85	2.50	8.46	11.07
Wyoming	4.95	7.71	6.64	8.12	8.55	19.83	26.43

Source: MAX Person Summary files, 2006.

Notes: Data for Maine were not available. Foster children and infants appear in multiple columns.

**Table II.8. CVs of Mean Inpatient Expenditures Among Records with Nonzero FFS Expenditures by State and Enrollee Subgroup, 2006, for a Five Percent Random Sample**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	0.70	1.49	0.84	2.33	1.15	7.34	1.94
Alabama	2.31	10.66	2.41	3.95	6.76	61.04	56.42
Alaska	9.15	12.61	10.21	37.00	24.13	68.95	14.34
Arizona	5.75	12.87	4.91	40.68	23.52	384.64	21.49
Arkansas	5.81	9.11	4.95	5.27	11.40	59.77	12.39
California	1.63	4.50	1.99	4.60	2.83	14.62	7.71
Colorado	5.14	8.82	4.43	23.74	11.88	40.51	15.51
Connecticut	6.58	14.91	12.54	16.61	9.11	66.34	20.09
Delaware	17.38	36.72	15.93	26.08	41.44	194.60	33.28
District of Columbia	8.19	28.56	18.53	36.93	9.34	46.09	34.95
Florida	2.19	4.33	2.07	6.12	3.83	18.64	6.35
Georgia	3.83	6.61	3.80	12.69	6.45	33.97	7.98
Hawaii	17.55	66.22	28.22	39.83	23.33	130.08	86.55
Idaho	8.55	17.23	7.39	24.56	17.29	123.37	22.52
Illinois	3.57	7.05	5.27	14.12	5.32	25.04	9.43
Indiana	5.76	15.77	7.05	18.66	7.00	55.25	18.80
Iowa	6.55	10.53	5.59	13.13	14.08	49.39	14.91
Kansas	8.11	13.49	8.79	19.88	13.51	33.39	18.43
Kentucky	4.78	11.22	4.38	27.17	7.13	36.10	14.72
Louisiana	4.14	11.14	4.02	8.80	6.17	145.13	14.08
Maryland	6.78	18.44	8.91	25.02	8.82	47.22	21.49
Massachusetts	3.25	5.06	5.54	10.73	5.85	363.87	5.46
Michigan	4.64	8.75	5.45	30.17	7.82	41.03	10.78
Minnesota	7.46	21.91	8.16	20.88	9.07	63.21	24.60
Mississippi	3.41	7.30	3.56	5.44	5.64	49.88	11.26
Missouri	3.89	7.88	5.18	18.00	5.29	18.16	9.29
Montana	13.13	24.77	10.46	20.81	22.04	65.05	30.23
Nebraska	10.02	15.33	17.26	17.85	18.42	44.46	20.61
Nevada	9.97	18.30	10.21	36.52	15.09	53.81	27.63
New Hampshire	12.14	26.35	13.34	18.70	17.13	142.20	35.02
New Jersey	4.80	14.57	5.44	10.71	6.75	39.26	16.98
New Mexico	18.10	37.17	14.68	32.66	29.57	371.19	44.10
New York	2.03	4.52	2.47	6.40	3.51	26.92	5.38
North Carolina	2.77	5.45	3.01	16.10	4.68	39.40	7.39
North Dakota	18.51	23.55	15.51	110.62	39.91	63.16	33.03
Ohio	3.79	8.57	5.79	11.52	5.02	30.20	11.86
Oklahoma	3.53	6.47	4.15	10.73	6.42	38.98	8.00
Oregon	6.84	10.50	8.78	34.53	14.06	46.06	12.34
Pennsylvania	4.17	9.55	5.63	7.24	6.96	40.45	14.71
Rhode Island	13.61	35.15	40.35	31.70	16.00	81.51	41.03
South Carolina	4.35	8.10	4.88	22.40	7.44	63.40	9.85
South Dakota	11.92	18.21	12.01	50.20	23.40	73.94	24.16
Tennessee	4.59	8.91	5.48	22.71	7.85	58.66	15.49
Texas	2.70	4.92	1.64	5.94	4.91	30.69	5.97
Utah	8.46	13.84	9.07	36.40	18.18	100.12	16.23
Vermont	12.07	30.58	12.10	24.04	22.84	102.31	50.07
Virginia	5.12	13.26	6.84	8.45	7.48	61.68	15.79
Washington	6.58	26.23	13.22	14.09	8.02	73.91	35.42
West Virginia	7.12	36.98	12.27	22.80	7.98	145.44	31.58
Wisconsin	7.64	18.75	10.50	13.38	10.64	40.38	21.68
Wyoming	14.33	20.14	13.54	30.25	33.57	70.75	31.65

Source: MAX Person Summary files, 2006.

Notes: Data for Maine were not available. Foster children and infants appear in multiple columns.

above 10 percent. Estimates for the adult subgroup show the greatest precision, with only 18 states having CVs in excess of 10 percent and only two of these over 20 percent. The CVs for the disabled exceeded 10 percent in fewer than half of the states, as well (22), but 10 of these were above 20 percent. At the opposite end, all 50 of the CVs for foster children exceeded 10 percent, and 47 exceeded 20 percent. For infants, 39 of the CVs exceeded 10 percent, and 22 exceeded 20 percent. For the aged, 40 CVs exceeded 10 percent, and almost half (24) exceeded 20 percent. Overall, the results for this expenditure item illustrate the problems in using a 5 percent sample to estimate highly variable outcomes for small subsets of the population.

### **C. Stratified Sample with Differential Selection by State**

To address the high variability of estimates for smaller states and determine how this would affect the precision of estimates for smaller subpopulations, we developed an alternative sample design stratified by state. The purpose of stratification is to control the composition of a sample rather than leave the composition to chance. Generally, this means ensuring that the representation of key subgroups in the sample is consistent with their representation in the population, but stratification may also be applied in order to increase the relative sizes of particular subgroups in the sample and thereby improve the precision with which their characteristics may be estimated.<sup>6</sup> If a sample is sufficiently large (and 5 percent of 60 million Medicaid enrollees is indeed a large sample), then stratification may be unnecessary to ensure that the sample's representation of key subgroups reflects the population. However, if there is a desire to improve the precision of estimates for small subgroups, then stratification coupled with

---

<sup>6</sup> A compensating reduction in the sample weights for members of these subgroups is required to maintain their proper representation in population estimates.

higher rates of selection for the subgroup members will be necessary to achieve this objective. The alternative sample design retains the five percent average selection probability of the first design, but selection probabilities vary by state, with enrollees from smaller states being assigned higher probabilities of selection than enrollees from larger states. This alternative design represents only one possible allocation of a five percent sample with differential selection by state.

For the smallest states, we increased the selection probability to 25 percent.<sup>7</sup> For California, we reduced the probability to 3.5 percent. For the remaining states, we assigned selection probabilities that declined incrementally from 25 percent down to 3.5 percent as the size of the state Medicaid population increased. Blocks of states with similar numbers of Medicaid enrollees were assigned the same sampling rate, with 5.0 percent remaining the most common selection probability. Subject to these constraints, rates were assigned by trial and error and adjusted, as necessary, to maintain an overall five percent selection rate. With this sampling scheme, the larger of two states always received a larger sample, even when their sampling rates were identical. The selection rates are shown in Table II.9, where the states are sorted by Medicaid enrollment, from smallest to largest.

With this alternative design, 12 states would retain the 5 percent selection rate from the first design. States with fewer than 850,000 Medicaid enrollees were assigned larger selection rates, ranging from 5.3 percent up to the ceiling of 25 percent. (The highest rate below 25 percent was 20 percent.) States with more than 1.3 million Medicaid enrollees were assigned selection rates

---

<sup>7</sup> A fourfold increase in sample size will reduce the standard error of an estimate, and thus its CV, by one-half. Increasing the sampling rate to 25 percent will reduce the standard errors for these states by slightly more than one-half.

**Table II.9. State Sampling Rates for Alternative Medicaid Sample Design**

State	Total Medicaid Enrollees	Sampling Rate	Sample Size
U.S. Total	59,770,821	0.050	2,988,528
North Dakota	74,877	0.250	18,719
Wyoming	80,397	0.250	20,099
Montana	112,766	0.200	22,553
South Dakota	130,651	0.185	24,170
Alaska	130,740	0.185	24,187
New Hampshire	142,225	0.180	25,601
Vermont	159,470	0.180	28,705
District of Columbia	167,174	0.180	30,091
Delaware	184,542	0.175	32,295
Rhode Island	221,050	0.153	33,821
Idaho	226,209	0.150	33,931
Hawaii	231,642	0.150	34,746
Nevada	254,747	0.140	35,665
Nebraska	262,014	0.140	36,682
Utah	300,166	0.130	39,022
Maine	332,411	0.118	39,224
Kansas	358,751	0.110	39,463
West Virginia	392,745	0.101	39,667
Iowa	467,347	0.085	39,724
New Mexico	516,862	0.077	39,798
Oregon	524,127	0.076	39,834
Connecticut	533,568	0.075	40,018
Colorado	571,704	0.070	40,019
Arkansas	758,906	0.053	40,222
Oklahoma	761,068	0.053	40,337
Mississippi	775,999	0.053	41,128
Minnesota	778,126	0.053	41,241
Maryland	855,745	0.050	42,787
Kentucky	871,842	0.050	43,592
Virginia	907,839	0.050	45,392
South Carolina	949,561	0.050	47,478
Alabama	967,777	0.050	48,389
Wisconsin	1,038,804	0.050	51,940
Indiana	1,049,760	0.050	52,488
New Jersey	1,081,498	0.050	54,075
Missouri	1,111,398	0.050	55,570
Washington	1,191,471	0.050	59,574
Louisiana	1,213,077	0.050	60,654
Massachusetts	1,256,946	0.050	62,847
Arizona	1,464,622	0.045	65,908
Tennessee	1,479,366	0.045	66,571
North Carolina	1,673,629	0.042	70,292
Georgia	1,758,788	0.041	72,110
Michigan	1,938,399	0.040	77,536
Pennsylvania	2,094,047	0.040	83,762
Ohio	2,157,048	0.040	86,282
Illinois	2,404,599	0.040	96,184
Florida	2,997,801	0.040	119,912
Texas	4,151,664	0.036	149,460
New York	5,092,937	0.036	183,346
California	10,611,919	0.035	371,417

Source: MAX Person Summary files, 2006.

below 5 percent, ranging from 4.5 percent to 3.5 percent. The two states that follow California in the size of their Medicaid enrollment were assigned selection rates of 3.6 percent. With these selection rates, the total sample size was within 13.5 percentage points of the sample size generated by a uniform 5 percent selection rate. However, under this alternative sample design, the smallest state sample (for North Dakota) was 18,719, compared to 3,744 with the uniform 5 percent design, and no other state samples were below 20,000 in size. With a uniform 5 percent sampling rate, 18 state samples were below 20,000 (see Table II.2).

Altogether, 27 states had higher sampling rates under the alternative design, compared to the first design, so their CVs would be reduced. Another 12 had identical sampling rates under the two designs, so their CVs would be unchanged. The remaining 12 states had lower sampling rates under the alternative design, so their CVs would be increased—but not appreciably. Furthermore, these 12 states tended to have the smallest CVs, so their estimates would continue to have comparatively high levels of precision despite the sample size reduction.<sup>8</sup>

The CVs for the three expenditure variables for this alternative sample design are reported in Tables II.10 through II.12. Three general outcomes are evident from a comparison of these new results with those reported in Tables II.6, II.7, and II.8. First, the CVs for the smallest states were greatly reduced. Second, the CVs for the national estimates were increased only marginally. Third, the CVs for the two small subpopulations, while reduced, remained unacceptably large in most states—especially for inpatient FFS expenditures.

The impact of the revised design is most evident for North Dakota, Wyoming, and Montana, which had the largest increases in sampling rates. For North Dakota, the CV for the percentage

---

<sup>8</sup> These 12 states have the most precise estimates for variables that use the full sample. For variables based on only part of the sample, the relevant sample sizes may be smaller for some of these states than for certain states with smaller total samples.

of Medicaid enrollees with no expenditures dropped from 5.23 percent (Table II.6) to 2.34 percent (Table II.10), and the CV for mean inpatient expenditures dropped from 18.51 percent (Table II.8) to 8.27 percent (Table II.12). The CV for North Dakota's mean inpatient expenditures for foster children dropped from 63.16 percent to 28.13 percent. Similar improvements were recorded for Wyoming and Montana, with several other small states showing appreciable reductions in all of their CVs as well. For estimates of the percentage of disabled persons with no expenditures, the number of CVs in excess of 10 percent decrease from 5 to 0. Similarly, for estimates of the percentage of infants with no expenditures, the number of CVs in excess of 20 percent declined from 17 to 8. For mean inpatient expenditures across all Medicaid enrollees, the number of state CVs in excess of 10 percent fell from 11 to only 2. The small impact of the alternative sample design on precision at the national level was evident for all enrollees and the six subgroups.

Despite the improved precision achieved for most states, the CVs for estimates for highly variable measures remained high for small subpopulations. These estimates were based on small samples in all states, so reallocating the sample toward smaller states would do little to address the weakness of the estimates. In particular, the CVs of state estimates of inpatient FFS expenditures for foster children were below 20 percent in California and Missouri only, and for 20 states, the CVs exceeded 50 percent. To reduce these CVs to a more acceptable range—for example, by a factor of four—would require a 16-fold increase in sample size—close to requiring the entire population. In the next section, we consider issues related to increasing the sample size of the Mini-MAX.

#### **D. Expansion of the Sample**

In discussions subsequent to the first TEP meeting, CMS asked us to consider how the sample might be expanded to 20 percent, and what implications this might have for construction

**Table II.10. CVs of Percentage with No Expenditures Among all Medicaid Enrollees by State and Enrollee Subgroup, 2006, with Differential Sampling by State**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	0.17	0.34	0.25	0.37	0.49	1.67	1.24
Alabama	1.05	11.46	1.13	1.76	2.98	102.43	37.72
Alaska	1.61	1.97	3.25	8.60	6.82	12.17	16.04
Arizona	0.58	2.21	0.55	1.27	2.02	471.48	19.19
Arkansas	1.64	4.47	1.81	4.45	6.11	48.56	65.42
California	0.38	0.87	0.41	2.78	5.36	9.36	2.83
Colorado	2.96	10.77	5.66	4.56	5.60	218.18	27.05
Connecticut	2.12	4.36	5.73	2.98	4.85	35.85	18.28
Delaware	1.76	4.48	2.64	2.99	4.67	49.96	29.32
District of Columbia	2.09	4.43	5.34	4.70	3.00	7.35	14.46
Florida	0.87	2.00	2.27	1.15	1.80	7.96	6.66
Georgia	1.64	3.89	5.05	2.32	2.86	23.53	11.95
Hawaii	2.62	6.04	5.02	4.54	5.31	19.99	9.88
Idaho	2.18	3.55	5.29	4.25	4.66	27.30	12.86
Illinois	0.68	1.02	1.32	1.31	2.31	4.55	5.96
Indiana	1.46	2.37	3.04	3.08	3.29	13.91	13.16
Iowa	1.49	3.11	1.83	4.02	7.07	33.48	17.88
Kansas	1.72	2.76	4.09	3.71	3.53	7.62	12.06
Kentucky	2.16	4.99	10.97	3.07	3.52	15.21	5.94
Louisiana	1.17	1.79	3.53	2.13	2.61	11.26	2.39
Maryland	1.23	3.94	1.22	3.36	4.09	20.12	14.36
Massachusetts	1.15	2.36	1.69	2.40	3.67	69.10	9.65
Michigan	1.11	2.66	1.25	3.69	4.31	28.38	11.83
Minnesota	1.53	2.83	3.23	2.20	5.03	15.99	13.56
Mississippi	0.98	1.49	1.84	2.60	2.40	13.05	9.74
Missouri	1.64	2.17	2.79	6.59	9.27	10.03	13.53
Montana	3.36	6.44	8.82	5.68	6.52	38.19	22.62
Nebraska	1.80	2.56	3.50	4.71	5.67	7.90	17.47
Nevada	1.87	3.50	5.06	3.02	3.61	15.66	11.40
New Hampshire	1.58	2.10	3.83	4.20	4.33	19.76	14.94
New Jersey	1.40	2.72	3.59	2.30	2.70	6.92	4.93
New Mexico	1.85	4.52	2.09	6.65	6.84	15.83	11.21
New York	0.65	1.08	1.06	1.60	2.02	6.43	6.62
North Carolina	1.39	3.36	2.80	2.26	2.72	14.73	14.31
North Dakota	2.34	3.91	5.04	4.35	5.94	15.82	20.11
Ohio	1.15	1.83	2.37	2.66	2.67	4.49	10.35
Oklahoma	2.06	4.67	3.34	3.87	4.58	26.83	14.28
Oregon	1.68	2.91	2.99	3.58	4.37	17.54	17.04
Pennsylvania	1.63	4.06	5.13	2.19	3.18	15.60	13.50
Rhode Island	2.17	5.00	5.98	3.35	3.77	11.39	22.24
South Carolina	1.11	2.06	1.60	3.15	2.81	11.01	14.50
Tennessee	2.27	17.66	33.83	2.28	4.68	157.08	24.41
Texas	0.76	1.57	2.35	0.98	1.46	9.39	7.90
Utah	1.04	1.31	2.06	4.47	3.92	8.25	5.23
Vermont	1.69	2.83	2.27	7.09	6.82	19.05	15.31
Virginia	1.37	2.08	3.41	2.76	3.23	9.19	11.13
Washington	1.18	2.67	1.46	3.71	3.29	7.36	10.17
West Virginia	1.72	4.15	5.35	2.92	2.58	11.52	9.49
Wisconsin	1.45	2.69	2.39	2.88	4.71	9.02	8.59
Wyoming	1.54	1.91	3.82	5.01	5.02	6.79	7.91

Source: MAX Person Summary files, 2006.

Notes: Data for Maine and South Dakota were not available. Foster children and infants appear in multiple columns.

**Table II.11. CVs of Mean Expenditures Among Records with Nonzero Expenditures by State and Enrollee Subgroup, 2006, with Differential Sampling by State**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	0.23	0.33	0.29	0.36	0.42	1.36	1.19
Alabama	1.31	1.06	1.69	2.86	2.38	6.59	4.61
Alaska	1.86	2.86	2.71	4.46	3.42	9.96	6.81
Arizona	2.10	2.96	3.32	11.74	5.12	4.92	3.62
Arkansas	1.53	2.47	2.59	2.31	2.65	11.87	10.23
California	0.70	1.08	0.99	1.20	1.24	4.17	5.96
Colorado	1.44	2.18	2.19	2.46	2.51	4.82	8.20
Connecticut	1.43	2.09	1.15	1.81	2.73	16.03	5.18
Delaware	1.60	2.22	1.15	3.85	3.47	11.61	6.86
District of Columbia	1.62	2.92	2.37	3.91	2.36	5.00	9.65
Florida	0.96	1.65	1.18	1.59	1.68	6.83	5.06
Georgia	1.31	1.74	1.43	2.36	2.68	5.86	5.41
Hawaii	1.63	1.49	1.05	3.55	3.66	6.82	10.29
Idaho	1.64	2.75	2.67	3.70	2.45	12.14	10.72
Illinois	1.38	3.02	2.43	2.22	2.01	9.04	8.62
Indiana	1.41	1.43	0.83	2.12	2.83	9.61	6.04
Iowa	1.50	2.34	2.01	1.96	2.48	6.02	8.75
Kansas	1.36	2.18	2.09	1.79	2.31	4.81	6.97
Kentucky	1.32	1.90	1.62	2.89	2.22	6.84	9.16
Louisiana	1.52	2.59	2.13	2.55	2.50	33.36	10.50
Maryland	1.46	2.73	2.28	2.89	2.18	8.54	10.58
Massachusetts	1.14	1.01	1.15	1.83	2.28	51.38	4.50
Michigan	1.05	1.67	1.27	2.20	1.74	7.24	7.88
Minnesota	1.29	2.14	1.39	1.56	2.07	10.16	7.85
Mississippi	1.57	2.71	2.13	2.82	2.74	18.10	8.60
Missouri	1.09	1.69	1.13	1.85	2.00	5.77	4.26
Montana	1.82	3.38	2.88	2.46	3.49	7.94	11.67
Nebraska	1.74	4.04	3.21	2.19	2.76	5.77	7.06
Nevada	2.03	2.91	2.39	3.50	3.55	7.27	10.04
New Hampshire	1.63	2.82	3.04	2.19	2.91	6.40	13.10
New Jersey	1.27	2.32	1.25	1.79	2.06	8.21	9.79
New Mexico	1.44	2.21	1.32	3.37	2.48	15.48	11.00
New York	0.94	1.41	0.89	1.30	1.71	7.82	3.95
North Carolina	1.12	1.85	1.64	1.99	1.92	8.62	6.07
North Dakota	1.97	4.09	3.14	2.31	3.43	9.39	12.10
Ohio	1.01	1.47	1.09	1.44	1.75	7.20	5.74
Oklahoma	1.69	2.37	2.45	2.34	3.32	8.21	6.72
Oregon	1.21	1.62	1.24	2.47	2.39	3.14	4.44
Pennsylvania	0.79	1.00	0.99	1.59	1.28	5.37	4.36
Rhode Island	1.74	2.69	1.15	2.82	2.89	5.70	13.88
South Carolina	1.48	2.36	2.17	3.10	2.63	8.94	7.86
South Dakota	1.90	3.12	3.07	2.92	3.42	6.89	9.98
Tennessee	1.51	1.85	1.93	3.19	2.71	11.54	11.77
Texas	0.91	1.41	0.99	1.22	1.77	7.16	3.89
Utah	2.06	4.39	2.57	4.04	3.27	6.36	8.80
Vermont	1.60	3.24	2.15	3.20	2.87	8.36	19.46
Virginia	1.37	2.31	1.35	2.56	2.26	9.42	8.22
Washington	1.56	1.74	1.85	2.07	3.23	15.22	12.16
West Virginia	1.34	2.39	1.94	2.85	1.94	11.29	7.21
Wisconsin	1.36	2.36	1.33	1.85	2.50	8.46	11.07
Wyoming	2.21	3.44	2.97	3.62	3.82	8.83	11.81

Source: MAX Person Summary files, 2006.

Notes: Data for Maine were not available. Foster children and infants appear in multiple columns.

**Table II.12. CVs of Mean Inpatient Expenditures Among Records with Nonzero FFS Expenditures by State and Enrollee Subgroup, 2006, with Differential Sampling by State**

State	Total Medicaid Enrollees	Non-Disabled Children	Non-Disabled Adults	Aged	Disabled, <65	Foster Children	Infants
U.S. Total	0.76	1.60	0.94	2.59	1.25	7.71	2.05
Alabama	2.31	10.66	2.41	3.95	6.76	61.04	56.42
Alaska	4.77	6.57	5.32	19.26	12.57	35.90	7.46
Arizona	6.06	13.57	5.17	42.87	24.78	421.35	22.64
Arkansas	5.64	8.85	4.81	5.12	11.08	58.07	12.03
California	1.94	5.37	2.37	5.50	3.38	17.48	9.21
Colorado	4.34	7.45	3.74	20.06	10.04	34.23	13.11
Connecticut	5.37	12.17	10.24	13.56	7.44	54.00	16.38
Delaware	9.16	19.35	8.39	13.74	21.84	102.15	17.51
District of Columbia	4.32	15.04	9.76	19.44	4.92	24.21	18.34
Florida	2.45	4.84	2.32	6.85	4.28	20.84	7.10
Georgia	4.23	7.30	4.20	14.01	7.12	37.52	8.81
Hawaii	10.13	38.23	16.29	22.98	13.47	74.91	49.86
Idaho	4.94	9.95	4.26	14.17	9.98	71.14	12.99
Illinois	3.99	7.89	5.89	15.78	5.95	28.00	10.54
Indiana	5.76	15.77	7.05	18.66	7.00	55.25	18.80
Iowa	5.02	8.07	4.28	10.07	10.80	37.86	11.43
Kansas	5.47	9.09	5.93	13.41	9.11	22.50	12.42
Kentucky	4.78	11.22	4.38	27.17	7.13	36.10	14.72
Louisiana	4.14	11.14	4.02	8.80	6.17	145.13	14.08
Maryland	6.78	18.44	8.91	25.02	8.82	47.22	21.49
Massachusetts	3.25	5.06	5.54	10.73	5.85	363.87	5.46
Michigan	5.18	9.78	6.09	33.73	8.74	45.89	12.06
Minnesota	7.25	21.28	7.92	20.28	8.81	61.40	23.90
Mississippi	3.31	7.09	3.46	5.28	5.48	48.37	10.94
Missouri	3.89	7.88	5.18	18.00	5.29	18.16	9.29
Montana	6.56	12.38	5.23	10.40	11.01	32.43	15.09
Nebraska	5.99	9.16	10.31	10.67	11.01	26.55	12.31
Nevada	5.96	10.94	6.10	21.81	9.01	32.12	16.49
New Hampshire	6.40	13.89	7.03	9.85	9.02	74.75	18.41
New Jersey	4.80	14.57	5.44	10.71	6.75	39.26	16.98
New Mexico	14.59	29.95	11.83	26.32	23.83	298.74	35.51
New York	2.39	5.32	2.91	7.54	4.14	31.73	6.34
North Carolina	3.02	5.95	3.28	17.57	5.11	42.99	8.06
North Dakota	8.27	10.53	6.93	49.40	17.83	28.13	14.74
Ohio	4.24	9.58	6.47	12.88	5.61	33.76	13.26
Oklahoma	3.43	6.28	4.03	10.42	6.24	37.88	7.77
Oregon	5.55	8.52	7.12	28.01	11.40	37.35	10.01
Pennsylvania	4.67	10.68	6.29	8.09	7.78	45.24	16.44
Rhode Island	7.86	20.29	23.29	18.29	9.23	47.03	23.65
South Carolina	4.35	8.10	4.88	22.40	7.44	63.40	9.85
South Dakota	6.21	9.49	6.26	26.13	12.20	38.51	12.57
Tennessee	4.84	9.39	5.78	23.94	8.28	61.82	16.33
Texas	3.19	5.80	1.93	7.00	5.79	36.18	7.03
Utah	5.25	8.58	5.62	22.57	11.27	62.07	10.06
Vermont	6.36	16.11	6.37	12.67	12.03	53.84	26.33
Virginia	5.12	13.26	6.84	8.45	7.48	61.68	15.79
Washington	6.58	26.23	13.22	14.09	8.02	73.91	35.42
West Virginia	5.01	26.02	8.63	16.04	5.61	102.23	22.22
Wisconsin	7.64	18.75	10.50	13.38	10.64	40.38	21.68
Wyoming	6.41	9.01	6.05	13.49	15.00	31.48	14.14

Source: MAX Person Summary files, 2006.

Notes: Data for Maine were not available. Foster children and infants appear in multiple columns.

and weighting of the expanded sample. Users of Medicare data, for example, can request 5, 10, or 20 percent samples, although only one 5 percent sample has been designed and documented for users.

A starting point for consideration of Mini-MAX sample expansion is to recognize that there are some research needs that cannot be addressed adequately by a sample. Researchers who wish to analyze selected states in great depth should be directed to the state MAX files. Mini-MAX will not be able to provide state samples of adequate size for such research except for the very largest states. Similarly, researchers who wish to examine very narrow subpopulations, especially when their analyses involve highly variable outcomes among these subpopulations or require statistically precise comparisons across states, should be directed to the population files. Mini-MAX cannot be expected to address such needs.

Another consideration is that increasing the sample size by four will reduce standard errors—and therefore, CVs—by one-half except where the sample begins to approach the size of the full population; in that case, the reduction will be greater. Thus, the precision obtainable from a 20 percent random sample can be determined by halving the CVs in Tables II.6 through II.8. The precision obtainable from a sample stratified by state with sampling rates four times those reported in Table II.9 can be determined by halving the CVs in Tables II.10 through II.12, except that the CVs for North Dakota and Wyoming would be reduced to zero (because the sample would include all records in these states), and the CVs for the next smallest states would be reduced to near zero.

Any potential expansion of the Mini-MAX sample must consider how the sample could better serve research needs focused on individual subpopulations rather than small states. Thus, it is likely that a 20 percent random sample would prove more useful to a broad range of users than a 20 percent sample stratified by state, but a targeted expansion assigning higher sampling

rates to specialized subpopulations would be even more useful. Any assessment of the expansion of Mini-MAX also should consider the reduced audience for this file associated with its resulting increased size and complexity, as well as any additional documentation needed to ensure its correct use.

Pooling samples across years is often used with annual surveys as a means to increase samples sizes for states and other subpopulations. With sample surveys, there is no recourse to a larger sample or population database, so pooling is attractive. However, pooling samples over time has clear drawbacks. In particular, the multi-year reference period makes interpretation of the timing of estimates problematic, and because of the sample overlap, estimates of year-to-year change using pooled samples are often misinterpreted. Mini-MAX users, however, will always have MAX as an alternative to pooling. Given that option, it is not clear why a user would find value in pooling Mini-MAX samples over time.

### III. OPTIONS FOR REDUCING FILE SIZE WITHOUT SIGNIFICANT LOSS OF RESEARCH CAPABILITIES

A five percent MAX sample of Medicaid beneficiaries results in Mini-MAX files that are far more manageable than the full MAX database but still too large for processing individually in most contexts. Files up to 3 to 6 GB can be processed on most desktop computers, and those up to 15 to 20 GB can be processed on mainframes. In comparison, a five percent sample of MAX PS records would be 9 GB, which is too large for standard desktop computer processing. The corresponding OT file is likely to reach 23 GB and would need to be split into several files to be processed even on a server. For this reason, participants of our TEP recommended developing “skinny” Mini-MAX files that contain only a subset of MAX variables. The TEP also requested that measures unique to MAX, including diagnostic, service, and program participation information, be kept in Mini-MAX.

#### A. Options for Reducing File Size by Excluding MAX Variables

The desire to create smaller Mini-MAX files coincides with the new storage format of MAX as a relational database, which allows far greater flexibility in selecting file subsets than was available in the past. Although users in the future may be able to select individual variables needed for their analyses, to simplify the selection process, we developed three options for the Mini-MAX—one containing all MAX variables and two containing a subset of increasingly critical measures.

**Sample Containing All MAX Data.** This option would have the same file structure and format as the MAX files from which it is drawn. The sample of records included in the Mini-MAX would be representative of all individuals in MAX, except S-CHIP-only enrollees, who are reported incompletely in MAX, and people with zero months of enrollment, who would be excluded. This approach would support MAX feasibility studies, program development testing,

and assessments of MAX data. It also would support a wide variety of Medicaid research studies.

**Skinny File Level 1: Excludes Duplicative, Rarely Used, and Unreliable Measures.** The Level 1 skinny file would exclude three types of MAX variables—measures duplicated elsewhere, rarely used or needed variables, and unreliable measures. Duplicative variables include those that summarize others in the file—for example, the race and sex combination and months of eligibility variables that researchers can create easily. Duplicative variables also include a series of eligibility measures replicated in each claims file.

The second type of variables excluded from proposed Level 1 includes measures rarely used or needed in research studies—for example, measures in the PS file summarizing the number of claims for each person by service type. Because the number of claims can vary by provider, state, region, and service, these measures are used primarily for state reporting assessment (better suited for full MAX files) and are of limited utility for research. Another example of a superfluous variable included in each MAX claims file is the Adjustment Code that documents how claims were combined.

Finally, Level 1 would exclude unreliable measures—including TANF enrollment, which is missing for most states, as well as claim third-party payment amounts and charge amounts that we understand are reported inconsistently.

**Skinny File Level 2: Excludes Linking and Proprietary Measures.** The Level 2 skinny file would exclude all variables excluded at Level 1, as well as MAX linkage variables and proprietary data. Linkage variables are identifiers that are often 8 to 12 bytes in length and use substantial space in a file. They include prepaid plan identifiers, waiver identifiers, and state-specific eligibility groups in the PS file—all of which, to be useful to researchers, require linkages to other data. In the claims files, identifiers include SSNs, billing provider, national

provider, managed care plan, and service provider identifiers. (Note that identifiers required to link across the Mini-MAX files—MSIS ID and State—would be retained). Also excluded from level 2 would be several inpatient and institutional long-term care summary variables in the PS file (for example, length of stays); these variables provide some aggregated information on topics that, to be studied appropriately (for example, to address censoring of stays), typically require claims-level analysis.

A complete list of MAX variables and proposed Level 1 and Level 2 exclusions for the MAX PS file and OT, RX, LT, and IP claims files are shown in Appendix Tables B.1 through B.5, respectively. Table III.1 shows the file sizes that would result from each option, by MAX file type. The Level 1 exclusions would reduce the PS file by a third, down from almost 9 to 6 GB. Level 2 exclusions would further reduce the PS file to 4 GB, manageable on most desktop computers.

**Table III.1. Estimated File Size of a Five Percent Sample of MAX and Levels 1 and 2 of MAX Skinny Files**

Measure	MAX Person Summary File	MAX Claim File			
		OT	RX	LT	IP
All MAX Records					
Number of records in millions	61.7	1,741.6	407.9	34.8	8.5
Size in GB	178.5	461.5	141.9	9.8	6.8
Potential MAX 5% sample					
Number of records in millions	3.0	87.1	20.4	1.7	0.4
Size in GB – all variables	8.9	23.1	7.1	0.5	0.3
Size in GB – level 1 skinny file	6.1	19.2	4.6	0.08	0.09
Size in GB – level 2 skinny file	4.0	12.5	2.5	0.06	0.09

The OT file is less frequently used but is by far the largest MAX file, due more to the large number of records than variables in the file. File size reductions thus are difficult to achieve by variable selection alone. Level 1 exclusions would reduce the all-variable 5 percent OT file by 17 percent, from 23 to 19 GB. Level 2 exclusions would reduce the file by an additional 29 percent to 13 GB. Although desktop users may need to request OT Level 2 records distributed

across several files, the PS and all other claims files could be processed on powerful desktop computers.

In addition to the all-variable and two skinny versions of the Mini-MAX, MAX variables could be aggregated further or summarized to create even smaller analytic files. One option would be to use the file layout of the MAX Enrollee Master File, which contains only summary demographic, enrollment, and service use variables but no expenditures. However, the file would have limited uses and many of the summary statistics it could produce already are available to the public in summary form.

Alternatively, if only one Mini-MAX format was desired, some combination of Level 1, Level 2, and more aggregated variables could be selected for Mini-MAX.

### **B. Options for Reducing File Size by Separating Files into Two or More Components**

Another option for reducing file size is splitting exceedingly large files into two or more components. This could be particularly useful for the OT file which, even when reduced to include only non-identifying, reliable, and unduplicated variables, exceeds 12 GB due to a large number of records. TEP members recommended considering splitting this file into multiple data files based on service type. Specifically, three to four comparably sized files containing claims for like services would result in file sizes below the 6 GB threshold under any of the three variable selection options.

We used MAX 2007 validation tables to determine which OT service or claims types could be separated into 6 GB or smaller files. We identified four non-overlapping claim groups, each about 5 GB in size, as promising candidates:

1. Capitated claims
2. Crossover claims

3. Remaining FFS and encounter claims for physician and other ambulatory services, corresponding to Part B services in Medicare (for example, physician, other practitioner, outpatient hospital, clinic, and dental services)
4. All other claims including community-based long-term care services, lab and X-ray, supplies, and other wraparound services

Please note that differential sampling rates by eligibility group may result in a different composition of claims than the 2007 data analyzed here. We therefore recommend that OT claim groupings be finalized after a Mini-MAX sampling design is selected to ensure that the files are an acceptable size.

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

#### **IV. TEP RECOMMENDATIONS**

In this chapter, we describe the recommendations made by the project's TEP on Mini-MAX design elements that would ensure the sample's usefulness to a broad set of users while minimizing the potential for user error. The TEP was composed of Medicaid researchers, sampling experts, and stakeholder agency representatives. There were seven formal TEP members. Three had experience using MSIS or MAX data, four had extensive knowledge of Medicaid, and three were experts in sample design and data development. Also contributing to the project were seven CMS representatives with extensive knowledge of MAX and other CMS data sources and who guided the aims of the project. Appendix A lists all TEP meeting participants.

Over the course of the five-month project, the TEP met three times. During the first in-person meeting, held on September 8, 2010, important background information on MAX was shared, and TEP members identified the critical elements of a Mini-MAX design. The second in-person meeting, held on October 19, was designed to evaluate the sampling and file size reduction options presented in Chapters II and III. During both meetings, more detailed design elements of Mini-MAX were discussed (for example, MAX records to be excluded, required documentation). To best characterize the TEP's views, at the end of the project, we conducted a survey of the seven formal TEP members on the most critical Mini-MAX design elements and convened a final meeting via phone on November 15 to discuss the survey findings. Features identified as critical by at least one TEP member are included in the final recommendations for the design of Mini-MAX. Below, we summarize the input we received from the TEP over the course of the project on four Mini-MAX design elements—expected audience, sample design, file structure, and documentation—focusing on the conclusions reached, rather than the issues

raised and opinions presented early in the project. Appendix C presents the survey instrument and a summary of the survey results.

### **A. Potential Users and Uses**

Ideally, Mini-MAX would be designed to serve a well-defined target population. However, this feasibility study did not include a market survey and, outside of CMS experience with the users of the Medicare five percent sample, little is known about the breadth, expertise, and Medicaid knowledge of potential Mini-MAX users. From past experience, TEP members suggested that it would be difficult to know the demand for such a data file before its construction; however, after such a file is created, collecting information that would indicate needed refinements for Mini-MAX would be essential to facilitate its use.

Four important potential audiences for Mini-MAX were discussed: (1) state government officials who could use the files to quickly compare outcomes in their state with those of other states, as well as identify issues with their state's reporting; (2) federal government officials, who in the past have relied on contractors to support their MAX studies but who could address ad hoc policy questions with Mini-MAX; (3) students, faculty, and other research staff with fewer resources than required for studies using complete MAX files; and (4) current MAX users who could use Mini-MAX to develop MAX study designs. TEP members noted that it is difficult to predict which of these user types would most benefit from Mini-MAX but that the audiences attracted would depend on the timing and access restrictions of the file.

The TEP disagreed on when to release Mini-MAX but agreed that reducing the lag in data production would be desirable. Specifically, at least one TEP member thought that the final MAX files should be incorporated into Mini-MAX (implying a two-year lag between the end of the represented calendar year and Mini-MAX release), whereas others viewed such a lag as unacceptable and would not support such a time lag. Two options were discussed for an earlier

release with a 16-month lag. The first option involves using early-release “Beta-MAX” files in Mini-MAX production.<sup>9</sup> Alternatively, one TEP member suggested that Mini-MAX be produced before MAX, using the current year’s data for states with complete data and that of the previous year for states with incomplete current-year data. Several TEP members expressed concerns about the complexity and additional documentation required to implement a pre-MAX release of Mini-MAX. The TEP did not reach consensus on the timing of the database’s production and, based on that, its content.

In addition to the timing of its production, the lag between initiating a request and receipt of Mini-MAX data also was discussed. As a RIF, requests for Mini-MAX will require privacy board review, which implies substantial delays in obtaining the data. Typically, there is a four- to six-month wait between initiating a file request at ResDAC and actual receipt of the requested files. Although the pre-specified Mini-MAX design may speed up the process by one to two months, the wait of several months to access the files will limit its widespread use to address ad-hoc Medicaid policy questions or develop study designs when timing is of concern.

Furthermore, use of RIF files is limited to that specified in the data use agreement. Thus, the RIF format likely will preclude use of Mini-MAX to address ad-hoc research questions or develop specifications for future studies. While TEP members agreed that there is an unmet need for data to address ad hoc questions on the Medicaid program within a limited timeframe (less than two months), they also expressed concern about greatly expanding the audience of MAX to include less experienced and knowledgeable persons without additional regard for user education and data privacy assurances.

---

<sup>9</sup> Beta-MAX is an early-release version of MAX (starting with calendar year 2009). Because some MSIS data may not be approved at the time of its production, data for all states may not be included in Beta-MAX. Therefore, Beta-MAX is expected to be less complete and contain more errors than MAX.

If Mini-MAX is created, the TEP members agreed that it would be valuable to create a log of file requestors and solicit their input on how the file might be modified to better meet their needs. CMS currently has a list of individuals who have received or expressed interest in receiving MAX data. These individuals could be contacted for input. Alternatively or in addition, a survey of Mini-MAX data requestors or a review of Mini-MAX data use agreements should be conducted to ensure that future alterations to Mini-MAX can be better informed by user needs.

## **B. Sampling Approach**

The TEP discussed three components of the sampling design: stratification of the sample by state and major eligibility group, the need for oversampling other Medicaid subgroups, and whether Mini-MAX should contain a longitudinal component.

**Stratified Sampling.** There was general consensus among the panel members that if Mini-MAX is developed, it should allow for precise estimates at the state level. All responding TEP Medicaid experts felt that stratified sampling by state, with higher sampling rates for smaller states, would be critical or important to the Mini-MAX design. National estimates remain important, however, so no one recommended that all state samples should be of equal size.

Views regarding whether the sample should be stratified by major eligibility group (child, adult, aged, and disabled) in addition to state were more varied, in part because achieving a more uniform distribution of sample observations across these four groups would imply an oversampling of aged and disabled enrollees, which would require reducing the sample of children, including infants and foster care children. Some TEP members thought this trade-off would produce a more useful file because the aged and disabled are high-cost populations. In

addition, many of the children and adults are enrolled in managed care and so would not have fee-for-service claims data for analysis.<sup>10</sup> One TEP member viewed stratified sampling by state and the four eligibility groups as a critical feature of Mini-MAX, one member identified it as an important feature, and two considered it unnecessary. Based on the recommendation of the one TEP member and the approval of other TEP members, stratified sampling by the four eligibility groups in addition to state is included as a critical feature of Mini-MAX.

**Oversampling Other Subgroups.** TEP members indicated that oversampling small subpopulations was important, but no member identified any of the specific subpopulations (infants, foster care children, dual Medicare and Medicaid enrollees) listed in the survey as critical. The precision estimates developed for foster children and infants in Chapter II indicate that, without substantial oversampling, a five percent Mini-MAX sample will not include sufficient numbers of individuals in these and other small subgroups to support their analysis.

CMS staff also discussed the privacy board restrictions requiring that researchers be provided with the minimum amount of data necessary to conduct their research. Given these restrictions, the privacy board likely will require researchers conducting analysis on only a small subpopulation to use an extract of the full MAX files specific to their population of interest rather than giving them approval to use Mini-MAX. The panel members concluded that, other than stratification on one or two dimensions, the sampling should be kept simple. Given this constraint on the design, any Mini-MAX documentation should specify that the file cannot support analysis of specific small populations and indicate the minimum cell size required to produce estimates of adequate precision under various circumstances.

---

<sup>10</sup> As the quality of encounter data improves, however, there may be more interest in studying the children and adults and so may require a re-evaluation of prioritized populations represented in Mini-MAX.

**Cross-Sectional Versus Longitudinal Sample.** The TEP did not reach a final conclusion regarding the importance of including a longitudinal component in the file. Sample design experts indicated that longitudinal components frequently were requested for other data sets. Furthermore, longitudinal data might be very useful to Medicaid researchers. (For example, for comparative effectiveness research, an initial set of individuals with similar health status is identified in one period. This initial sample then is followed over time to determine whether differences in treatment result in substantial differences in outcome.) However, there are drawbacks to including a longitudinal component in Mini-MAX. The TEP cited two concerns. First, such a component would imply that independent random samples could not be pooled across multiple years to increase precision. However, the TEP noted that the benefit of pooling would be rather limited in the Medicaid context and that researchers needing larger samples of specific subpopulations would be better served by extracts of the full MAX files. Second, individuals expressed concern about the potential complexity of analysis weights or other challenges to using the file. Two TEP members identified a longitudinal component as important for Mini-MAX, one person felt it was unnecessary, while others did not express an opinion. The guidance we take from the TEP's input is that, provided it did not further complicate the sample, a longitudinal component would be desirable but not critical.

### **C. File Structure**

Although TEP members expressed potential interest in a variety of formats for Mini-MAX (for example, 5 percent, 10 percent, or 20 percent files, and a file containing all MAX variables), during our final meeting, they indicated a desire for only one Mini-MAX design. A single Mini-MAX would produce replicable results, require a single source of documentation, and minimize sample product confusion. The TEP also expressed various levels of interest in decreasing the

size of MAX, but emphasized the importance of having a simple file structure of one file per file type.

Of the two slim file options, the TEP members preferred the option that includes the identifiers. Specifically, the TEP concluded that the optimal Mini-MAX design was the skinny file version 1 described in Chapter III, which excludes only duplicative, less frequently used, and unreliable measures in MAX. Although mentioning concern about loss of information, the TEP agreed that the benefits to file size outweighed the loss of measures used by few if any MAX researchers. In contrast, identifiers potentially excluded in skinny file version 2 were of greater interest to the TEP. One TEP member noted that currently there is a great deal of interest in continuity of care in the researcher community, and these identifiers would be useful for that type of analysis even if not linked to identifying information or provider characteristics. However, at least one TEP member and CMS suggested that a few additional measures might be considered for exclusion from Level 1. For example, CMS may want to consider also excluding some lesser used linking variables, proprietary information, and potentially a few other Level 2 excluded variables from Level 1. TEP members did not view the release of a Mini-MAX option with aggregated or summarized information as important.

All TEP members agreed that a simpler file structure including only one PS file and one or two files for each claim type would be an important advantage to researchers using Mini-MAX rather than full MAX files. Currently, there are five MAX files for each state (PS, IP, LT, RX, OT), and reading through the 255 files to produce national estimates is a time-consuming process. Despite the desire for a simpler file structure, TEP members agreed that the OT claims file should be divided into three or more components (into multiple claims types) to keep its size less than 6 GB so that individuals with a basic personal computer would be able to process them.

## D. Documentation

TEP members agreed that Mini-MAX should not be produced unless it is accompanied by a high-quality, thorough, accessible, and readily citable user's guide. All responding TEP members thought it was critical or important that the user's guide include the following elements:

- Information on reporting anomalies across states
- A guide for applying file weights and calculating standard errors
- Sample programs
- A discussion of research uses that cannot be supported by Mini-MAX
- Information on whether the sample sizes in Mini-MAX will be sufficient to support specific types of research

TEP members initially disagreed on whether or not a summary of Medicaid program differences across states would be a critical component of the user's guide. Two TEP members considered such information critical. However, comprehensive information on Medicaid program differences currently is not available in a readily accessible format. For example, complete state plans are stored in paper files at CMS regional offices.<sup>11</sup> Other TEP members did not want the lack of such data to preclude the development of Mini-MAX. During the teleconference, CMS suggested that the Mini-MAX documentation contain a link to data currently being collected under CMS's Medicaid/CHIP Environmental Scanning and Program Characteristics Database (ESPC) contract. This database will contain information from existing Medicaid program characteristic and summary outcome data sources for each of the 50 states and the District of Columbia. While it should be noted that the one TEP member who indicated

---

<sup>11</sup> See Borden et al. (2010), page 13, for additional details.

information on program differences was a critical component of the user's guide was not present, all others thought reference to these existing sources would be sufficient.

The TEP also agreed that the user's guide must provide information to potential users on whether the sample sizes in the file will be sufficient to support their research. Specifically, CMS and TEP members expressed concern that researchers may go through the trouble and cost of accessing Mini-MAX to learn that the sample cannot support their analysis. Therefore, this documentation component is particularly important, since users will need to assess the sufficiency of the sample for their research before purchasing the file.

However, the group did not come to a consensus regarding the format in which this information should be provided. TEP members suggested that CVs were difficult to interpret. One panel member said that metadata (for example, sample sizes, means) comprised a critical component to the documentation and could serve this purpose.

One panel member suggested that documentation prepared by NORC for other data sources would be an excellent model. Also, CMS recently identified the National Center for Health Statistics' documentation for the National Health and Nutrition Examination Survey (NHANES) as another potential model. The NHANES documentation includes an on-line tutorial that assists users to assess the adequacy of the size of the sample for a potential study. Such a component, if possible to implement for Mini-MAX, would strengthen its documentation.

## **E. Summary**

This feasibility study was supported by a TEP of Medicaid and sampling experts that provided a broad research perspective on critical Mini-MAX features and helped to ascertain whether a design could be developed that would meet broader user needs without compromising research quality. Although all participants supported the Mini-MAX effort, substantial concern was expressed by the TEP about potential user error as a result of lack of sufficient knowledge of

data reporting anomalies, Medicaid program differences, and complex sampling schemes. Overall, the TEP supported the creation of Mini-MAX, provided that five critical elements were included in its design, with the details to be developed in a separate design study:

1. A single design (one Mini-MAX database)
2. Timely access to Mini-MAX
3. Stratified sampling by state and eligibility group (child, adult, aged, and disabled)
4. A simplified file structure, excluding unreliable and duplicative MAX measures and splitting the OT file into three or more files by service type, each less than 6 GB
5. A user's guide containing six elements—data anomalies, links to information on program differences, a guide for applying weights, sample programs, discussion of research not supported by Mini-MAX, and information needed to determine what research *is* supported by Mini-MAX

However, the TEP did not reach a consensus regarding how best to achieve timely access to Mini-MAX or document research supported by Mini-MAX in a user's guide.

The following elements were considered by the TEP as important but not critical for Mini-MAX:

- Oversampling of foster care children, infants, and other smaller groups
- A longitudinal (cross-year) sample component

Finally, the TEP recommended that Mini-MAX exclude S-CHIP-only enrollees and people with no months of eligibility because data for these groups are incomplete. The TEP also recommended that CMS solicit input about the needs of potential Mini-MAX users once the sample is created.

We should note that some members of the TEP also expressed interest in the development of a Limited Data Set (LDS) with fewer identifiers and speedier access to the files, while others expressed interest in a Public Use File (PUF) based on MAX. Although further investigation of the feasibility of developing such data sets was not within the scope of this project, one TEP member suggested that CMS's decision to produce Mini-MAX should include considerations of

opportunity costs. The TEP suggested that, while there may be value to producing a sample file, CMS also should consider alternative approaches to increasing the volume of Medicaid research before making a decision to produce Mini-MAX.

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

## V. CONCLUSIONS

Creating a Medicaid sample has long been viewed as infeasible, not because of technical reasons, but rather because of the Medicaid program's complexity, which makes it difficult to identify a broadly useful sample design. In addition to complex Medicaid eligibility and service coverage rules that vary across states, reporting anomalies make any file design that may be useful for one research study inappropriate for another. That is, a Medicaid sample cannot serve all Medicaid research interests. This study assessed the feasibility of specifying a sample design for a research identifiable file (RIF) that has the potential to extend information in MAX to a broader audience and expand the scope of high-quality Medicaid research.

The study consisted of three main components: (1) an analysis of MAX data to assess the precision of various sample-based estimates, (2) the development of options for reducing file size and simplifying the file structure to address barriers to MAX analyses and thus maximize the number of potential users of Mini-MAX, and (3) a TEP that advised the project on the optimal Mini-MAX design. The three components led to conclusions regarding options for a feasible Mini-MAX and recommendations for its design. Below, we summarize these conclusions and recommendations and describe the next steps required to implement Mini-MAX.

### A. Options for Mini-MAX

Our analyses of MAX data and its contents suggested that some options for Mini-MAX are better than others:

- **Differential sampling by state is critical.** Although national estimates from a Medicaid sample may be useful, variations in how state Medicaid programs are designed and in the data reported by states suggest that users of Mini-MAX will want to produce state estimates. To support such usage, Mini-MAX will need to sample smaller states at higher rates than larger states.
- **Small population studies cannot be served by a five percent sample.** Some subgroups, such as foster children and infants, are so small that they require full MAX data. Analyses focusing on in-depth analyses of small populations or select states cannot be supported by a five percent sample.

- **A five percent sample could support state-level estimates of use and expenditures.** Our analyses suggest that a five percent sample with differential sampling by state could be used to estimate a wide range of service use and expenditure outcomes at the state level and for larger subgroups.
- **A five percent sample (without substantial loss of information) could be analyzed on a desktop computer.** A five percent sample that excludes some non-critical MAX variables and splits the OT file into several components would result in Mini-MAX files under 6 GB that could be processed on most desktop computers. A larger sample would require some loss of information or increased file complexity.
- **A 10 percent MAX sample would require substantial tradeoffs.** A 10 percent sample would enable a wider range of analyses but would contain files too large to be analyzed on desktop computers, multiple files per file type, or files that exclude a larger number of variables.
- **Insufficient information is available about the needs of potential users.** The priorities for Mini-MAX should be determined by the needs of its audience, but little is known regarding whether time delays, the size and complexity of the files, costs, or other factors are the principal obstacles to MAX data use. To increase users' access, additional information is needed about their possible needs.

Overall, the feasibility study indicates that selecting a five percent sample and removing less commonly used, unreliable, and linking variables would substantially reduce the computing resources required to conduct analysis of MAX data and produce a file capable of addressing many common research questions with an acceptable level of precision. A five percent sample of MAX data with differential sampling by state and major eligibility group could support analyses of Medicaid utilization and expenditures for a wide range of services by state, as well as each of four eligibility groups (child, adult, aged, and disabled) within states.

## **B. Mini-MAX Design Recommendations**

Given these findings, the TEP members supported production of a five percent Medicaid sample to increase the volume of research conducted with MAX data, provided it has the following features:

- **Single design.** Mini-MAX should include only one “best” design to minimize any confusion about data products or their associated documentation and ensure that research is replicable.

- **Timely release.** The file should be released as close as possible to the release of Beta-MAX or MAX.
- **Stratified by state and four eligibility groups (child, adult, aged, disabled).** The need to produce state-level estimates for costly groups like aged or disabled enrollees will require that the sample be stratified to yield appropriate sample sizes for each state and eligibility group.
- **Simple file structure.** Mini-MAX should include only one PS file and one file for each claim type (with the very large OT file split into three or more files by detailed claim type). The size of each individual file should be under 6 GB.
- **User's guide.** Since one goal of producing Mini-MAX is to expand the community of MAX users, such a sample should only be produced if it will be accompanied by user-friendly documentation. This documentation should include information on reporting anomalies across states, links to sources of information on Medicaid program differences across states, a guide for applying file weights and calculating standard errors, sample programs, a discussion of research uses that cannot be supported by the file, and information sufficient for users without direct access to the data to determine whether the sample sizes in the file will be sufficient to support their research.

The TEP also recommended that Mini-MAX exclude S-CHIP-only enrollees and people with no months of eligibility, and that CMS solicit input about the needs of potential Mini-MAX users once the sample is created.

Other file features were discussed at the TEP meetings and considered part of the feasibility study, but the features listed above were considered critical or important. In addition, CMS privacy board requirements that researchers be provided with only the minimum data necessary to conduct their research imply that the niche for Mini-MAX comprises studies that will produce estimates for all states and major subpopulations. Researchers proposing to focus on narrow subpopulations are likely to be asked to request extracts from the full MAX files rather than being given access to Mini-MAX. This reinforces the TEP recommendation that Mini-MAX not oversample subpopulations other than the four larger eligibility subgroups (child, adult, aged, and disabled). However, TEP members suggested that to best serve future needs, the design of Mini-MAX should be reassessed as user needs become more clearly established.

### C. Next Steps

This study assessed the feasibility of creating a sample Medicaid RIF and recommended Mini-MAX design elements. However, the study was not designed to provide specifications or documentation sufficient for its implementation, both of which will be needed before work to create a sample can begin.

Assuming that CMS approves the recommendations made by the TEP, decisions on several additional aspects of the sample design are required before specifications can be developed. Specifically, the following questions must be addressed:

- What are the implications of privacy board requirements that researchers be provided only the minimum data necessary to conduct their analyses? How does this limit potential uses of Mini-MAX, and what does this suggest about the design of the sample?
- Will the sample include a longitudinal component?
- What sampling rates will be implemented by state and eligibility group?
- Which year and source of data (MAX or Beta-MAX) will provide the base for the sample?
- Will SSNs, when present, be used to select the sample and, if so, what will be used for the 10 percent of records that lack SSNs?

It is common among agencies selecting samples from administrative records to use SSNs. Furthermore, in every case that we are aware of, this is done in such a way that the resulting sample is longitudinal. Absent objections by CMS to including a longitudinal component, we would favor using SSNs for the 90 percent of records with SSNs and incorporates a longitudinal component in Mini-MAX. Given CMS's input on this issue and the remaining questions above, specifications for Mini-MAX could be developed quickly.

Finally, the most critical next step before implementing a Medicaid sample is to create a plan for designing and implementing a high-quality user's guide and documentation for Mini-MAX. We emphasize the importance of clear, high-quality, and complete documentation for the

sample to ensure that the aims of Mini-MAX—to expand the high-quality research conducted using MAX—are met.

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

**REFERENCES**

- Borden, William S., Cheryl A. Camillo, Matthew Potts, Megan Reilly, Cara Weiss, and Amy Wodarek O'Reilly. "Improving the Completeness, Accuracy, and Timeliness of Medicaid Data." Report submitted to the Centers for Medicare & Medicaid Services, Disabled and Elderly Health Programs Group. Princeton, NJ: Mathematica Policy Research, July 13, 2010.
- Perez, Victoria, Bob Schmitz, Audra T. Wenzlow, Kathy Shepperson, Dave Baugh, and Susan Radke. "The Medicaid Analytic eXtract 2004 Chartbook." Report submitted to the Centers for Medicare & Medicaid Services. Cambridge, MA: Mathematica Policy Research, May 2008.
- Schneider, Andy, Risa Elias, Rachel Garfield, David Rousseau, and Victoria Wachino. *The Medicaid Resource Book*. Washington, DC: Kaiser Commission on Medicaid and the Uninsured, July 2002.
- Stuart, Alan. *The Ideas of Sampling*. New York: Oxford University Press, 1984.

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

## GLOSSARY OF SAMPLING TERMS

Coefficient of variation (CV) = A CV expresses the standard error of an estimate as a percentage of the estimate. For example, a CV of one percent on an estimated mean of \$1,000 implies a standard error of just \$10. Since the 95 percent confidence interval of an estimate is plus or minus approximately two standard deviations, a CV of one percent in this case implies a confidence interval of plus or minus two percent of the estimate, or plus or minus \$20. CVs are useful in comparing precision across estimates with different means. Furthermore, the absolute value of the CV is informative about the degree of precision. A CV below one percent is indicative of a high level of precision, whereas a CV above 10 percent begins to raise concern about the precision of an estimate; a CV of 20 percent or more (indicating a confidence interval of plus or minus 40 percent) indicates a low level of precision for most purposes.

Confidence interval = A confidence interval for a parameter (for example, a mean or a percentage) expresses the likely range of the unknown true value relative to the value estimated from a single sample. For example, saying that the 90 percent confidence interval for a sample mean of 3.5 is  $\pm 0.8$  implies that we can be 90 percent confident that the true value lies between 2.7 ( $3.5-0.8$ ) and 4.3 ( $3.5+0.8$ ). That is, the 90 percent confidence interval is (2.7, 4.3). The notion of 90 percent confidence derives from the fact that if we were to draw repeated samples and estimate the mean and the interval from each one, then 90 percent of the intervals would include the true value.

Cross-sectional sample design = A cross-sectional sample design is one in which data are collected from participating subjects at a single point of time or period. If the sample is repeated (for example, a year later), the subjects will be sampled regardless of whether they were in the previous year's sample. That is, the subjects in both samples may overlap somewhat or contain data from a completely different set of subjects.

Longitudinal sample design = A longitudinal sample design is one in which the data are collected from the same subjects at more than one point in time. We can compare the subjects to themselves over time to measure change that is gross rather than net—for example, what proportion of the population experienced an increase in income versus how much the average income changed.

Oversampling = Oversampling is systematically sampling any particular group of population at a higher rate than the other groups. Oversampling is usually done for small groups. This helps ensure that we have a large enough sample from a particular group of interest to achieve the desired precision in estimation. For example, since Wyoming and North Dakota have the fewest disabled Medicaid enrollees, we may need to oversample these states to produce estimates for disabled enrollees with the desired precision.

Precision = Precision refers to how closely the value estimated from a sample approximates the true value. Precision increases as the sample size increases. Precision is usually measured by the standard error—the smaller the standard error, the greater the precision.

Probability proportional to size (PPS) = PPS is a sampling technique in which the probability of selecting a unit is proportional to the size of the unit. For example, if there are 60 males and 40 females and we are to select a PPS sample of 20, we would select 12 females and 8 males.

Random sample = A random sample is “a sample selected by a chance mechanism with known chances of selection.” (Stuart, 1984, p. 3)

Sample size = Sample size is the total number of units in a sample selected from a population.

Sampling frame = “Sampling frame is the actual set of units from which the sample is drawn” (<http://www.statistics.com/resources/glossary/>). For example, the sampling frame can be the list of all Medicaid enrollees in a state in a given year

Sampling unit = The fundamental unit of sampling is called the sampling unit. For example, for the Mini-MAX study, “persons” are the sampling unit.

Selection probability = The selection probability is the likelihood of selecting a unit. The probability is always between 0 and 1. For example, if there are 20 students and we are to sample 6 students, the selection probability for each student is 6 out of 20, or 0.3 (or 30 percent).

Simple random sample = A simple random sample is one selected by a process that gives every possible sample (of that size from that population) the same chance of selection. (Stuart, 1984, p. 4)

Standard error = The standard error of a sample estimator (for example, the mean or a percentage) is a measure of how widely the estimates from repeated samples of the same size will be distributed around the true value. The standard error diminishes as the sample size increases. It is used to calculate the confidence interval around a sample statistic.

Strata = Strata are two or more mutually exclusive subgroups of a sampling frame. For example, if we are sampling from a list of people, we might define male and female as separate strata.

Stratified sampling = Stratified sampling is a method of drawing a sample from a population in which the sampling frame is divided into mutually exclusive subgroups, called strata, and a sample is selected from each of these strata. To maximize the benefits of stratified sampling, the members of each stratum should be similar to each other (homogeneous) and different from the members of other strata with respect to the characteristics that the sample will be used to measure.

Variance = “The variance is a numerical value used to indicate how widely individuals in a group” are distributed around the group mean. “If individual observations vary greatly from the group mean, the variance is big” (<http://stattrek.com/Help/Glossary.aspx>).

**APPENDIX A**

**MINI-MAX TEP MEETING PARTICIPANTS**

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

## TEP Members

*Stephen Ash*

Victimization and Expenditures Branch  
Demographic Statistical Methods Division  
Bureau of the Census  
U.S. Census Bureau

*Stephen Crystal*

Director, Center for Health Services Research on Pharmacotherapy, Chronic Disease Management, & Outcomes; & Center for Education & Research on Mental Health Therapeutics  
Research Professor & Chair, Division on Aging  
Rutgers University

*Michael Davern*

Director of Public Health Research  
National Opinion Research Center (NORC)  
University of Chicago

*April Grady*

Principal Analyst  
Medicaid and CHIP Payment and Access Commission (MACPAC)

*Genevieve (Jenny) Kenney*

Senior Fellow & Health Economist  
Urban Institute

*Kenneth Finegold*

Program Analyst  
Division of Health Care for Low Income Populations  
Office of Health Policy, Office of the Assistant Secretary for Planning and Evaluation (ASPE)

*Tom Selden*

Economist  
Agency for Healthcare Research and Quality (AHRQ)

**CMS Representatives from the Office of Research, Development, and Information,  
Centers for Medicare & Medicaid Services**

*Susan Radke*  
Mini-MAX Project Officer

*Dave Baugh*  
Senior Technical Advisor

*William Clark*  
Director, Division of Research on State Programs & Special Populations

*Christine Cox*  
Director, Division of Research & Information Dissemination

*Mary Kapp*  
Deputy Director, Research and Evaluation Group

*Kim Lochner*  
Technical Advisor

*Renee Mentnech*  
Director, Research & Evaluation Group

**Project Staff at Mathematica Policy Research**

*Ellen Bouchery*  
Principal Program Analyst

*John Czajka*  
Senior Fellow

*Julie Sykes*  
MAX-PDQ Project Director

*Shinu Verghese*  
Senior Systems Analyst

*Audra Wenzlow*  
Senior Researcher

## **APPENDIX B**

### **MAX 2007 RECORD LAYOUT AND POTENTIAL VARIABLES TO BE EXCLUDED FROM MINI-MAX**

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

Table B.1. MAX 2007 PS File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
***	MEDICAID ANALYTIC EXTRACT (MAX) PERSON SUMMARY RECORD	1	2895	2895					
***	MEDICAID ELIGIBILITY VARIABLES	1	1175	1175					
**	IDENTIFYING GROUP	1	81	81					
1.	MSIS IDENTIFICATION NUMBER	1	20	20					
2.	STATE ABBREVIATION CODE	21	22	2					
3.	MAX YEAR DATE	23	26	4					
4.	SOCIAL SECURITY NUMBER - FROM MSIS	27	35	9					
5.	SSN HIGH GROUP TEST - SSN FROM MSIS	36	36	1					
6.	SOCIAL SECURITY NUMBER FROM EXTERNAL SOURCE	37	45	9					
7.	EXTERNAL SOCIAL SECURITY NUMBER (SSN) SOURCE	46	46	1					
8.	STATE CASE NUMBER	47	58	12	1	12			Can use MSIS ID for linking
9.	MEDICARE HEALTH INSURANCE CLAIM (HIC) NUMBER - FROM MSIS	59	70	12	1	12			Use more reliable Medicare code
10.	MEDICARE HEALTH INSURANCE CLAIM (HIC) NUMBER - FROM MEDICARE	71	81	11					
**	DEMOGRAPHIC GROUP	82	126	45					
11.	BIRTH DATE	82	89	8					
12.	AGE GROUP CODE	90	90	1					
13.	SEX CODE	91	91	1					
14.	RACE/ETHNICITY CODE	92	92	1					
15.	RACE - WHITE	93	93	1					
16.	RACE - BLACK/AFRICAN AMERICAN	94	94	1					
17.	RACE - AMERICAN INDIAN/ALASKAN NATIVE	95	95	1					
18.	RACE - ASIAN	96	96	1					
19.	RACE - NATIVE HAWAIIAN/OTHER PACIFIC ISLANDER	97	97	1					
20.	ETHNICITY - HISPANIC OR LATINO	98	98	1					
21.	MEDICARE RACE/ETHNICITY CODE	99	99	1	1	1			Use more complete Medicaid code
22.	MEDICARE LANGUAGE CODE	100	100	1					
23.	SEX-RACE CODE	101	101	1	1	1			Easily created by researchers
24.	MEDICAID DEATH DATE	102	109	8					
25.	MEDICARE DEATH DATE	110	117	8					
26.	MEDICARE DEATH DAY SWITCH	118	118	1					
27.	DATE OF DEATH (FROM SSA DEATH MASTER FILE)	119	126	8					
28.	RESIDENCE COUNTY CODE	127	129	3					
29.	RESIDENCE ZIP CODE	130	134	5					
**	ANNUAL MEDICAID AND OTHER HEALTH INSURANCE GROUP	135	152	16					
30.	STATE SPECIFIC ELIGIBILITY CODE - MOST RECENT	135	140	6					
31.	MAX UNIFORM ELIGIBILITY CODE - MOST RECENT	141	142	2					
32.	MISSING MEDICAID ELIGIBILITY DATA SWITCH	143	143	1	1	1			All people will be enrollees
33.	MONTHS OF ELIGIBILITY	144	145	2	1	2			Easily created by researchers
34.	PRIVATE INSURANCE MONTHS COUNT	146	147	2	1	2			Easily created by researchers
35.	MEDICARE DUAL CODE - ANNUAL	148	149	2					
36.	MEDICARE BENEFICIARY MONTHS COUNT	150	151	2	1	2			Easily created by researchers
37.	MEDICARE ORIGINAL ENTITLEMENT REASON CODE	152	152	1					
**	MEDICARE DUAL GROUP - MONTHLY (OCCURS 12 TIMES)	153	176	24					

Table B.1. MAX 2007 PS File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
38.	MEDICARE DUAL CODE - FIRST MONTH	153	154	2					
**	STATE SPECIFIC ELIGIBILITY GROUP - MONTHLY (OCCURS 12 TIMES)	177	248	72			1	72	Forgo state for national groups
39.	STATE SPECIFIC ELIGIBILITY CODE - FIRST MONTH	177	182	6					
**	MAX UNIFORM ELIGIBILITY GROUP - MONTHLY (OCCURS 12 TIMES)	249	272	24					
40.	MAX UNIFORM ELIGIBILITY CODE - FIRST MONTH	249	250	2					
**	PRIVATE INSURANCE GROUP - MONTHLY (OCCURS 12 TIMES)	273	284	12					
41.	PRIVATE INSURANCE CODE - FIRST MONTH	273	273	1					
**	MEDICARE BENEFICIARY GROUP - MONTHLY (OCCURS 12 TIMES)	285	296	12					
42.	MEDICARE BENEFICIARY CODE - FIRST MONTH	285	285	1					
**	PRE-PAID PLAN MONTHS COUNT GROUP - PLAN TYPE (OCCURS 7 TIMES)	297	310	14					
43.	PRE-PAID PLAN MONTHS COUNT - FIRST PLAN TYPE	297	298	2					
**	PRE-PAID PLAN ENROLLMENT GROUP - MONTHLY (OCCURS 12 TIMES)	311	982	672					
44.	PRE-PAID PLAN TYPE-1 CODE - FIRST MONTH	311	312	2					
45.	PRE-PAID PLAN IDENTIFIER-1 - FIRST MONTH	313	324	12			12	144	Forgo linking in Level 2
46.	PRE-PAID PLAN TYPE-2 CODE - FIRST MONTH	325	326	2					
47.	PRE-PAID PLAN IDENTIFIER-2 - FIRST MONTH	327	338	12			12	144	Forgo linking in Level 2
48.	PRE-PAID PLAN TYPE-3 CODE - FIRST MONTH	339	340	2					
49.	PRE-PAID PLAN IDENTIFIER-3 - FIRST MONTH	341	352	12			12	144	Forgo linking in Level 2
50.	PRE-PAID PLAN TYPE-4 CODE - FIRST MONTH	353	354	2					
51.	PRE-PAID PLAN IDENTIFIER-4 - FIRST MONTH	355	366	12			12	144	Forgo linking in Level 2
**	MEDICAID MANAGED CARE COMBINATIONS GROUP - MONTHLY (OCCURS 12 TIMES)	983	1006	24	1	24			Can use unique codes but combinations miss important MC types
52.	MEDICAID MANAGED CARE COMBINATIONS - FIRST MONTH	983	984	2					
**	DAYS OF ELIGIBILITY GROUP - MONTHLY (OCCURS 12 TIMES)	1007	1030	24	1	24			Recommend using months instead
53.	DAYS OF ELIGIBILITY - FIRST MONTH	1007	1008	2					
**	TEMPORARY ASSISTANCE FOR NEEDY FAMILIES (TANF) CASH FLAG GROUP - MONTHLY (OCCURS 12 TIMES)	1031	1042	12	1	12			Incomplete for most states
54.	TEMPORARY ASSISTANCE FOR NEEDY FAMILIES (TANF) CASH FLAG - FIRST MONTH	1031	1031	1					
**	RESTRICTED BENEFITS FLAG GROUP - MONTHLY (OCCURS 12 TIMES)	1043	1054	12					
55.	RESTRICTED BENEFITS FLAG - FIRST MONTH	1043	1043	1					
**	CHILD HEALTH INSURANCE PROGRAM (CHIP) CODE GROUP - MONTHLY (OCCURS 12 TIMES)	1055	1066	12					
56.	CHILD HEALTH INSURANCE PROGRAM (CHIP) CODE - FIRST MONTH	1055	1055	1					
**	MEDICAID WAIVER GROUP - MONTHLY (OCCURS 12 TIMES)	1067	1174	108					
57.	MAX WAIVER TYPE CODE - 1 - FIRST MONTH	1067	1067	1					
58.	WAIVER ID - 1 - FIRST MONTH	1068	1069	2			1	2	Forgo linking in Level 2
59.	MAX WAIVER TYPE CODE - 2 - FIRST MONTH	1070	1070	1					
60.	WAIVER ID - 2 - FIRST MONTH	1071	1072	2			1	2	Forgo linking in Level 2
61.	MAX WAIVER TYPE CODE - 3 - FIRST MONTH	1073	1073	1					
62.	WAIVER ID - 3 - FIRST MONTH	1074	1075	2			1	2	Forgo linking in Level 2
63.	ANNUAL 1915C MAX WAIVER TYPE - MOST RECENT	1175	1175	1					
***	UTILIZATION AND PAYMENT SUMMARY/REGION	1176	2895	1720					

Table B.1. MAX 2007 PS File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
64.	RECIPIENT INDICATOR	1176	1176	1					
**	INPATIENT HOSPITAL UTILIZATION SUMMARY	1177	1194	18					
65.	TOTAL INPATIENT DISCHARGE COUNT	1177	1179	3			1	3	Forgo claims detail in Level 2
66.	TOTAL INPATIENT STAY COUNT	1180	1182	3			1	3	Forgo claims detail in Level 2
67.	TOTAL INPATIENT LENGTH OF STAY (LOS), IN DAYS (FOR DISCHARGES)	1183	1185	3			1	3	Forgo claims detail in Level 2
68.	TOTAL INPATIENT LENGTH OF STAY (LOS), IN DAYS (FOR STAYS)	1186	1188	3			1	3	Forgo claims detail in Level 2
69.	TOTAL INPATIENT COVERED DAY COUNT (FOR DISCHARGE S)	1189	1191	3			1	3	Forgo claims detail in Level 2
70.	TOTAL INPATIENT COVERED DAY COUNT (FOR STAYS)	1192	1194	3			1	3	Forgo claims detail in Level 2
**	INSTITUTIONAL LONG TERM CARE UTILIZATION SUMMARY GROUP	1195	1209	15					
71.	LONG TERM CARE MENTAL HOSPITAL FOR THE AGED COVERED DAY COUNT	1195	1197	3			1	3	Forgo claims detail in Level 2
72.	LONG TERM CARE INPATIENT PSYCHIATRIC FACILITY (AGE < 21) COVERED DAY COUNT	1198	1200	3			1	3	Forgo claims detail in Level 2
73.	INTERMEDIATE CARE FACILITY FOR THE MENTALLY RETARDED - ICF-MR COVERED DAY COUNT	1201	1203	3			1	3	Forgo claims detail in Level 2
74.	NURSING FACILITY - NF - COVERED DAY COUNT	1204	1206	3			1	3	Forgo claims detail in Level 2
75.	LONG TERM CARE COVERED DAY COUNT	1207	1209	3			1	3	Forgo claims detail in Level 2
**	CLAIMS PAYMENT SUMMARY GROUP	1210	1269	60					
76.	TOTAL MEDICAID RECORD COUNT	1210	1214	5	1	5			Questionable utility
77.	TOTAL MEDICAID FEE-FOR-SERVICE CLAIM COUNT	1215	1219	5	1	5			Questionable utility
78.	TOTAL MEDICAID PRE-PAID PLAN PREMIUM PAYMENT RECORD COUNT	1220	1224	5	1	5			Questionable utility
79.	TOTAL MEDICAID ENCOUNTER RECORD COUNT	1225	1229	5	1	5			Questionable utility
80.	TOTAL MEDICAID PAYMENT AMOUNT	1230	1237	8					
81.	TOTAL MEDICAID FEE-FOR-SERVICE PAYMENT AMOUNT	1238	1245	8					
82.	TOTAL MEDICAID PRE-PAID PLAN PREMIUM PAYMENT AMOUNT	1246	1253	8					
83.	TOTAL MEDICAID CHARGE AMOUNT	1254	1261	8	1	8			Inconsistent reporting
84.	TOTAL THIRD PARTY PAYMENT AMOUNT	1262	1269	8	1	8			Inconsistent reporting
**	PROGRAM TYPE SUMMARY GROUP - TYPE OF PROGRAM 2 - 7 (OCCURS 6 TIMES)	1270	1599	330					
85.	INPATIENT HOSPITAL RECORDS - FIRST TYPE OF PROGRAM	1270	1272	3	6	18			Questionable utility
86.	INPATIENT HOSPITAL PAYMENTS - FIRST TYPE OF PROGRAM	1273	1280	8					
87.	INSTITUTIONAL LONG TERM CARE RECORDS - FIRST TYPE OF PROGRAM	1281	1283	3	6	18			Questionable utility
88.	INSTITUTIONAL LONG TERM CARE PAYMENTS - FIRST TYPE OF PROGRAM	1284	1291	8					
89.	OTHER SERVICE RECORDS - FIRST TYPE OF PROGRAM	1292	1294	3	6	18			Questionable utility
90.	OTHER SERVICE PAYMENTS - FIRST TYPE OF PROGRAM	1295	1302	8					
91.	PRESCRIPTION DRUG RECORDS - FIRST TYPE OF PROGRAM	1303	1305	3	6	18			Questionable utility
92.	PRESCRIPTION DRUG PAYMENTS - FIRST TYPE OF PROGRAM	1306	1313	8					
93.	TOTAL RECORDS - FIRST TYPE OF PROGRAM	1314	1316	3	6	18			Questionable utility
94.	TOTAL PAYMENTS - FIRST TYPE OF PROGRAM	1317	1324	8					
95.	DELIVERY CODE	1600	1600	1					
**	TYPE OF SERVICE GROUP - MAX TOS (OCCURS 31 TIMES)	1601	2685	1085					
96.	RECIPIENT INDICATOR - FIRST MAX TOS	1601	1601	1	31	31			Computable from \$
97.	FEE-FOR-SERVICE CLAIM COUNT - FIRST MAX TOS	1602	1606	5	31	155			Not critical for PS analyses
98.	FEE-FOR-SERVICE MEDICAID PAYMENT AMOUNT - FIRST MAX TOS	1607	1614	8					

Table B.1. MAX 2007 PS File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
99.	FEE-FOR-SERVICE CHARGE AMOUNT - FIRST MAX TOS	1615	1622	8	31	248			Inconsistent reporting
100.	FEE-FOR-SERVICE THIRD PARTY PAYMENT AMOUNT - FIRST MAX TOS	1623	1630	8	31	248			Inconsistent reporting
101.	ENCOUNTER RECORD COUNT - FIRST MAX TOS	1631	1635	5					
**	MEDICAID COMMUNITY-BASED LONG-TERM CARE (CLTC) PAYMENT SUMMARY GROUP - CLTC (OCCURS 21 TIMES BY CLTC INDICATOR FLAG)	2686	2853	168					
102.	MEDICAID PAYMENT AMOUNT - FIRST TYPE OF CLTC	2686	2693	8					
**	PREMIUM PAYMENT GROUP (OCCURS 3 TIMES, ONE FOR EACH MAX TYPE OF PREMIUM TOS = 20 - 22)	2854	2895	42					
103.	PREMIUM PAYMENT INDICATOR - FIRST TYPE OF PREMIUM	2854	2854	1	3	3			Computable from record count
104.	PREMIUM PAYMENT RECORD COUNT - FIRST TYPE OF PREMIUM	2855	2859	5					
105.	MEDICAID PREMIUM PAYMENT AMOUNT - FIRST TYPE OF PREMIUM	2860	2867	8					

Table B.2. MAX 2007 OT File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
***	MEDICAID ANALYTIC EXTRACT OTHER SERVICES RECORD	1	265	265					
***	MEDICAID ELIGIBILITY VARIABLES	1	79	79					
1.	MSIS IDENTIFICATION NUMBER	1	20	20					
2.	STATE ABBREVIATION CODE	21	22	2					
3.	SOCIAL SECURITY NUMBER - FROM MSIS	23	31	9			1	9	Forgo linking in level 2
4.	MEDICARE HEALTH INSURANCE CLAIM (HIC) NUMBER - FROM MSIS	32	43	12			1	12	Forgo linking in level 2
5.	BIRTH DATE	44	51	8			1	8	Forgo linking in level 2
6.	SEX CODE	52	52	1			1	1	Forgo linking in level 2
7.	RACE/ETHNICITY CODE	53	53	1	1	1			Available in PS file
8.	RACE - WHITE	54	54	1	1	1			Available in PS file
9.	RACE - BLACK/AFRICAN AMERICAN	55	55	1	1	1			Available in PS file
10.	RACE - AMERICAN INDIAN/ALASKAN NATIVE	56	56	1	1	1			Available in PS file
11.	RACE - ASIAN	57	57	1	1	1			Available in PS file
12.	RACE - NATIVE HAWAIIAN/OTHER PACIFIC ISLANDER	58	58	1	1	1			Available in PS file
13.	ETHNICITY - HISPANIC OR LATINO	59	59	1	1	1			Available in PS file
14.	STATE SPECIFIC ELIGIBILITY CODE - MOST RECENT	60	65	6	1	6			Available in PS file
15.	STATE SPECIFIC ELIGIBILITY CODE - FOR MONTH OF SERVICE	66	71	6	1	6			Available in PS file
16.	MAX UNIFORM ELIGIBILITY CODE - MOST RECENT	72	73	2	1	2			Available in PS file
17.	MAX UNIFORM ELIGIBILITY CODE - FOR MONTH OF SERVICE	74	75	2	1	2			Available in PS file
18.	MISSING ELIGIBILITY DATA	76	76	1	1	1			Available in PS file
19.	MEDICARE DUAL CODE - CLAIM-BASED	77	77	1	1	1			Available in PS file
20.	MEDICARE DUAL CODE - ANNUAL	78	79	2	1	2			Available in PS file
***	UTILIZATION AND PAYMENT SUMMARY REGION	80	265	186					
**	SERVICE GROUP	80	122	43					
21.	MSIS TYPE OF SERVICE CODE	80	81	2					
22.	MSIS TYPE OF PROGRAM CODE	82	82	1					
23.	MAX TYPE OF SERVICE CODE	83	84	2					
24.	COMMUNITY-BASED LONG-TERM CARE (GLTC) FLAG	85	86	2					
25.	BILLING PROVIDER IDENTIFICATION NUMBER	87	98	12			1	12	Forgo linking in level 2
26.	NATIONAL PROVIDER IDENTIFIER	99	110	12			1	12	Forgo linking in level 2
27.	PROVIDER TAXONOMY	111	122	12					
**	CLAIMS AND PAYMENT GROUP	123	194	72					
28.	TYPE OF CLAIM CODE	123	123	1					
29.	ADJUSTMENT CODE	124	124	1	1	1			Questionable utility
30.	MANAGED CARE TYPE OF PLAN CODE	125	126	2					
31.	MANAGED CARE PLAN IDENTIFICATION NUMBER	127	138	12			1	12	Forgo linking in level 2
32.	MEDICAID PAYMENT AMOUNT	139	146	8					
33.	THIRD PARTY PAYMENT AMOUNT	147	154	8	1	8			Inconsistent reporting
34.	PAYMENT DATE	155	162	8					
35.	CHARGE AMOUNT	163	170	8	1	8			Inconsistent reporting
36.	PREPAID PLAN SERVICE VALUE	171	178	8					
37.	MEDICARE COINSURANCE PAYMENT AMOUNT	179	186	8					
38.	MEDICARE DEDUCTIBLE PAYMENT AMOUNT	187	194	8					
**	OTHER SERVICES GROUP	195	265	71					

Table B.2. MAX 2007 OT File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
39.	SERVICE BEGINNING DATE	195	202	8					
40.	ENDING DATE OF SERVICE	203	210	8					
41.	PROCEDURE CODING SYSTEM CODE	211	212	2					
42.	PROCEDURE (SERVICE) CODE	213	220	8					
43.	PROCEDURE (SERVICE) MODIFIER CODE	221	222	2					
44.	DIAGNOSIS CODE-1	223	230	8					
45.	DIAGNOSIS CODE-2	231	238	8					
46.	QUANTITY OF SERVICE	239	243	5					
47.	SERVICING PROVIDER IDENTIFICATION NUMBER	244	255	12			1	12	Forgo linking in level 2
48.	SERVICING PROVIDER SPECIALTY CODE	256	259	4					
49.	PLACE OF SERVICE CODE	260	261	2					
50.	UB-92 REVENUE CODE	262	265	4					

Table B.3. MAX 2007 RX File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
***	MEDICAID ANALYTIC EXTRACT DRUG RECORD	1	348	348					
***	MEDICAID ELIGIBILITY VARIABLES	1	78	78					
1.	MSIS IDENTIFICATION NUMBER	1	20	20					
2.	STATE ABBREVIATION CODE	21	22	2					
3.	SOCIAL SECURITY NUMBER - FROM MSIS	23	31	9			1	9	Forgo linking in level 2
4.	MEDICARE HEALTH INSURANCE CLAIM (HIC) NUMBER - FROM MSIS	32	43	12			1	12	Forgo linking in level 2
5.	BIRTH DATE	44	51	8			1	8	Forgo linking in level 2
6.	SEX CODE	52	52	1			1	1	Forgo linking in level 2
7.	RACE/ETHNICITY CODE	53	53	1	1	1			Available in PS file
8.	RACE - WHITE	54	54	1	1	1			Available in PS file
9.	RACE - BLACK/AFRICAN AMERICAN	55	55	1	1	1			Available in PS file
10.	RACE - AMERICAN INDIAN/ALASKA NATIVE	56	56	1	1	1			Available in PS file
11.	RACE - ASIAN	57	57	1	1	1			Available in PS file
12.	RACE - NATIVE HAWAIIAN/OTHER PACIFIC ISLANDER	58	58	1	1	1			Available in PS file
13.	ETHNICITY - HISPANIC OR LATINO	59	59	1	1	1			Available in PS file
14.	STATE SPECIFIC ELIGIBILITY CODE - MOST RECENT	60	65	6	1	6			Available in PS file
15.	STATE SPECIFIC ELIGIBILITY CODE - FOR MONTH OF SERVICE	66	71	6	1	6			Available in PS file
16.	MAX UNIFORM ELIGIBILITY CODE - MOST RECENT	72	73	2	1	2			Available in PS file
17.	MAX UNIFORM ELIGIBILITY CODE - FOR MONTH OF SERVICE	74	75	2	1	2			Available in PS file
18.	MISSING ELIGIBILITY DATA	76	76	1	1	1			Available in PS file
19.	MEDICARE DUAL CODE - ANNUAL	77	78	2	1	2			Available in PS file
***	UTILIZATION AND PAYMENT SUMMARY REGION	79	348	270					
**	SERVICE GROUP	79	119	41					
20.	MSIS TYPE OF SERVICE CODE	79	80	2					
21.	MSIS TYPE OF PROGRAM CODE	81	81	1					
22.	MAX TYPE OF SERVICE CODE	82	83	2					
23.	BILLING PROVIDER IDENTIFICATION NUMBER	84	95	12			1	12	Forgo linking in level 2
24.	NATIONAL PROVIDER IDENTIFIER	96	107	12			1	12	Forgo linking in level 2
25.	PROVIDER TAXONOMY	108	119	12					
**	CLAIMS AND PAYMENT GROUP	120	191	72					
26.	TYPE OF CLAIM CODE	120	120	1					
27.	ADJUSTMENT CODE	121	121	1	1	1			Questionable utility
28.	MANAGED CARE TYPE OF PLAN CODE	122	123	2					
29.	MANAGED CARE PLAN IDENTIFICATION NUMBER	124	135	12			1	12	Forgo linking in level 2
30.	MEDICAID PAYMENT AMOUNT	136	143	8					
31.	THIRD PARTY PAYMENT AMOUNT	144	151	8	1	8			Inconsistent reporting
32.	PAYMENT DATE	152	159	8					
33.	CHARGE AMOUNT	160	167	8	1	8			Inconsistent reporting
34.	PREPAID PLAN SERVICE VALUE	168	175	8					
35.	FILLER	176	183	8	1	8			Filler
36.	FILLER	184	191	8	1	8			Filler
**	PRESCRIPTION DRUG GROUP	192	241	50					
37.	PRESCRIBING PHYSICIAN IDENTIFICATION NUMBER	192	203	12					

Table B.3. MAX 2007 RX File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
38.	PRESCRIBED DATE	204	211	8					
39.	PRESCRIPTION FILLED DATE	212	219	8					
40.	NEW OR REFILL INDICATOR	220	221	2	1	2			Unreliable
41.	NATIONAL DRUG CODE (NDC)	222	233	12					
42.	QUANTITY OF SERVICE	234	238	5					
43.	DAYS SUPPLY	239	241	3					
*	FIRST DATA BANK/MEDISPAN GROUP (PROPRIETARY - ACCESS LIMITED TO LICENSE HOLDERS)	242	348	107					
44.	NATIONAL DRUG CODE FORMAT INDICATOR	242	242	1			1	1	Proprietary
45.	DRUG CLASS	243	243	1			1	1	Proprietary
46.	MULTI-SOURCE CODE	244	244	1			1	1	Proprietary
47.	FILLER	245	298	54	1	54			Filler
48.	HIERARCHICAL SPECIFIC THERAPEUTIC CLASS CODE	299	301	3			1	3	Proprietary
49.	THERAPEUTIC CLASS CODE - GENERIC	302	303	2			1	2	Proprietary
50.	FILLER	304	309	6	1	6			Filler
51.	CLINICAL FORMULATION ID	310	315	6			1	6	Proprietary
52.	INGREDIENT LIST IDENTIFIER	316	321	6			1	6	Proprietary
53.	HIERARCHICAL SPECIFIC THERAPEUTIC CLASS CODE SEQUENCE NUMBER	322	327	6			1	6	Proprietary
54.	FILLER	328	333	6	1				Filler
55.	MEDI-SPAN THERAPEUTIC CLASSIFICATION SYSTEM CODE	334	347	14			1	14	Proprietary
56.	OVER-THE-COUNTER INDICATOR CODE	348	348	1			1	1	Proprietary

Table B.4. MAX 2007 LT File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
***	MEDICAID ANALYTIC EXTRACT LONG TERM CARE RECORD	1	281	281					
***	MEDICAID ELIGIBILITY VARIABLES	1	79	79					
1.	MSIS IDENTIFICATION NUMBER	1	20	20					
2.	STATE ABBREVIATION CODE	21	22	2					
3.	SOCIAL SECURITY NUMBER - FROM MSIS	23	31	9			1	9	Forgo linking in level 2
4.	MEDICARE HEALTH INSURANCE CLAIM (HIC) NUMBER - FROM MSIS	32	43	12			1	12	Forgo linking in level 2
5.	BIRTH DATE	44	51	8			1	8	Forgo linking in level 2
6.	SEX CODE	52	52	1			1	1	Forgo linking in level 2
7.	RACE/ETHNICITY CODE	53	53	1	1	1			Available in PS file
8.	RACE - WHITE	54	54	1	1	1			Available in PS file
9.	RACE - BLACK/AFRICAN AMERICAN	55	55	1	1	1			Available in PS file
10.	RACE - AMERICAN INDIAN/ALASKA NATIVE	56	56	1	1	1			Available in PS file
11.	RACE - ASIAN	57	57	1	1	1			Available in PS file
12.	RACE - NATIVE HAWAIIAN/ OTHER PACIFIC ISLANDER	58	58	1	1	1			Available in PS file
13.	ETHNICITY - HISPANIC OR LATINO	59	59	1	1	1			Available in PS file
14.	STATE SPECIFIC ELIGIBILITY CODE - MOST RECENT	60	65	6	1	6			Available in PS file
15.	STATE SPECIFIC ELIGIBILITY CODE - FOR MONTH OF SERVICE	66	71	6	1	6			Available in PS file
16.	MAX UNIFORM ELIGIBILITY CODE - MOST RECENT	72	73	2	1	2			Available in PS file
17.	MAX UNIFORM ELIGIBILITY CODE - FOR MONTH OF SERVICE	74	75	2	1	2			Available in PS file
18.	MISSING ELIGIBILITY DATA	76	76	1	1	1			Available in PS file
19.	MEDICARE DUAL CODE - CLAIM-BASED	77	77	1	1	1			Available in PS file
20.	MEDICARE DUAL CODE - ANNUAL	78	79	2	1	2			Available in PS file
***	UTILIZATION AND PAYMENT SUMMARY REGION	80	281	202					
**	SERVICE GROUP	80	120	41					
21.	MSIS TYPE OF SERVICE CODE	80	81	2					
22.	MSIS TYPE OF PROGRAM CODE	82	82	1					
23.	MAX TYPE OF SERVICE CODE	83	84	2					
24.	BILLING PROVIDER IDENTIFICATION NUMBER	85	96	12			1	12	Forgo linking in level 2
25.	NATIONAL PROVIDER IDENTIFIER	97	108	12			1	12	Forgo linking in level 2
26.	PROVIDER TAXONOMY	109	120	12					
**	CLAIMS AND PAYMENT GROUP	121	192	72					
27.	TYPE OF CLAIM CODE	121	121	1					
28.	ADJUSTMENT CODE	122	122	1	1	1			Questionable utility
29.	MANAGED CARE TYPE OF PLAN CODE	123	124	2					
30.	MANAGED CARE PLAN IDENTIFICATION NUMBER	125	136	12			1	12	Forgo linking in level 2
31.	MEDICAID PAYMENT AMOUNT	137	144	8					
32.	THIRD PARTY PAYMENT AMOUNT	145	152	8	1	8			Inconsistent reporting
33.	PAYMENT DATE	153	160	8					
34.	CHARGE AMOUNT	161	168	8	1	8			Inconsistent reporting
35.	PREPAID PLAN SERVICE VALUE	169	176	8					
36.	MEDICARE COINSURANCE PAYMENT AMOUNT	177	184	8					
37.	MEDICARE DEDUCTIBLE PAYMENT AMOUNT	185	192	8					
**	INSTITUTIONAL LONG TERM CARE GROUP	193	281	89					

Table B.4. MAX 2007 LT File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
38.	INSTITUTIONAL LONG TERM CARE ADMISSION DATE	193	200	8					
39.	SERVICE BEGINNING DATE	201	208	8					
40.	ENDING DATE OF SERVICE	209	216	8					
*	DIAGNOSIS CODE GROUP (OCCURS 5 TIMES)	217	256	40					
41.	DIAGNOSIS CODE - FIRST DIAGNOSIS	217	224	8					
42.	MENTAL HOSPITAL FOR THE AGED DAY COUNT	257	259	3					
43.	INPATIENT PSYCHIATRIC FACILITY (AGE < 21) DAY COUNT	260	262	3					
44.	INTERMEDIATE CARE FACILITY FOR THE MENTALLY RETARDED DAY COUNT	263	265	3					
45.	NURSING FACILITY DAY COUNT	266	268	3					
46.	LONG TERM CARE LEAVE DAY COUNT	269	271	3					
47.	PATIENT STATUS CODE	272	273	2					
48.	PATIENT LIABILITY AMOUNT	274	281	8					

Table B.5. MAX 2007 IP File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
***	MEDICAID ANALYTIC EXTRACT INPATIENT RECORD	1	807	807					
***	MEDICAID ELIGIBILITY VARIABLES	1	79	79					
1.	MSIS IDENTIFICATION NUMBER	1	20	20					
2.	STATE ABBREVIATION CODE	21	22	2					
3.	SOCIAL SECURITY NUMBER - FROM MSIS	23	31	9			1	9	Forgo linking in level 2
4.	MEDICARE HEALTH INSURANCE CLAIM (HIC) NUMBER - FROM MSIS	32	43	12			1	12	Forgo linking in level 2
5.	BIRTH DATE	44	51	8			1	8	Forgo linking in level 2
6.	SEX CODE	52	52	1			1	1	Forgo linking in level 2
7.	RACE/ETHNICITY CODE	53	53	1	1	1			Available in PS file
8.	RACE - WHITE	54	54	1	1	1			Available in PS file
9.	RACE - BLACK/AFRICAN-AMERICAN	55	55	1	1	1			Available in PS file
10.	RACE - AMERICAN INDIAN/ALASKAN NATIVE	56	56	1	1	1			Available in PS file
11.	RACE - ASIAN	57	57	1	1	1			Available in PS file
12.	RACE - NATIVE HAWAIIAN/OTHER PACIFIC ISLANDER	58	58	1	1	1			Available in PS file
13.	ETHNICITY - HISPANIC OR LATINO	59	59	1	1	1			Available in PS file
14.	STATE SPECIFIC ELIGIBILITY CODE - MOST RECENT	60	65	6	1	6			Available in PS file
15.	STATE SPECIFIC ELIGIBILITY CODE - FOR MONTH OF SERVICE	66	71	6	1	6			Available in PS file
16.	MAX UNIFORM ELIGIBILITY CODE - MOST RECENT	72	73	2	1	2			Available in PS file
17.	MAX UNIFORM ELIGIBILITY CODE - FOR MONTH OF SERVICE	74	75	2	1	2			Available in PS file
18.	MISSING ELIGIBILITY DATA	76	76	1	1	1			Available in PS file
19.	MEDICARE DUAL CODE - CLAIM-BASED	77	77	1	1	1			Available in PS file
20.	MEDICARE DUAL CODE - ANNUAL	78	79	2	1	2			Available in PS file
***	UTILIZATION AND PAYMENT SUMMARY REGION	80	807	728					
**	SERVICE GROUP	80	120	41					
21.	MSIS TYPE OF SERVICE CODE	80	81	2					
22.	MSIS TYPE OF PROGRAM CODE	82	82	1					
23.	MAX TYPE OF SERVICE CODE	83	84	2					
24.	BILLING PROVIDER IDENTIFICATION NUMBER	85	96	12			1	12	Forgo linking in level 2
25.	NATIONAL PROVIDER IDENTIFIER	97	108	12			1	12	Forgo linking in level 2
26.	PROVIDER TAXONOMY	109	120	12					
**	CLAIMS AND PAYMENT GROUP	121	192	72					
27.	TYPE OF CLAIM CODE	121	121	1					
28.	ADJUSTMENT CODE	122	122	1	1	1			Questionable utility
29.	MANAGED CARE TYPE OF PLAN CODE	123	124	2					
30.	MANAGED CARE PLAN IDENTIFICATION NUMBER	125	136	12			1	12	Forgo linking in level 2
31.	MEDICAID PAYMENT AMOUNT	137	144	8					
32.	THIRD PARTY PAYMENT AMOUNT	145	152	8	1	8			Inconsistent reporting
33.	PAYMENT DATE	153	160	8					
34.	CHARGE AMOUNT	161	168	8	1	8			Inconsistent reporting
35.	PREPAID PLAN SERVICE VALUE	169	176	8					
36.	MEDICARE COINSURANCE PAYMENT AMOUNT	177	184	8					
37.	MEDICARE DEDUCTIBLE PAYMENT AMOUNT	185	192	8					
**	INPATIENT HOSPITAL GROUP	193	807	615					

Table B.5. MAX 2007 IP File Record Layout and Potential Variables to Be Excluded from Mini-MAX

ELEMENT NUMBER	ELEMENT NAME	BEG	END	LENGTH	Exclude from Skinny File Level 1	Length Excluded	Exclude from Skinny File Level 2	Length Excluded	Reason
38.	ADMISSION DATE	193	200	8					
39.	SERVICE BEGINNING DATE	201	208	8					
40.	ENDING DATE OF SERVICE	209	216	8					
41.	PRINCIPAL DIAGNOSIS CODE	217	224	8					
*	DIAGNOSIS CODE GROUP - ADDITIONAL DIAGNOSIS 2 - 9 (OCCURS 8 TIMES)	225	288	64					
42.	DIAGNOSIS CODE-2	225	232	8					
43.	PRINCIPAL PROCEDURE DATE	289	296	8					
44.	PROCEDURE CODING SYSTEM CODE - PRINCIPAL	297	298	2					
45.	PROCEDURE CODE - PRINCIPAL	299	306	8					
*	PROCEDURE CODE GROUP - ADDITIONAL PROCEDURES 2 - 6 (OCCURS 5 TIMES)	307	356	50					
46.	PROCEDURE CODING SYSTEM CODE - ADDITIONAL PROCEDURES	307	308	2					
47.	PROCEDURE CODE - ADDITIONAL PROCEDURES	309	316	8					
48.	DELIVERY CODE	357	357	1					
49.	MEDICAID COVERED INPATIENT DAYS	358	360	3					
50.	PATIENT STATUS CODE	361	362	2					
51.	DIAGNOSIS RELATED GROUP INDICATOR	363	366	4					
52.	DIAGNOSIS RELATED GROUP	367	370	4					
*	UB-92 REVENUE CODE GROUP (OCCURS 23 TIMES)	371	807	437					
53.	UB-92 REVENUE CODE - FIRST REVENUE CODE	371	374	4					
54.	UB-92 REVENUE CODE CHARGE - FIRST REVENUE CODE	375	382	8					
55.	UB-92 REVENUE CODE UNITS - FIRST REVENUE CODE	383	389	7					

## **APPENDIX C**

### **TEP SURVEY INSTRUMENT AND SUMMARY OF RESULTS**

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

**National Medicaid Sample - Mini-MAX**

**SURVEY INSTRUMENT AND SUMMARY OF TECHNICAL EXPERT PANEL MEMBER RESPONSES**

**1. Assuming Mini-MAX will be produced and will be restricted-access/research-identifiable files (DUA and privacy board review required) containing a 5 percent sample of annual Medicaid enrollees, which of the following components should or should not be included as part of Mini-MAX?**

(Please mark an "X" in the appropriate box to the right of each entry and its sub-elements.)	Critical (Do Not Release Without This Component)	Important	Not Necessary	Exclude This Component from Mini-MAX	No Opinion	Total # of Responses
1. A user's guide that strives to reduce potential misuse of the data	5	1				6
a. Includes reporting anomalies across states	3	2				5
b. Summarizes program differences across states	2	1	2			5
c. Includes guide for applying file weights and calculating standard errors	3	2				5
d. Contains examples of research uses that cannot be supported by the file	1	4				5
e. Includes a table of coefficients of variation		3	1		1	5
f. Other (specify) _____ Metadata _____	1					1
2. Timely file availability		1				1
a. Two year lag (one to two months after MAX)	1	3		1		5
b. 16-month lag (may require use of incomplete files)	1	3	1			5
c. Other (specify) _____						0
3. A simpler file structure (one Person Summary file, one or two files for each claim type)		5				5
a. The OT ("other") claims file (the largest file) split into two or three files by service type, each < 6 GB		3				3
4. Over-sampling of small subpopulations (note that researchers can currently request data extracts for these populations)		2				2
a. Infants		4	1			5
b. Foster-care children		3	1			4
c. Dual Medicare and Medicaid enrollees		5				5
d. Other (specify) _____						
e. Other (specify) _____						
5. Stratified sampling by state with state sample sizes sufficient to enable state-level estimates	3	3				6

(Table 1 continued)

(Please mark an "X" in the appropriate box to the right of each entry and its sub-elements.)	Critical (Do Not Release Without This Component)	Important	Not Necessary	Exclude This Component from Mini-MAX	No Opinion	Total # of Responses
6. Stratified sampling by state and basis of eligibility (BOE - aged, disabled, child, adult) (oversampling smaller BOE groups would improve estimates for the aged and disabled but would reduce the size of child and adult subgroups such as foster children and infants unless they are oversampled separately)	1	1	2		1	5
7. Longitudinal sample component		2	1		2	5
8. "Slim" version of files	1	1				2
a. Option that includes all MAX variables			4		1	5
b. Option to exclude unreliable/duplicative information (version 1 in Oct 19th handout)	1	3	1			6
c. Option that also excludes linking/identifier variables (version 2 in Oct 19th handout)		2	2	1	1	6
d. Option with aggregated/summarized data with substantial loss of information (not yet specified)			2	1	1	4

88

**2. Provided that all components you identified as "Critical" above are included in the file (and those listed as "Exclude" are excluded), do you support the release of Mini-MAX by CMS?**

# of No Responses	# of Don't Know Responses	# of Yes Responses	Total Number of Responses
		3	3

**Please explain:**

One respondent suggested that Mini-MAX would be particularly useful for existing MAX users to conduct exploratory work to determine whether to conduct full-fledged MAX analyses. Another expressed interest in seeing the data used to a greater extent.

**3. Do you have any other concerns about the release of Mini-MAX or recommendations for the Mini-MAX files?**

---

**MATHEMATICA**  
Policy Research, Inc.

---

[www.mathematica-mpr.com](http://www.mathematica-mpr.com)

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research