MINI-SENTINEL DATA QUALITY AND CHARACTERIZATION PROCEDURES AND FINDINGS REPORT

VERSION 1.0

September 2011

Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the Sentinel Initiative, a multi-faceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. Mini-Sentinel Collaborators include Data and Academic Partners that provide access to health care data and ongoing scientific, technical, methodological, and organizational expertise. The Mini-Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.

Mini-Sentinel Data Quality and Characterization Procedures And Findings Report

Version 1.0

I. PURPOSE 3

A. SCOPE 3
B. ASSUMPTIONS AND DEPENDENCIES 3

II. DATA QUALITY CHECKS 4

A. LEVELS 1 AND 2 DATA QUALITY CHECKS 4
B. LEVEL 3 DATA QUALITY PROFILING 4
C. LEVEL 4 DATA PROFILING 5

III. ERROR REPORTING 5

A. ERROR CONDITIONS AND CODES 5
B. SUMMARY DATA QUALITY REPORT 7

IV. APPENDICES 9

A. APPENDIX A: LEVEL ONE DATA CHECK REQUIREMENTS 9
1. Table: Enrollment 9
2. Table: Demographic 10
3. Table: Dispensing 10
4. Table: Encounter 11
5. Table: Diagnosis 13
6. Table: Procedure 14
7. Table: Death 15
8. Table: Cause of Death 16
B. APPENDIX B: LEVEL TWO: DATA CHECK REQUIREMENTS 17
1. Table: Enrollment 17
2. Table: Demographic 17
3. Table: Dispensing 17
4. Table: Encounter 18
5. Table: Diagnosis 18
6. Table: Procedure 19
7. Table: Death 19
8. Table: Cause of Death 19
C. APPENDIX C: LEVEL THREE: DATA QUALITY PROFILING 20
1. Table: Enrollment 20
2. Table: Demographic 20
3. Table: Dispensing 21
4. Table Encounter 21
5. Table: Diagnosis 23
6. Table: Procedure 24
7. Table: Death 25
8. Table: Cause of Death 25
D. APPENDIX D. MINI-SENTINEL DATA QUALITY CHECKING SUMMARY 26
E. APPENDIX E. MINI-SENTINEL DATA CHARACTERIZATION REPORT 34

I. PURPOSE

This document describes the approach used during Year 1 of the Mini-Sentinel project by the Mini-Sentinel Operations center (MSOC) for data characterization and quality checking of the Mini-Sentinel Distributed Database (MSDD). The initial extract of the MSDD contained administrative and claims data from participating Mini-Sentinel Data Partners; additional data types will be added in subsequent years. To create the MSDD, each Data Partner transformed their local source data into the Mini-Sentinel Common Data Model (MSCDM) format.

The MSOC worked closely with each Data Partner to ensure the MSDD conforms to the MSCDM and to understand their data resources. To guarantee that the MSDD data meets expectations needed for a distributed health data network, the MSOC created a set of standard and fully tested Data Characterization and Quality Review programs to assess the data in the MSDD. The design and the scope of these programs take into account the following considerations:

- Mini-Sentinel Data Partners have a variety of disparate sources and local business rules and procedures for using their data, and each Data Partner must follow their own unique process for populating the MSDD using their own extract, transform, and load (ETL) procedure.

- The Mini-Sentinel distributed analytic approach requires that the MSDD files match the defined MSCDM requirements.

Data checking and characterization is an ongoing process that is undertaken by the MSOC with respect to the entire MSDD, and also a requirement for specific queries and evaluations undertaken by Mini-Sentinel investigators.

The rest of this document defines the Data Characterization and Quality Review process, the specific checks performed, and how inconsistencies with the MSCDM were identified.

A. SCOPE

The main driver for the Data Characterization and Quality Review process is to ensure that the MSDD conforms to the MSCDM and that the data included in the MSDD meet reasonable standards for data transformation consistency and quality (e.g., no missing months, reasonable trends in utilization and enrollment, anomalies explained by Data Partners). To evaluate data characteristics and quality, each Data Partner executed distributed code developed by the MSOC and returned aggregated results to the MSOC. Some files generated by the distributed code - files that contained either proprietary or confidential information - were kept by the Data Partners for review and to help investigate anomalies or questions posed by the MSOC. MSOC reviewed results within and across Data Partners and examined findings to determine if MSCDM and MSDD requirements have been met.

B. ASSUMPTIONS AND DEPENDENCIES

The Data Characterization and Quality Review programs were executed after completion of the initial Mini-Sentinel ETL, and for each subsequent update of the MSDD. The evaluation is based on the degree

to which the MSDD conforms to the MSCDM and an assessment of overall data quality. The MSOC distributed code executes against the MSDD held by each Data Partner.

II. DATA QUALITY CHECKS

The Data Characterization and Quality Review programs were written by MSOC staff and distributed to the Data Partners to be run on the MSDD.

The Data Characterization and Quality Review activities for the first extract are organized into three levels, based on the type of checks being performed. A fourth level of data quality is planned and described in more detail in this document.

A. LEVELS 1 AND 2 DATA QUALITY CHECKS

- **Level 1 –** Reviews completeness and content of each variable, within each file in the MSDD, to ensure that the required variables are populated and the content and formats are consistent with the MSCDM data dictionary.

- **Level 2 –** Reviews completeness and integrity between variables within a table, or variables between tables, to identify possible inconsistencies across variables or tables.

Level 1 and 2 data quality checks generate a Check Summary table that is sent to MSOC for review. After the first extract, MSOC staff developed data checking result standardization programs to organize the data for comparison and highlight anomalies for further review. The Check Summary reports and standardized comparison data also were manually inspected by MSOC staff. All identified data anomalies were reported to the Data Partners. A detailed listing of data fields included in the Level 1 and 2 data checks are provided in Appendices A and B below.

B. LEVEL 3 DATA QUALITY PROFILING

Level 3 data profiling checks are intended to provide high-level qualitative and quantitative reviews of the data. The data profiling output is a collection of tables containing either frequencies and cross frequencies of most of the data fields of the MSCDM (e.g., frequency of procedure code types and procedure code values) or specific data aggregates (e.g., mean, median, min, max number of dispensing per patient by year).

These tables were shared with MSOC for quality assurance purposes (e.g., conformity with MSCDM expectations) and also to analyze patterns, trends and characteristics of the data fields. The Level 3 output tables are used to determine variability within each data field and confirm whether it meets expectations. For example, monthly counts of total number of dispensings (dispensing table) can help identify an unusual peak associated with duplicated data entries. Similarly, shifts in the yearly distribution of encounter types (encounter table) can reveal inconsistencies in the way the underlying have been transformed to MSCDM format over the years. A detailed listing of data fields included in the Level 3 is described in Appendix C below.

C. LEVEL 4 DATA PROFILING

Additional data checks are planned (Level 4) that will consist of a set of more targeted data analyses to further investigate data characterization within and across Data Partners. Examples of Level 4 checks include:

- Members with drug coverage have higher rate of dispensing than those without
- Trends in dispensing rates of specific medications by age and sex
- Frequency of pregnancy codes for men and women
- Frequency with prostate cancer codes for men and women
- Prevalence and incidence rate of major disease categories (e.g., asthma, COPD, diabetes, depression, anxiety, obesity, cancer) by age and sex
- Frequency of specific procedures by age and sex over time (e.g., hip replacement, Coronary Artery Bypass Graft (CABG))

III. ERROR REPORTING

A. ERROR CONDITIONS AND CODES

For all levels of data checks and profiling, potential data errors (e.g., nonconformity within MSCDM), anomalies, and unusual patterns (e.g., large change in enrollment) were flagged as a "warning" for discussion and resolution with Data Partners.

Rules and thresholds for warnings were determined on a field by field. Data quality error statuses include the following:

Status	Status Description
Passed	Data quality check ran without any errors.
Failed	Data quality check ran with fatal errors. Issue must be resolved.
Warning	Data quality check ran without fatal error but minor issues. No action items.

All violations of Level 1 or Level 2 checks generated a warning for that specific check identified by the error code listed in Appendix A and Appendix B. Because Level 3 checks (appendix C) are not binary (pass or fail), checks were assigned a warning status (identified by the error check code in the Appendix) based on manual review and comparison within and across Data Partners. For example, although members over 120 years of age or dispensings with over 100 days supplied are not unexpected and are not a violation of the data model, the number and percent of those occurrences were reported for review. The MSOC reports all warnings to the Data Partners for review and resolution. Resolution can range from documentation of the "anomaly" as valid to requiring a new extract of the files(s) as needed.

The tables from Appendices A, B, and C below lists error codes and the error code descriptions. A full list of errors identified and communicated with the Data Partners within in the MSDD is in Appendix D. The tables in Appendix D present the alerts that were generated, the number of Data Partners that generated the alert, and MSOC comment regarding the alert. An alert does not mean there is an error in the data, only that the anomaly requires discussion and adjudication or documentation. The adjudication and documentation process involved several rounds of discussion with each data partner, including written responses for each alert and discussion regarding whether the alert should be documented as a data anomaly found in the source data (e.g., an incorrect date of birth or missing sex) or should be corrected in subsequent data extracts (e.g., duplicate records or values inconsistent with the MSCDM). Please refer to the MSCDM for details regarding the specifications of the variable and acceptable values.

Please note that alerts can be generated by a single instance of an out-of-range value or inconsistency. The error reports are summarized below, by each table in the MSCDM. Additional details on variable level checks are illustrated in Appendix D and Appendix E.:

Enrollment: Eleven different alerts were generated for the Enrollment table, most occurring at only a few Data Partners. Alerts included some records with duplicate information and inconsistent enrollment start and end dates. Other alerts related to the benefit coverage flags that found partners with all members having both medical and drug coverage (which was correct) and a high rate of one flag as “unknown”.

Demographics: Nearly all Data partners had at least 1 person with an age that was out of expected range. Other alerts were generated for duplicate records and having a person (identifier) in the demographic table that was not found in the enrollment table.

Dispensing: Utilization tables such as dispensings are expected to have out of range values and some level of outliers. There were 8 unique alerts generated for the dispensing table. Most sites had at least some dispensings with over 100 days supplied and non-positive values for the dispensed amount. Dispensings with over 100 days supply is not necessarily incorrect, but the rate of such dispensings should be low and consistent across partners. Many partners had some dispensings with non-standard coding for the National Drug Code.

Encounter: There were 25 unique alerts generated for the Encounter Table. The most common alerts were for duplicate records, missing values, and values that did not conform to the MSCDM.

Diagnosis: There were 8 unique alerts generated for the diagnosis table. These included invalid characters in the diagnosis code filed, duplicate records, missingness, unusual trends, and high rate of diagnoses per person. The most common alerts were for duplicate records and the use of the primary diagnosis flag for non-inpatient encounters.

Procedure: There were 8 unique alerts generated for the procedure table. The most common alert was for duplicate records. Other alerts included invalid characters, missingness, unusual trends, and high rate of procedures per person.

Death tables: There were 5 unique alerts generated for the death table. Alerts included missingness, identifiers not found in the enrollment table, and dates of death prior to the begin data for the data extract. The cause of death table had similar alerts.

B. SUMMARY DATA QUALITY REPORT

The MSOC uses the output tables from the Levels 1-3 to create data quality checking and profiling reports for use by MSOC, FDA, and workgroups to help guide understanding of the MSDD.

A summary of the data characterization presented in Appendix E is provided below. The tables represent **uncorrected** output of the initial data checking process. **Many of the anomalies shown in the tables have subsequently been corrected; subsequent reports will reflect those corrections.** *No Mini-sentinel analyses were impacted by these anomalies*.

Enrollment (Table 1, Figure 1-2): In the first data extract there were 62 million unique members in the enrollment table. Table 1 presented enrollment periods per member, types of enrollment, and length of enrollment across Data Partners. The number of enrollment periods per person ranged from 1 to 4, with an overall average of 2. Most enrollment periods had both medical and drug coverage, and this varied across partners. About 45% of members had over 2 years of enrollment available. Interpretation of the rate of short enrollment periods is complicated by right censoring of newly enrolled members.

Demographics (Table 2, Figures 3-17): There were 83 million unique individuals in the demographic table. There are more people in the demographic table than in the enrollment table because some Data Partners have utilization information for non-members. The sex information is consistent across partners, with all sites having slightly more females than males. Age varies across Data Partners, ranging from about 36 years of age to 59 years of age; most partners had an average age between 40 and 50. Few partners have race and ethnicity information.

Dispensing (Tables 3, Figures 18-22): There were 47 million unique individuals in the dispensing table. Table 4 presents dispensings per person and the distribution of days supplied and dispensed amount per dispensing. Information across sites is comparable, with few dispensings having missing, negative or zero values for days supplied and dispensed amount. Dispensings per person per year (among those with at least 1 dispensing) were stable within and across Data Partners. Several anomalies were identified, including partners with large proportion of dispensings for 34 to 66 days and 67 to 99 days. These were verified as correct based on formulary guidelines that promoted dispensing of two and 3 month supplies. A small portion of large values for amount dispensed is expected because the unit of measure can be a volume (e.g., 1000 ml) as compared to a number of pills or tablets. Note that the most recent year of data reported may be incomplete for some Data Partners so the rate of dispensings per user is not comparable with prior years.

Encounter (Table 4, Figures 23-28): There were 49.9 million unique individuals in the encounter table. Encounters per person per year (among those with at least 1 encounter) were stable within and across Data Partners. Ambulatory visits accounted for over 90% of all encounters, followed by emergency department visits and inpatient stays. Some Data partners populated the “Other ambulatory visit” field and others did not, resulting in variation within the ambulatory visit fields. One site had a substantial

number of non-acute institutional stays that can be explained by the underlying age of the population at that Data Partner.

Diagnosis: (Table 5, Figures 29-31): There were 49.2 million unique individuals in the diagnosis table. There were about 2 distinct diagnoses per encounter, with the most diagnoses found in the inpatient setting. All diagnoses were ICD-9-CM codes except one partner that had 2% of codes identified as "other".

Procedure (table 6, Figures 32-34): There were 49.6 million unique individuals in the procedure table. There were about 2 procedures per encounter, with the most procedures found in the inpatient setting. The vast majority of procedure codes were HCPCS Level I and level II. Some partners had a substantial portion of revenue codes. The use of ICD-9 procedure codes – generally used in the inpatient setting - matched the rate of inpatient encounters.

Death tables (Tables 7 and 8, Figures 35-39): There were 4.2 million unique individuals in the death table and 1.6 million unique individuals in the cause of death table. Most information was derived from state death registries, with a mix of other local sources. Confidence in the matching of death information to health plan members was typically excellent or fair.

Cross Table Comparison (Table 9): Table 9 illustrates the proportion of unique members who are identified in the each of the MSDD tables who are also in the enrollment table. For most Data Partners the proportion is over 90%, indicating that most individuals with utilization are found in the enrollment file. Some Data Partners provide care for non-members, meaning that there will be a lower proportion of matches.

The descriptions above are meant as a brief summary of the Year 1 data characterization information. Once again, we note that the data presented are **uncorrected**. Subsequent data characterization reports will reflect corrections to the data extract process.

IV. APPENDICES

A. APPENDIX A: LEVEL ONE DATA CHECK REQUIREMENTS

The tables list the Level One data checks (by table and by variable).

1. Table: Enrollment

	Variable Name	Rule	Error Code
1	PatID	Must be character data type	Enr1.1
	PatID	Must be non-missing	Enr1.2
	PatID	Must be left justified	Enr1.3
2	Enr_Start	Must be a SAS date value of numeric data type	Enr1.4
	Enr_Start	Must be of SAS length 4	Enr1.5
	Enr_Start	Must be non-missing	Enr1.6
	Enr_Start	Earliest permitted date is January 1, 2007; latest permitted date is current date	Enr1.7
3	Enr_End	Must be a SAS date value of numeric data type	Enr1.8
	Enr_End	Must be of SAS length 4	Enr1.9
	Enr_End	Must be non-missing	Enr1.10
	Enr_End	Earliest permitted date is January 1, 2007; latest permitted date is current date	Enr1.11
4	Med_Cov	Must be character data type	Enr1.12
	Med_Cov	Must be exactly 1 character in length	Enr1.13
	Med_Cov	Must be non-missing	Enr1.14
	Med_Cov	Must have the values of either “Y”, “N”, or “U” only	Enr1.15
5	Drug_Cov	Must be character data type	Enr1.16
	Drug_Cov	Must be exactly 1 character in length	Enr1.18
	Drug_Cov	Must be non-missing	Enr1.19
	Drug_Cov	Must have the values of either “Y”, “N”, or “U” only	Enr1.20

2. Table: Demographic

	Variable Name	Rule	Error Code
1	PatID	Must be character data type	Dem1.1
	PatID	Must be non-missing	Dem1.2
	PatID	Must be left justified	Dem1.3
2	Birth_Date	Must be a SAS date value of numeric data type	Dem1.4
	Birth_Date	Must be of SAS length 4	Dem1.5
	Birth_Date	Must be non-missing	Dem1.6
	Birth_Date	Earliest permitted date is January 1, 1885; latest permitted date is current date	Dem1.7
3	Sex	Must be character data type	Dem1.8
	Sex	Must be exactly 1 character in length	Dem1.9
	Sex	Must be non-missing	Dem1.10
	Sex	Must have the values of either “F”, “M”, “A”, or “U” only	Dem1.11
4	Hispanic	Must be character data type	Dem1.12
	Hispanic	Must be exactly 1 character in length	Dem1.13
	Hispanic	Must be non-missing	Dem1.14
	Hispanic	Must have the values of either “Y”, “N”, or “U” only	Dem1.15
5	Race	Must be character data type	Dem1.16
	Race	Must be exactly 1 character in length	Dem1.17
	Race	Must be non-missing	Dem1.18
	Race	Must have the values of either “0”, “1”, “2”, “3”, “4”, or “5” only	Dem1.19

3. Table: Dispensing

	Variable Name	Rule	Error Code
1	PatID	Must be character data type	Dis1.1
	PatID	Must be non-missing	Dis1.2
	PatID	Must be left justified	Dis1.3
2	RxDate	Must be a SAS date value of numeric data type	Dis1.4
	RxDate	Must be of SAS length 4	Dis1.5
	RxDate	Must be non-missing	Dis1.6

	Variable Name	Rule	Error Code
3	NDC	Must be character data type	Dis1.7
	NDC	Must be exactly 11 characters in length	Dis1.8
	NDC	Must be non-missing	Dis1.9
	NDC	Must only contain digits from 0-9 (i.e., no space or other characters)	Dis1.10
4	RxSup	Must be a SAS date value of numeric data type	Dis1.11
	RxSup	Must be of SAS length 4	Dis1.12
	RxSup	Must be non-negative	Dis1.13
5	RxAmt	Must be a SAS date value of numeric data type	Dis1.15
	RxAmt	Must be of SAS length 4	Dis1.16
	RxAmt	Must be non-negative	Dis1.17

4. Table: Encounter

	Variable Name	Rule	Error Code
1	PatID	Must be character data type	Enc1.1
	PatID	Must be non-missing	Enc1.2
	PatID	Must be left justified	Enc1.3
2	EncounterID	Must be character data type	Enc1.4
	EncounterID	Must be non-missing	Enc1.5
	EncounterID	Must be left justified	Enc1.6
3	ADate	Must be a SAS date value of numeric data type	Enc1.7
	ADate	Must be of SAS length 4	Enc1.8
	ADate	Must be non-missing	Enc1.9
4	DDate	Must be a SAS date value of numeric data type	Enc1.10
	DDate	Must be of SAS length 4	Enc1.11
5	Provider	Must be character data type	Enc1.12
	Provider	Must be non-missing	Enc1.13
	Provider	Must be left justified	Enc1.14
6	Facility_Location	Must be character data type	Enc1.15
	Facility_Location	Must be exactly 3 characters in length	Enc1.16

	Variable Name	Rule	Error Code
	Facility_Location	Must be left justified	Enc1.17
	Facility_Location	Must only contain digits from 0-9 (i.e., no space or other characters)	Enc1.18
7	EncType	Must be character data type	Enc1.19
	EncType	Must be exactly 2 characters in length	Enc1.20
	EncType	Must be non-missing	Enc1.21
	EncType	Must have the values of either “IP”, “IS”, ”ED”, “AV" or “OA” only	Enc1.22
8	Facility_Code	Must be character data type	Enc1.23
	Facility_Code	Must be at least 5 and no more than 25 characters in length	Enc1.24
	Facility_Code	Must be left justified	Enc1.25
	Facility_Code	Must include only uppercase letters and/or digits; no embedded blanks or special characters	Enc1.26
9	Discharge_Disposition	Must be character data type	Enc1.27
	Discharge_Disposition	Must be exactly 1 character in length	Enc1.28
	Discharge_Disposition	Must have the values of either “A”, “E” or “U” only	Enc1.29
10	Discharge_Status	Must be character data type	Enc1.30
	Discharge_Status	Must be exactly 2 characters in length	Enc1.31
	Discharge_Status	Must have the values of either “AF”, “AL”, “AM”,“AW”, “EX”, “HH”, “HS”, “HO”, “IP”, “NH”, “OT”, “RS”, “RH”, “SH”, or “SN” only	Enc1.32
11	DRG	Must be character data type	Enc1.33
	DRG	Must be exactly 3 characters in length	Enc1.34
	DRG	Must only contain digits from 0-9 (i.e., no space or other characters)	Enc1.35
12	DRG_Type	Must be character data type	Enc1.36
	DRG_Type	Must be exactly 1 character in length	Enc1.37
	DRG_Type	Must have the values of exactly “1” or “2” only	Enc1.38
13	Admitting_Source	Must be character data type	Enc1.39
	Admitting_Source	Must be exactly 2 characters in length	Enc1.40
	Admitting_Source	Must have the values of either “AV”, “ED”, “AF”, “AL”, “HH”, “HS”, “HO”, “IP”, “NH”, “OT”, “RS”, “RH”, “SN” or “UN” only	Enc1.41

5. Table: Diagnosis

	Variable Name	Rule	Error Code
1	PatID	Must be character data type	Dia1.1
	PatID	Must be non-missing	Dia1.2
	PatID	Must be left justified	Dia1.3
2	EncounterID	Must be character data type	Dia1.4
	EncounterID	Must be non-missing	Dia1.5
	EncounterID	Must be left justified	Dia1.6
3	ADate	Must be a SAS date value of numeric data type	Dia1.7
	ADate	Must be of SAS length 4	Dia1.8
	ADate	Must be non-missing	Dia1.9
4	Provider	Must be character data type	Dia1.10
	Provider	Must be non-missing	Dia1.11
	Provider	Must be left justified	Dia1.12
5	EncType	Must be character data type	Dia1.13
	EncType	Must be exactly 2 characters in length	Dia1.14
	EncType	Must be non-missing	Dia1.15
	EncType	Must have the values of either “IP”, “IS”, ”ED”, “AV" or “OA” only	Dia1.16
6	DX	Must be character data type	Dia1.17
	DX	Must be exactly 6 characters in length	Dia1.18
	DX	Must be non-missing	Dia1.19
	DX	Must contain only uppercase letters and/or digits; no embedded blanks or special characters other than decimal point if needed; no local suffix	Dia1.20
7	DX_Codetype	Must be character data type	Dia1.21
	DX_Codetype	Must be exactly 2 characters in length	Dia1.22
	DX_Codetype	Must have the values of either “09”, “10”, “11” or “OT” only	Dia1.23
8	OrigDX	Must be character data type	Dia1.24
	OrigDX	Must be at least 2 and no more than 25 characters in length	Dia1.25
9	PDX	Must be character data type	Dia1.26
	PDX	Must be exactly 1 character in length	Dia1.27
	PDX	Must have the values of either “P”, “S” or “X” only	Dia1.28

6. Table: Procedure

	Variable Name	Rule	Error Code
1	PatID	Must be character data type	Pro1.1
	PatID	Must be non-missing	Pro1.2
	PatID	Must be left justified	Pro1.3
2	EncounterID	Must be character data type	Pro1.4
	EncounterID	Must be non-missing	Pro1.5
	EncounterID	Must be left justified	Pro1.6
3	ADate	Must be a SAS date value of numeric data type	Pro1.7
	ADate	Must be of SAS length 4	Pro1.8
	ADate	Must be non-missing	Pro1.9
4	Provider	Must be character data type	Pro1.10
	Provider	Must be non-missing	Pro1.11
	Provider	Must be left justified	Pro1.12
5	EncType	Must be character data type	Pro1.13
	EncType	Must be exactly 2 characters in length	Pro1.14
	EncType	Must be non-missing	Pro1.15
	EncType	Must have the values of either “IP”, “IS”, ”ED”, “AV" or “OA” only	Pro1.16
6	PX	Must be character data type	Pro1.17
	PX	Must be exactly 6 characters in length	Pro1.18
	PX	Must be non-missing	Pro1.19
	PX	Must contain only uppercase letters and/or digits; no embedded blanks or special characters; no local suffix	Pro1.20
7	PX_Codetype	Must be character data type	Pro1.21
	PX_Codetype	Must be exactly 2 characters in length	Pro1.22
	PX_Codetype	Must be non-missing	Pro1.23
	PX_Codetype	Must have the values of either “09”, “10”, “11”, “C4”, “HC”, “H3”, “C2”, “C3”, “RE”, “LO” or “OT” only	Pro1.24
8	OrigPX	Must be character data type	Pro1.25
	OrigPX	Must be at least 2 and no more than 25 characters in length	Pro1.26

7. Table: Death

	Variable Name	Rule	Error Code
1	PatID	Must be character data type	Dth1.1
	PatID	Must be non-missing	Dth1.2
	PatID	Must be left justified	Dth1.3
2	DeathDt	Must be a SAS date value of numeric data type	Dth1.4
	DeathDt	Must be of SAS length 4	Dth1.5
	DeathDt	Must be non-missing	Dth1.6
	DeathDt	Earliest permitted date is January 1, 1885; latest permitted date is current date	Dth1.7
3	DtImpute	Must be character data type	Dth1.8
	DtImpute	Must be exactly 1 character in length	Dth1.9
	DtImpute	Must be non-missing	Dth1.10
	DtImpute	Must have the values of either “B”, “D”, “M”, or “N” only	Dth1.11
4	Source	Must be character data type	Dth1.12
	Source	Must be exactly 1 character in length	Dth1.13
	Source	Must be non-missing	Dth1.14
	Source	Must have the values of either “L”, “N”, “S”, or “T” only	Dth1.15
5	Confidence	Must be character data type	Dth1.16
	Confidence	Must be exactly 1 character in length	Dth1.17
	Confidence	Must be non-missing	Dth1.18
	Confidence	Must have the values of either “E”, “F” or “P” only	Dth1.19

8. Table: Cause of Death

	Variable Name	Rule	Error Code
1	PatID	Must be character data type	COD1.1
	PatID	Must be non-missing	COD1.2
	PatID	Must be left justified	COD1.3
2	COD	Must be character data type	COD1.4
	COD	Must be non-missing	COD1.5
	COD	Must be exactly 3, 5 or 6 characters in length	COD1.6
	COD	Must be left justified	COD1.7
	COD	Can consist of only digits or one decimal point	COD1.8
	COD	If decimal point is present, must be positioned only as the 4th character from the left	COD1.9
3	CodeType	Must be character data type	COD1.10
	CodeType	Must be exactly 1 character in length	COD1.11
	CodeType	Must be non-missing	COD1.12
	CodeType	Must have the values of either “A” or “B” only	COD1.13
4	CauseType	Must be character data type	COD1.14
	CauseType	Must be exactly 1 character in length	COD1.15
	CauseType	Must be non-missing	COD1.16
	CauseType	Must have the values of either “C”, “I”, “O”, or “U” only	COD1.17
5	Source	Must be character data type	COD1.18
	Source	Must be exactly 1 character in length	COD1.19
	Source	Must be non-missing	COD1.20
	Source	Must have the values of either “L”, “N”, “S”, or “T” only	COD1.21
6	Confidence	Must be character data type	COD1.22
	Confidence	Must be exactly 1 character in length	COD1.23
	Confidence	Must be non-missing	COD1.24
	Confidence	Must have the values of either “E”, “F” or “P” only	COD1.25

B. APPENDIX B: LEVEL TWO: DATA CHECK REQUIREMENTS

1. Table: Enrollment

	Variable Name	Rule	Error Code
1	PatID	Can occur more than once in the file	Enr2.1
	PatID	Must have a corresponding value in the Demographic table	Enr2.2
2	Enr_Start	Must be earlier than or equal to Enr_End	Enr2.3
	Enr_Start	In combination with PatID, MedCov, and DrugCov, must occur only once in the file	Enr2.4
3	Enr_End	In combination with PatID, MedCov, and DrugCov, must occur only once in the file (implemented in Year 2)	Enr2.5

2. Table: Demographic

	Variable Name	Rule	Error Code
1	PatID	Must have at least one corresponding value in the Enrollment table	Dem2.1
	PatID	Can occur only once in the file	Dem2.2

3. Table: Dispensing

	Variable Name	Rule	Error Code
1	PatID	Must have at least one corresponding value in the Enrollment table	Dis2.1
2	NDC	In combination with PatID and RxDate, must occur only once in the table	Dis2.2

4. Table: Encounter

	Variable Name	Rule	Error Code
1	PatID	Must have at least one corresponding value in the Enrollment table	Enc2.1
2	EncounterID	Must occur only once in the table	Enc2.2
3	DDate	Must be populated for EncType = “IP”, “IS” or “ED” only	Enc2.3
4	Provider	In combination with PatID, ADate and EncType, must occur only once in the table	Enc2.4
5	EncType	In combination with PatID, Adate, and DDate, must occur only once in the table	Enc2.5
6	Discharge_Disposition	Must be populated for EncType = “IP”, “IS” or “ED” only	Enc2.6
7	Discharge_Status	Must be populated for EncType = “IP”, “IS” or “ED” only	Enc2.7
8	DRG	Must be populated for EncType = “IP” only	Enc2.8
9	DRG_Type	Must be populated for EncType = “IP” only	Enc2.9
10	Admitting_Source	Must be populated for EncType = “IP”, “IS” or “ED” only	Enc2.10

5. Table: Diagnosis

	Variable Name	Rule	Error Code
1	PatID	Must have at least one corresponding value in the Enrollment table	Dia2.1
2	EncounterID	Must occur only once in the table	Dia2.2
	EncounterID	Must have exactly one corresponding value in the Encounter table (implemented in Year 2)	Dia2.3
3	EncType	In combination with PatID, Adate, DDate, and DX, must occur only once in the table	Dia2.4
4	DX	Value must be consistent with Dx_CodeType	Dia2.5
5	PDX	Must be populated for EncType = “IP”, “IS” or “ED” only	Dia2.6

6. Table: Procedure

	Variable Name	Rule	Error Code
1	PatID	Must have at least one corresponding value in the Enrollment table	Pro2.1
2	EncounterID	Must occur only once in the table	Pro2.2
	EncounterID	Must have exactly one corresponding value in the Encounter table (implemented in Year 2)	Pro2.3
3	EncType	In combination with PatID, Adate, DDate, and PX, must occur only once in the table	Pro2.4
4	PX	Value must be consistent with PX_CodeType	Pro2.5

7. Table: Death

	Variable Name	Rule	Error Code
1	PatID	Must have at least one corresponding value in the Enrollment table	Dth2.1
2	DeathDt	Must be on or after Birth_Date in Demographic table (implemented in Year 2)	Dth2.2
	DeathDt	In combination with PatID, must occur only once in the table	Dth2.3
	DeathDt	Year and Month of Death must be contemporaneous to Year and Month of latest month observed in the Enrollment table (implemented in Year 2)	Dth2.4

8. Table: Cause of Death

	Variable Name	Rule	Error Code
1	PatID	Must have at least one corresponding value in the Enrollment table	COD2.1
	PatID	Must have only one corresponding value in the Death table	COD2.2
2	CodeType	For any one PatID, all values for CodeType should be identical (implemented in Year 2)	COD2.3
3	CauseType	With PatID, value of “U” must occur once and only once in this file (implemented in Year 2)	COD2.4

C. APPENDIX C: LEVEL THREE: DATA QUALITY PROFILING

Level 3 checks enable comparisons of descriptive statistics (e.g., trends, distributions) for each variable. Appendix E provides a detailed presentation by Data Partner of the information collected as part of the Level 3 data checks.

There are two primary comparisons: the first is within and across Data Partner for an individual data extracts, and the second compares individual Data Partner data between data refreshes. Level 3 checks are described below.

1. Table: Enrollment

	Statistic	Item/Variable	Error Code
1	Statistics on table	Number of records	Enr3.1
		Number of unique PatIDs (includes number/percent with missing if any)	Enr3.2
2	Distribution of Enrollment Dates	Enrollment Start: min, max, distribution by year, year-month	Enr3.3
		Enrollment End: min, max, distribution by year, year-month	Enr3.4
3	Distribution of Enrollment Months per Patient (by coverage status)	MedCov="Y": mean, standard deviation, and median, and distribution (0-6, 7-12, 13-18,19-24, 25-36, 37-48, 48+)	Enr3.5
		DrugCov="Y": mean, standard deviation, and median, and distribution	Enr3.6
		MedCov="Y" and DrugCov="Y": mean, standard deviation, and median, and distribution	Enr3.7
4	Distribution of Medical and Drug Coverage Indicators	MedCov	Enr3.8
		DrugCov	Enr3.9
		MedCov *DrugCov	Enr3.10

2. Table: Demographic

	Statistic	Item/Variable	Error Code
1	Statistics on table	Number of records	Dem3.1
		Number of unique PatIDs (includes number/percent with missing if any)	Dem3.2
2	Distribution of Birth Dates	Min, max, distribution by year	Dem3.3

	Statistic	Item/Variable	Error Code
3	Distribution of Age as of last refresh date[1]	Overall by age, mean, standard deviation, median	Dem3.4
		Overall by age groups (in years: 0-1, 2-4, 5-9, 10-14, 15-18, 19-21, 22-44, 45-64, 65-74, 75+)	Dem3.5
4	Distribution of Other Demographics	Sex	Dem3.6
		Race	Dem3.7
		Hispanic	Dem3.8

3. Table: Dispensing

	Statistic	Item/Variable	Error Code
1	Statistics on table	Number of records	Dis3.1
		Number of unique PatIDs (includes number/percent with missing if any)	Dis3.2
2	Distribution of RxDate (i.e., total number of dispensings per RxDate):	Overall for any NDC by Year, and by Year-month	Dis3.3
3	Average number of Prescriptions per PatID	By Year	Dis3.4
4	Distribution of Rx	RxAmt—all years, overall	Dis3.5
		RxSup—all years, overall	Dis3.6

4. Table Encounter

	Statistic	Item/Variable	Error Code
1	Statistics on table	Number of records	Enc3.1
		Number of unique PatIDs (includes number/percent with missing if any)	Enc3.2
		Number of unique EncounterIDs (includes number/percent with missing if any)	Enc3.3
		Number of unique Provider values (includes number/percent with missing if any)	Enc3.4
2	Distribution of ADate	By Year, and Year-Month	Enc3.5

[1] The last refresh date is a fixed date determined by individual Data Partners.

	Statistic	Item/Variable	Error Code
3	Distribution of EncType	All years – overall by Year	Enc3.6
		Percent encounters by Year - Enctype (must sum to 100%)	Enc3.7
		Percent encounters by Year – Month - Enctype (must sum to 100%)	Enc3.8
		Percent encounters by Year – Month - Enctype (must sum to 100%)	Enc3.9
		Number of Encounters per member by Year by EncType (mean, median, min, max…)	Enc3.10
4	Distribution of DDate	All years – overall by Year, and by Year-month	Enc3.11
		By EncType	Enc3.12
5	Distribution of Length of Stay (DDate – Adate +1)	By EncType (must include % of missing for non hospitals EncType)	Enc3.13
6	Distribution of Discharge_Disposition	All years – by Year	Enc3.14
		By EncType	Enc3.15
7	Distribution of Discharge_Status	All years – by Year	Enc3.16
		By EncType	Enc3.17
8	Distribution of Admitting_Source	All years – by Year	Enc3.18
		By EncType	Enc3.19
9	Distribution of DRG and DRG_Type	Overall	Enc3.20
10	Distribution of DRG_Type	All years – by Year	Enc3.21
		By EncType	Enc3.22

5. Table: Diagnosis

	Statistic	Item/Variable	Error Code
1	Statistics on table	Number of records	Dia3.1
		Number of unique PatIDs (includes number/percent with missing if any)	Dia3.2
		Number of unique EncounterIDs (includes number/percent with missing if any)	Dia3.3
		Number of unique Provider values (includes number/percent with missing if any)	Dia3.4
2	Distribution of ADate	All years – overall by ADate	Dia3.5
		Year-Month	Dia3.6
		Year	Dia3.7
3	Distribution of EncType	All years – overall by ADate	Dia3.8
		Year-Month	Dia3.9
		Year	Dia3.10
4	Distribution of DX_Codetype*DX	Overall	Dia3.11
5	Distribution of DX	Overall	Dia3.12
6	Distribution of DX_Codetype	Year-Month	Dia3.13
		Year	Dia3.14
		By EncType	Dia3.15
		By EncType-Year	Dia3.16
7	Distribution of OrigDX	Overall (all years, must include percent of missing)	Dia3.17
		OrigDx indicator * EncType	Dia3.18
		OrigDx indicator * EncType-Year	Dia3.19
8	Distribution of PDX	Overall (all years, must include percent of missing)	Dia3.20
		PDX * EncType	Dia3.21
		PDX * EncType-Year	Dia3.22
9	Number of Diagnosis per Encounter Visit	Overall by EncType (all years, mean, std, min, p5, median, p95, max	Dia3.23

6. Table: Procedure

	Statistic	Item/Variable	Error Code
1	Statistics on table	Number of records	Pro3.1
		Number of unique PatIDs (includes number/percent with missing if any)	Pro3.2
		Number of unique EncounterIDs (includes number/percent with missing if any)	Pro3.3
		Number of unique Provider values (includes number/percent with missing if any)	Pro3.4
2	Distribution of ADate	All years – overall by ADate	Pro3.5
		Year-Month	Pro3.6
		Year	Pro3.7
3	Distribution of EncType	All years – overall by ADate	Pro3.8
		Year-Month	Pro3.9
		Year	Pro3.10
4	Distribution of PX_Codetype*PX	Overall	Pro3.11
5	Distribution of PX	Overall	Pro3.12
6	Distribution of PX_Codetype	Year-Month	Pro3.13
		Year	Pro3.14
		By EncType	Pro3.15
		By EncType-Year	Pro3.16
7	Distribution of OrigPX	Overall (all years, must include percent of missing)	Pro3.17
		OrigPx indicator * EncType	Pro3.18
		OrigPx indicator * EncType-Year	Pro3.19
8	Number of Unique Procedure per Encounter Visit (not necessarily unique)	Overall by EncType (all years, mean, std, min, p5, median, p95, max)	Pro3.20

7. Table: Death

	Statistic	Item/Variable	Error Code
1	Statistics on table	Number of records	Dth3.1
		Number of unique PatIDs (includes number/percent with missing if any)	Dth3.2
2	Distribution of Death Dates	Min, max, distribution by year	Dth3.3
3	Distribution of Age at Death	Overall by age, mean, standard deviation, median	Dth3.4
		Overall by age groups (0-4 weeks, 5-16 weeks, 17-36 weeks, 37-52 weeks, 1-4 5-9 10-19 20-39 40-64 65-74 75-99 100+)	Dth3.5
4	Distribution of Death Variables (Overall)	Dt Impute	Dth3.6
		Source	Dth3.7
		Confidence	Dth3.8
		Source*Confidence	Dth3.9

8. Table: Cause of Death

	Statistic	Item/Variable	Error Code
1	Statistics on table	Number of records	COD3.1
		Number of unique PatIDs (includes number/percent with missing if any)	COD3.2
2	Distribution of CodeType and COD	CodeType	COD3.3
		CodeType*COD	COD3.4
3	Distribution of Death Variables (Overall)	CauseType	COD3.5
		Source	COD3.6
		Confidence	COD3.7
		Source*Confidence	COD3.8

D. APPENDIX D. MINI-SENTINEL DATA QUALITY CHECKING SUMMARY

Table. Summary of MSCDM Data Quality Review for All Data Partners

Err#	MSCDM Item	Dataset	MSOC Comment	# of Data Partners
Enrollment				
ENR2.1	Definition of the table: 1 record per enrollment period (i.e., unique combination of PatID, Enr_Start, Enr_End, MedCov and DrugCov)	Enr_l2_def	Some records have duplicate information.	3
ENR2.3	Enr_Start, Enr_End	Flags_l1_l2	Some records have Enr_End < Enr_Start.	3
ENRL3	MedCov, Drug Cov	Enr_l3_meddrugcov	Number of records with MedCov=No and DrugCov=Yes is low.	1
ENRL3	MedCov, Drug Cov	Enr_l3_meddrugcov	Number of records with MedCov=Yes and DrugCov=No is high.	2
ENRL3	MedCov, Drug Cov	Enr_l3_meddrugcov	Number of records with MedCov=Yes and DrugCov=U (Unknown) is high.	3
ENRL3	MedCov, DrugCov (Dataset: Enr_l3_enrmd_y)	Enr_l3_meddrugcov	All records have MedCov=Yes and DrugCov=Yes.	1
ENRL3	MedCov, Drug Cov	Enr_l3_enrmd_y	All enrollment periods in a given period of time are MedCov=Yes and DrugCov=No only. There are no Drug only or Med and Drug periods.	1
ENRL3	Months of Membership	Enr_l3_stats_enrd, Enr_l3_stats_enrm,	Maximum number of enrolled months per member > number of available months in enrollment table.	
ENRL3	Membership Rate	Enr_l3_enrmd_ym, Enr_l3_enrmd_y	Inconsistent trends in membership rates (i.e., decreasing or spiking at different points in time)	
ENRL3	Enr_Start	Enr_l3_dist_start	Max value is later than current date.	1
ENRL3	Enr_End	Enr_L3_dist_end	Max value is later than current date.	1

Table. Summary of MSCDM Data Quality Review for All Data Partners

Err#	MSCDM Item	Dataset	MSOC Comment	# of Data Partners
Demographic				
ENR0.0	Table must exist	Should create something?	Table does not exist	
DEM2.1	Definition of the table: 1 record per member (each PatID encountered only once)	Dem_l2_def	Some records have duplicate information.	2
DEM1.9	Birth_Date: Between 1/1/1885 - Current Date	Flags_l1_l2	Some records (patients) have birth dates outside expected range (1/1/1885 - Current Date)	14
DEM2.1b	PatID: Must have a corresponding value in the Enrollment table	Dem_l2_PatIDmatch	There are some members without matching PatID value in the Enrollment table.	1
DEML3	Age	Dem_l3_ageyrsdist1	There are outliers in the age distribution.	14
DEML3	Age (Dataset: DEM_l3_ageyrsdist2)	Dem_l3_ageyrsdist2	Some age groups are empty.	1

Table. Summary of MSCDM Data Quality Review for All Data Partners

Err#	MSCDM Item	Dataset	MSOC Comment	# of Data Partners
Dispensing				
COD0.0	Table must exist	Should create something?	Table does not exist	1
DIS2.1	Definition of the table: 1 record per dispensing (i.e., unique combination of PatID, RxDate, NDC)	Dis_n_all	Some records have duplicate information.	13
DIS1.10	NDC: Length of NDC= 11	Flags_l1_l2	Some NDC values have incorrect lengths (not = 11).	9
DIS1.12	NDC: no special characters	Flags_l1_l2	Some NDC values have special characters or non-digits.	4
DIS1.15	RxSup: non-negative	Flags_l1_l2, Dis_l3_rxsup	Some records have non-positive RxSup values (i.e., negative, zero, or missing).	12
DIS1.15	RxSup: high	Dis_l3_rxsup	Some records have RxSup > 100.	13
DIS1.18	RxAmt: non-negative	Flags_l1_l2, Dis_l3_rxamt	Some records have non-positive RxAmt values (i.e., negative, zero, or missing).	12
DIS1.18	RxAmt: high	Dis_l3_rxamt	Some records have RxAmt > 100.	13
DISL3	Max Rx/Yr	Dis_l3_rxptyr	Max number of RX per member per year is high (>200).	15

Table. Summary of MSCDM Data Quality Review for All Data Partners

Err#	MSCDM Item	Dataset	MSOC Comment	# of Data Partners
Encounter				
ENC2.1	Definition of the table: 1 record per encounter (i.e., unique combination of PatID, ADate, Provider and EncType); EncounterID must be unique across PatIDs.	Enc_n_all	Some records have duplicate information.	10
ENC	Adate	Enc_l3_adate_y, Enc_l3_adate_ym	Some records have empty values. Please populate this field.	1
ENC	DDate	Enc_l3_ddate_y, Enc_l3_ddate_ym	Some records have empty values. Please populate this field.	1
ENC2.3	DDate	Enc_l3_enctype_ddate_y	Should only be populated for inpatient hospital and overnight encounters (IP, IS, and ED). Should be missing for ambulatory visits.	10
ENCL3	DDate and EncType	Enc_l3_enctype_ddate_y	DDate for EncTypes= AV and OA is set to 1/1/1900. Please set to missing (per MSCDM specification).	1
ENC1.22	Facility Location: 3 characters in length	Flags_l1_l2	All values are missing/blank. Please confirm this is accurate.	15
ENC1.24	Facility Location: Only digits	Flags_l1_l2	Should only contain digits.	1
ENC1.35	Discharge_Disposition: values of "A", "E" or "U"	Enc_l3_DisDisp_Enc	Some records contain invalid values.	1
ENC1.37	Discharge Status: must be 2 characters	Flags_l1_l2	All values are missing/blank. Please confirm this is accurate.	1
ENC1.47	Admitting Source: must be selected values	Enc_l3_admsrc	Some records contain invalid values.	10
ENCL3	Discharge Disposition	Enc_l3_DisDisp_Enc	Should only be populated for inpatint (IP) records.	9
ENCL3	Discharge Disposition	Enc_l3_DisDisp_Enc	Some inpatient (IP) records have values that are blank/missing or not allowed in the MSCDM.	2
ENCL3	Discharge Status	Enc_l3_Disstat_Enc	Should only be populated for inpatint (IP) records.	9
ENCL3	Discharge Status	Enc_l3_Disstat	Some inpatient (IP) records have values that are blank/missing or not allowed in the MSCDM.	2
ENCL3	EncType	Enc_l3_encdate_y	No record with EncType = IS (Non-Acute Institutional Stay) were found. Is that accurate?	1
ENCL3	EncType	Enc_l3_encdate_y	No record with EncType = OT (Other Ambulatory Visits). Is that accurate?	1
ENCL3	EncType	Enc_l3_encdate_y	% of records with EncType=ED unusually low for some years.	1
ENCL3	DRG codes and Encounter Types	Enc_l3_drg_enc_type	Some non inpatient (IP) records have DRG codes. What are those records?	1
ENC1.40	DRG: must be 3 characters	Enc_l3_drg_drg_type	Some records have values with more or less than 3 characters.	6
ENC1.41	DRG: Only digits	Enc_l3_drg_type_y	Should only contain digits.	7

Table. Summary of MSCDM Data Quality Review for All Data Partners

Err#	MSCDM Item	Dataset	MSOC Comment	# of Data Partners
ENCL3	DRG: MS-DRG starting in 2007	Enc_l3_drg_type_y	Old DRG system (CMS-DRG) ended in 2007. Some records contain new DRG codes (MS-DRG) prior to 2007.	1
ENCL3	DRG: MS-DRG starting in 2007	Enc_l3_drg_type_y	Old DRG system (CMS-DRG) ended in 2007 but most records post 2007 contain old DRG codes.	1
ENCL3	Number of monthly or yearly encounters	Enc_l3_adate_ym, Enc_l3_adate_y, Enc_l3_encdate_ym,	Inconsistent trends in number of records per month or year (i.e., decreasing or spiking at different points in time).	3
ENCL3	Encounter/Patient per year	Enc_l3_stats_y	Max number of Encounter per patient per year is high (>200).	7
ENCL3	Distribution of EncType by year	Enc_l3_encdate_ym, Enc_l3_encdate_y	Inconsistent trends in number of records per month or year by EncType (i.e., decreasing or spiking at different points in time).	1

Table. Summary of MSCDM Data Quality Review for All Data Partners

Err#	MSCDM Item	Dataset	MSOC Comment	# of Data Partners
Diagnosis				
DIA0.0	Table must exist	Should create something?	Table does not exist	
DIAL2	Definition of the table: 1 record per DX per encounter (i.e., unique combination of PatID, ADate, Provider, EncType and DX).	Dia_n_all	Some records have duplicate information.	10
DIA1.25	Dx: no missing values	Flags_l1_l2	Some records have missing values.	1
DIAL3	Dx: special characters	Flags_l1_l2	A number of records have invalid values or contain special characters such as *, ' , $, +, /, or spaces.	2
DIAL3	PDx: for IP encounters only.	Dia_l3_pdx_et	Primary diagnosis flag is a concept associated with inpatient (IP) encounter type only. This flag should only be populated for records withg EncType=IP.	11
DIAL3	PDx	Dia_l3_pdx_et	No values of X (Unable to classify) were recorded. Is this accurate?	1
DIAL3	PDx	Dia_l3_pdx_et	Some IP records have missing values.	2
DIAL3	PDx	Dia_l3_pdx_et	Some records have values not defined in the MSCDM.	1
DIAL3	Number of records per month (or year)	Dia_l3_adate_ym, Dia_l3_encdate_y	Inconsistent trends in number of records per month (i.e., decreasing or spiking at different points in time).	1
Procedure				
PRO0.0	Table must exist	Should create something?	Table does not exist	
PROL2	Definition of the table: 1 record per PX per encounter (i.e., unique combination of PatID, ADate, Provider, EncType and PX).	Pro_n_all	Some records have duplicate information.	12
PROL3	Adate	Pro_l3_adate, Pro_l3_adate_ym	Some records have ADate > current date.	1
PRO1.24	Px: special characters	Flags_l1_l2	A number of records contain special characters such as *, ' , $, +, /, or spaces.	1
PROL3	Number of records per month (or year)	Pro_l3_adate_ym, Pro_l3_encdate_y	Inconsistent trends in number of records per month or year (i.e., decreasing or spiking at different points in time).	1
PROL3	Px_Codetype: lookup for HCPCS Level III	Pro_l3_ct_y	Does your site have access to lookup/description for the HCPCS level III codes (Px_Codetype=H3) encountered in your data?	1
PROL3	Px_Codetype: Local/homegrown codes	Pro_l3_ct_y	Does your site have access to lookup/description for the local codes (Px_Codetype=LO) encountered in your data?	1
PROL3	PX_CodeType	Pro_l3_ct_y	There are no records with PX_CodeType=HC for HCPCS Level II (i.e., the "J Codes"). That is unusual.	1
PROL3	Revenue Code, EncType	Pro_l3_ct_enct, Pro_l3_ct_y	Revenue codes populated for AV: what are those visits exactly?	3

Table. Summary of MSCDM Data Quality Review for All Data Partners

Err#	MSCDM Item	Dataset	MSOC Comment	# of Data Partners
Death				
DTH0.0	Table must exist	Should create something?	Table does not exist	
DTHL2	Definition of the table: 1 record per member (each PatID encountered only once)	Dth_n_all	Some records have duplicate information.	1
DTH1.9	Death Date: no missing values	Dth_l3_dthdt_ym	Some records have missing values.	3
DTH2.1	PatID match: (Dataset: Dth_l2_patidmatch)	Dth_l2_PatIDmatch	There are some members without matching PatID value in the Enrollment table.	1
DTHL3	Death date	Dth_l3_dthdt_ym	Death records have a date of death prior to beginning of data.	7
DTHL3	Confidence: no missing values	Dth_l3_confidence	Some records have missing values.	1
COD				
COD0.0	Table must exist	Should create something?	Table does not exist	1
CODL2	Definition of the table: 1 record per cause of death per member (i.e., unique combination of PatID and COD)	Cod_n_all	Some records have duplicate information.	1
COD2.1	PatID match: (Dataset: Cod_l2_patidmatch)	Cod_l2_PatIDmatch	There are some members without matching PatID value in the Enrollment table.	1
CODL3	Causetype: no missing values	Cod_l3_causet	Some records have missing values.	1
CODL3	Codetype: no missing values	Cod_l3_codet	Some records have missing values.	1
CODL3	Confidence: no missing values	Cod_l3_confidence	Some records have missing values.	1

E. APPENDIX E. MINI-SENTINEL DATA CHARACTERIZATION REPORT

Mini-Sentinel Year 1 Data Partner Quality Check and Characterization Output

Prepared by: Mini-Sentinel Data Core

Author Affiliations: Harvard Pilgrim Health Care Institute, Boston, MA

January 12, 2011

Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the Sentinel Initiative, a multi-faceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. Mini-Sentinel Collaborators include Data and Academic Partners that provide access to health care data and ongoing scientific, technical, methodological, and organizational expertise. The Mini-Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.

Table of Contents

Table/Figure	Description
Table 1	Snapshot of the Mini-Sentinel Distributed Database Enrollment Table in Extract 1
Figure 1	Distribution of Medical Coverage in MSDD Enrollment Table in Extract 1
Figure 2	Distribution of Length of Enrollment (Months) in MSDD Enrollment Table in Extract 1
Table 2	Snapshot of the Mini-Sentinel Distributed Database Demographic Table in Extract 1
Figure 3	Distribution of Race in Demographic Table in Extract 1
Figure 4	Distribution of Sex in MSDD Demographic Table in Extract 1
Figure 5	Distribution of Hispanic in Demographic Table in Extract 1
Figure 6	Distribution of Age Group in MSDD Demographic Table in Extract 1
Figure 7	Distribution of Age Group in Demographic Table in Extract 1, by Data Partner
Figure 8	Percentage of Persons in Demographic Table, Aged 0-4 Weeks in Extract 1
Figure 9	Percentage of Persons in Demographic Table, Aged 5-52 Weeks in Extract 1
Figure 10	Percentage of Persons in Demographic Table, Aged 1-4 Years in Extract 1
Figure 11	Percentage of Persons in Demographic Table, Aged 5-9 Years in Extract 1
Figure 12	Percentage of Persons in Demographic Table, Aged 10-19 Years in Extract 1
Figure 13	Percentage of Persons in Demographic Table, Aged 20-39 Years in Extract 1
Figure 14	Percentage of Persons in Demographic Table, Aged 40-64 Years in Extract 1
Figure 15	Percentage of Persons in Demographic Table, Aged 65-74 Years in Extract 1
Figure 16	Percentage of Persons in Demographic Table, Aged 75-84 Years in Extract 1
Figure 17	Percentage of Persons in Demographic Table, Aged 85+ Years in Extract 1
Table 3	Snapshot of the Mini-Sentinel Distributed Database Dispensing Table in Extract 1
Figure 18	Total Number of Dispensings in MSDD Dispensing Table in Extract 1, by Year
Figure 19	Average Number of Dispensings per Patient in MSDD Dispensing Table in Extract 1, by Year
Figure 20	Average Number of Dispensings per Patient in MSDD Dispensing Table in Extract 1, by Data Partner
Figure 21	Distribution of Days Supplied in MSDD Dispensing Table in Extract 1
Figure 22	Distribution of Rx Amount in MSDD Dispensing Table in Extract 1
Table 4	Snapshot of the Mini-Sentinel Distributed Database Encounter Table in Extract 1
Figure 23	Average Number of Encounters per Patient in MSDD Encounter Table in Extract 1, by Year
Figure 24	Distribution of Encounter Type in MSDD Encounter Table in Extract 1
Figure 25	Distribution of Admission Source in MSDD Encounter Table in Extract 1
Figure 26	Distribution of Discharge Disposition in MSDD Encounter Table in Extract 1
Figure 27	Distribution of Discharge Status in MSDD Encounter Table in Extract 1
Figure 28	Distribution of Length of Stay in MSDD Encounter Table in Extract 1
Table 5	Snapshot of the Mini-Sentinel Distributed Database Diagnosis Table in Extract 1
Figure 29	Number Diagnoses per Encounter in MSDD Diagnosis Table in Extract 1, by Year
Figure 30	Distribution of Encounter Type in MSDD Diagnosis Table in Extract 1, by Data Partner
Table 6	Snapshot of the Mini-Sentinel Distributed Database Procedure Table in Extract 1
Figure 31	Number Procedures per Encounter, in MSDD Procedure Table in Extract 1, by Year
Figure 32	Distribution of Code Type in MSDD Procedure Table in Extract 1

Table of Contents (Cont.)

Table/Figure	Description
Table 7	Snapshot of the Mini-Sentinel Distributed Database Death Table in Extract 1
Figure 33	Distribution of Source in MSDD Cause of Death Table in Extract 1
Figure 34	Distribution of Confidence in MSDD Cause of Death Table in Extract 1
Table 8	Snapshot of the Mini-Sentinel Distributed Database Cause of Death Table in Extract 1
Figure 35	Distribution of Cause Type in MSDD Cause of Death Table in Extract 1
Figure 36	Distribution of Source in MSDD Cause of Death Table in Extract 1
Table 9	Snapshot of the Mini-Sentinel Distributed Database Patient Match in Extract 1

Table 1. Snapshot of the Mini-Sentinel Distributed Database Enrollment Table in Extract 1 (Unique Members = 62,704,250)

	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10	DP11	DP12	DP13	DP14	DP15	Total
Number of enrollment periods per person	3.0	1.2	3.1	3.4	1.8	1.0	1.9	2.3	4.0	3.6	3.8	3.8	1.5	2.5	1.7	2.0
Records with *only* medical coverage	6.3%	9.2%	0.0%	10.2%	11.4%	1.5%	6.0%	4.2%	0.0%	10.0%	12.3%	37.7%	0.0%	51.3%	5.0%	15.1%
Records with *only* drug coverage	0.0%	0.3%	0.0%	0.0%	0.0%	43.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.3%
Records with medical + unknown drug coverage	16.4%	0.0%	87.7%	0.0%	0.0%	0.0%	0.0%	0.0%	52.5%	14.3%	35.2%	0.0%	0.0%	0.0%	0.0%	6.4%
Records with *both* medical and drug coverage	77.3%	90.4%	12.3%	89.8%	88.6%	54.9%	94.0%	95.8%	47.5%	75.6%	52.5%	62.3%	100.0%	48.7%	95.0%	75.2%
Length of enrollment:																
Statistics (months):																
Min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	---
P1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	---
P5	0	2	0	0	0	0	2	0	0	2	1	1	3	0	2	---
P25	7	9	0	8	9	0	11	7	9	16	11	12	12	0	13	---
Median	24	21	0	22	24	3	28	23	30	47	32	36	32	21	36	---
P75	61	38	8	48	53	18	67	60	84	99	77	87	54	57	72	---
P95	120	73	121	112	125	30	127	122	129	120	126	126	114	126	110	---
P99	120	73	124	122	125	30	127	123	129	120	126	128	120	129	111	---
Max	120	134	182	185	128	32	133	312	135	120	126	163	174	137	113	---
Distribution (months):																
0 - 6	23.8%	19.3%	74.6%	21.8%	20.6%	55.3%	15.3%	23.3%	20.9%	11.8%	17.6%	17.1%	9.6%	34.4%	13.0%	24.1%
7 - 12	13.0%	15.6%	2.1%	15.1%	12.6%	14.3%	14.4%	13.3%	10.8%	8.9%	11.4%	9.2%	18.9%	7.1%	12.0%	13.5%
13 - 24	14.7%	21.8%	3.4%	17.9%	19.0%	14.1%	16.7%	16.0%	13.9%	12.9%	14.4%	13.9%	16.4%	11.7%	15.9%	17.7%
25 - 36	10.4%	16.4%	2.8%	12.0%	11.9%	16.2%	10.8%	10.3%	8.9%	9.5%	10.1%	9.9%	8.4%	9.4%	10.1%	13.8%
37 - 48	7.6%	8.8%	2.2%	8.3%	8.6%	0.0%	8.0%	7.3%	6.8%	7.9%	7.8%	8.1%	8.5%	7.7%	9.3%	7.2%
49 - 60	5.4%	5.9%	2.0%	6.2%	6.3%	0.0%	6.5%	5.4%	5.3%	6.7%	6.5%	6.4%	19.6%	6.1%	7.8%	5.4%
61 - 72	4.4%	4.4%	1.7%	4.2%	4.4%	0.0%	5.5%	4.2%	4.5%	5.8%	5.5%	5.1%	6.1%	5.0%	7.1%	4.0%
73 - 84	3.6%	7.7%	1.7%	4.2%	3.4%	0.0%	4.2%	3.7%	4.0%	5.3%	4.6%	4.4%	3.4%	4.9%	5.1%	5.2%
85 - 96	3.1%	0.0%	1.3%	2.8%	2.7%	0.0%	3.5%	3.3%	3.7%	5.2%	4.0%	4.0%	1.9%	3.6%	5.9%	1.6%
97 - 108	3.8%	0.0%	1.6%	1.8%	2.5%	0.0%	3.2%	3.2%	3.4%	4.8%	3.9%	3.7%	1.6%	2.4%	3.8%	1.5%
109 - 120	10.2%	0.0%	1.5%	3.0%	2.0%	0.0%	2.6%	4.3%	3.9%	21.0%	3.5%	3.6%	5.5%	2.0%	10.1%	3.4%
120 & Up	0.0%	0.0%	5.0%	2.5%	6.0%	0.0%	9.1%	5.7%	13.9%	0.0%	10.7%	14.6%	0.1%	5.6%	0.0%	2.7%

Figure 1. Distribution of Medical Coverage in MSDD Enrollment Table in Extract 1 (Unique Members = 62,704,250)

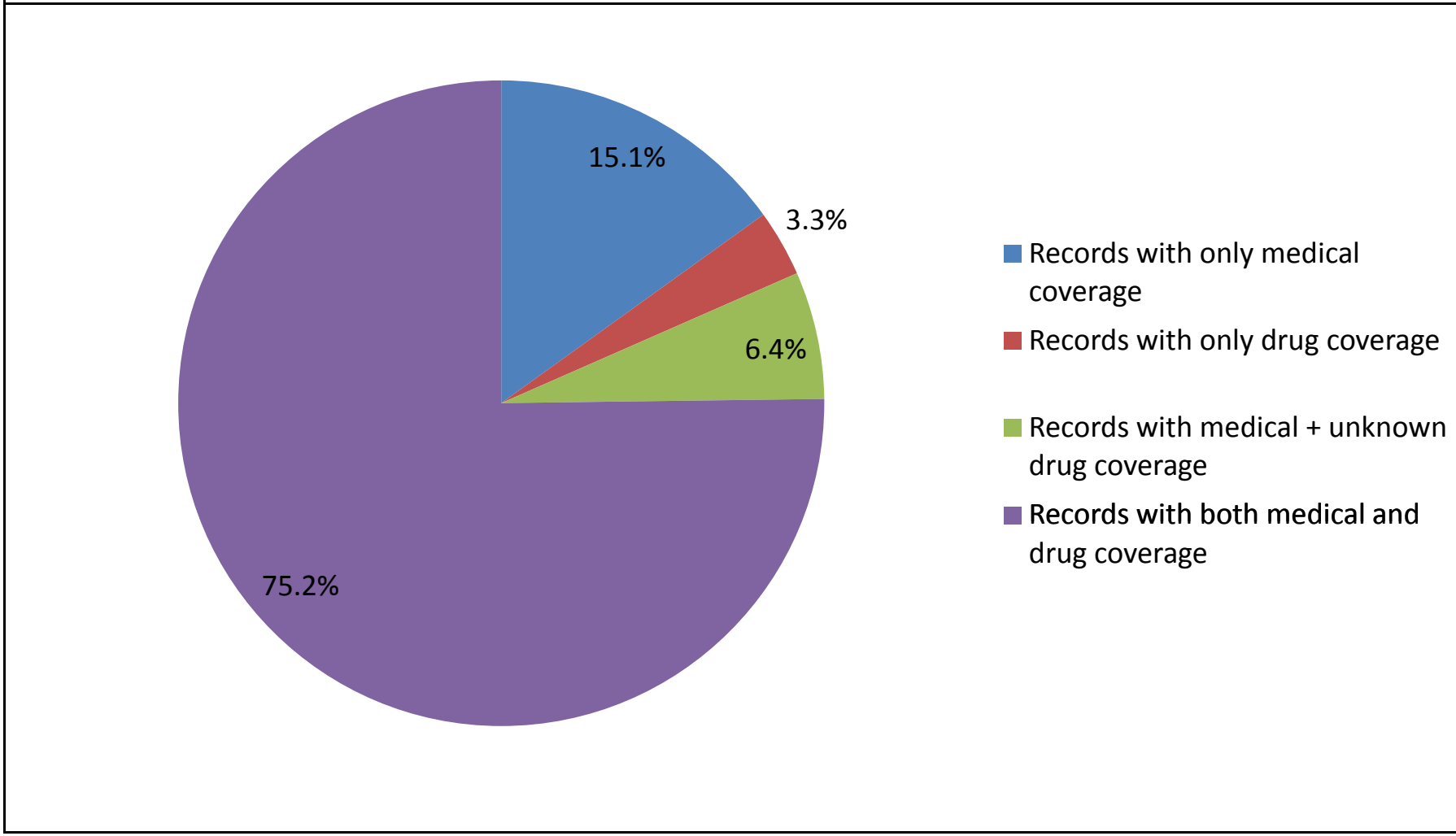


Figure 2. Distribution of Length of Enrollment (Months) in MSDD Enrollment Table in Extract 1 (Unique Members = 62,704,250)

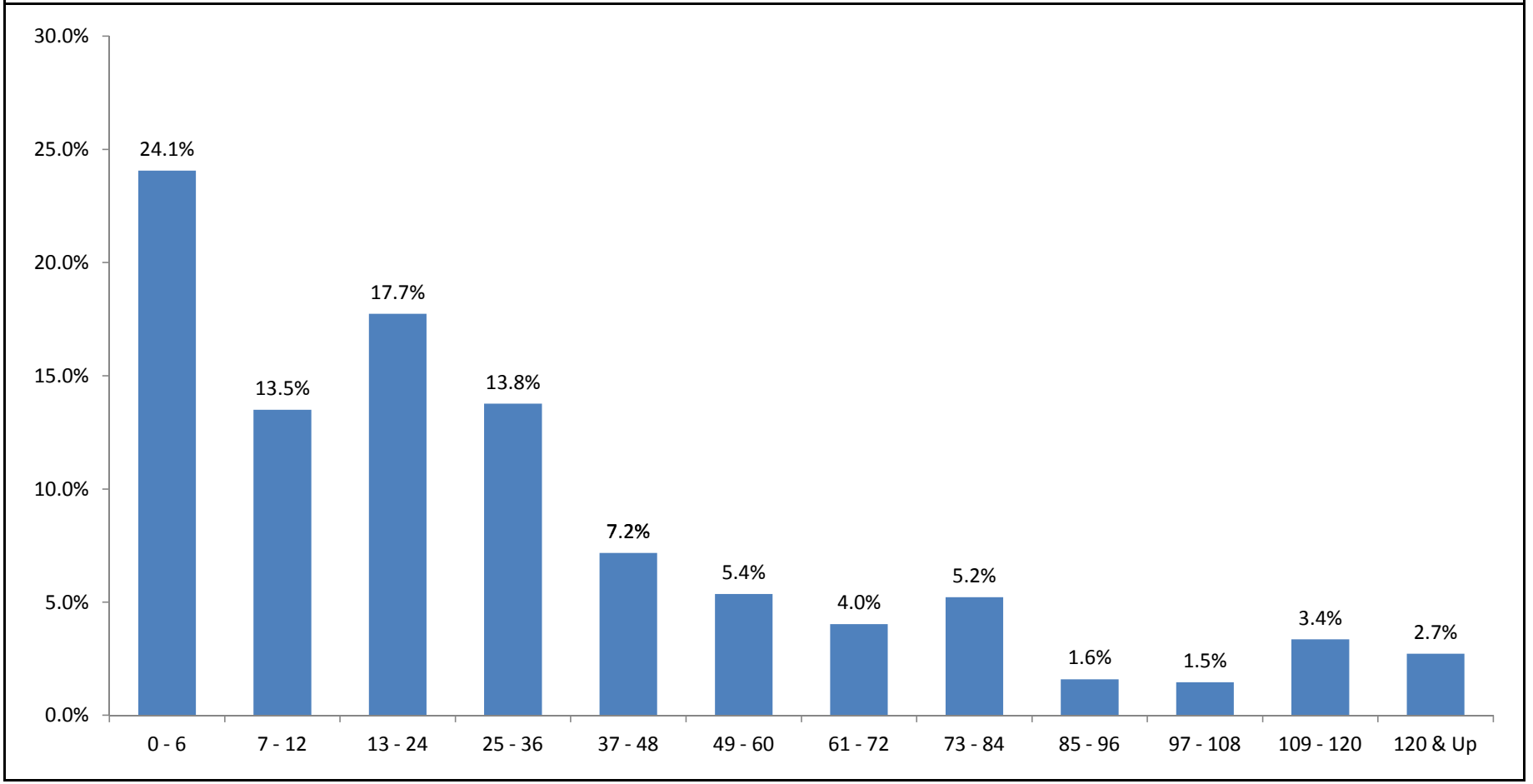


Table 2. Snapshot of the Mini-Sentinel Distributed Database Demographic Table in Extract 1 (Unique Individuals = 83,003,100)

	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10	DP11	DP12	DP13	DP14	DP15	Total
Race																
Unknown	90.0%	100.0%	16.5%	91.8%	86.2%	36.7%	81.5%	89.3%	84.1%	67.3%	79.8%	72.4%	100.0%	100.0%	93.0%	79.6%
American Indian / Alaska Native	0.2%	0.0%	0.4%	0.0%	0.1%	0.3%	0.1%	0.0%	0.3%	0.2%	0.2%	0.1%	0.0%	0.0%	0.0%	0.1%
Asian	0.6%	0.0%	3.2%	0.4%	0.8%	0.9%	0.7%	0.5%	5.9%	5.4%	0.8%	3.0%	0.0%	0.0%	0.0%	1.7%
African American	0.4%	0.0%	30.7%	0.6%	1.6%	6.1%	1.2%	4.5%	0.4%	3.4%	0.7%	4.3%	0.0%	0.0%	0.0%	2.6%
Native Hawaiian / Other Pacific Islander	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	0.1%	0.0%	3.9%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%	0.2%
White	8.7%	0.0%	49.3%	7.2%	11.3%	55.0%	16.4%	5.7%	5.5%	23.5%	18.5%	20.3%	0.0%	0.0%	7.0%	15.9%
Sex																
Male	47.4%	49.3%	46.4%	48.3%	48.9%	44.0%	48.1%	47.9%	49.9%	49.6%	48.5%	49.7%	46.6%	48.1%	49.2%	48.6%
Female	52.4%	50.7%	53.6%	51.7%	50.9%	56.0%	51.9%	52.1%	49.9%	50.4%	51.4%	50.3%	53.4%	51.9%	50.8%	51.4%
Ambiguous	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Unknown	0.1%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Hispanic																
Yes	0.3%	0.0%	1.6%	0.4%	0.4%	1.5%	3.4%	0.6%	1.3%	8.6%	0.8%	14.9%	42.7%	0.0%	0.0%	4.4%
No	5.7%	0.0%	0.0%	7.9%	0.0%	63.3%	14.0%	6.6%	10.7%	0.0%	10.1%	18.4%	51.5%	0.0%	7.1%	11.6%
Unknown	94.0%	100.0%	98.4%	91.7%	99.6%	35.3%	82.6%	92.8%	88.0%	91.4%	89.1%	66.7%	5.9%	100.0%	92.9%	84.1%
Age																
Statistics																
Mean	43.7	36.8	44.6	37.3	40.2	58.8	41.9	37.8	47.1	47.2	46.1	41.6	35.8	48.1	41.9	---
STD	194.4	535.7	134.3	167.6	230.1	356.0	155.4	113.3	127.2	408.4	175.5	360.1	102.1	84.4	77.3	---
Min	-1.0	-2,290.0	2.0	-2.0	-7,889.0	-1.0	-1.0	-1.0	-1.0	0.0	-38.0	-1.0	0.0	-1.0	1.0	---
Max	210.0	232.0	257.0	128.0	169.0	170.0	161.0	127.0	164.0	150.0	141.0	210.0	136.0	204.0	114.0	---
Distribution																
0 - 4 wks	0.1%	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.1%	0.0%	0.2%	0.0%	0.0%
5 - 52 wks	0.3%	0.3%	0.0%	0.4%	0.4%	0.2%	0.3%	0.3%	0.2%	0.2%	0.2%	0.3%	0.6%	0.6%	0.0%	0.3%
1 - 4 yrs	1.6%	3.2%	1.4%	2.8%	2.8%	1.6%	2.0%	1.9%	1.3%	1.7%	1.5%	1.9%	4.2%	3.1%	1.6%	2.4%
5 - 9 yrs	2.7%	6.1%	4.5%	5.5%	4.6%	2.3%	3.7%	3.8%	2.3%	2.9%	2.8%	3.3%	6.6%	4.6%	4.6%	4.2%
10 - 19 yrs	10.0%	12.6%	10.7%	13.3%	11.5%	4.8%	10.4%	12.8%	6.7%	8.0%	8.6%	10.7%	17.4%	11.3%	12.4%	10.3%
20 - 39 yrs	29.3%	33.3%	27.3%	31.1%	30.6%	13.2%	29.8%	34.4%	26.9%	25.7%	27.9%	31.2%	29.9%	23.6%	28.6%	28.7%
40 - 64 yrs	40.2%	35.6%	35.5%	37.9%	37.8%	22.5%	40.0%	39.3%	42.9%	40.1%	39.6%	38.7%	29.5%	27.4%	36.9%	36.1%
65 - 74 yrs	7.3%	5.4%	8.1%	5.5%	5.6%	27.4%	6.8%	4.7%	8.2%	8.5%	7.9%	6.7%	5.9%	7.8%	6.8%	8.8%
75 - 84 yrs	3.8%	2.1%	6.8%	2.1%	2.9%	18.7%	3.8%	2.0%	5.2%	5.4%	4.6%	3.8%	3.3%	7.3%	5.0%	5.1%
85 + yrs	4.7%	1.2%	5.8%	1.0%	3.7%	9.3%	3.1%	0.8%	6.2%	7.5%	6.8%	3.3%	2.4%	14.3%	4.1%	4.1%

Figure 3. Distribution of Race in Demographic Table in Extract 1 (Unique Individuals = 83,003,100)

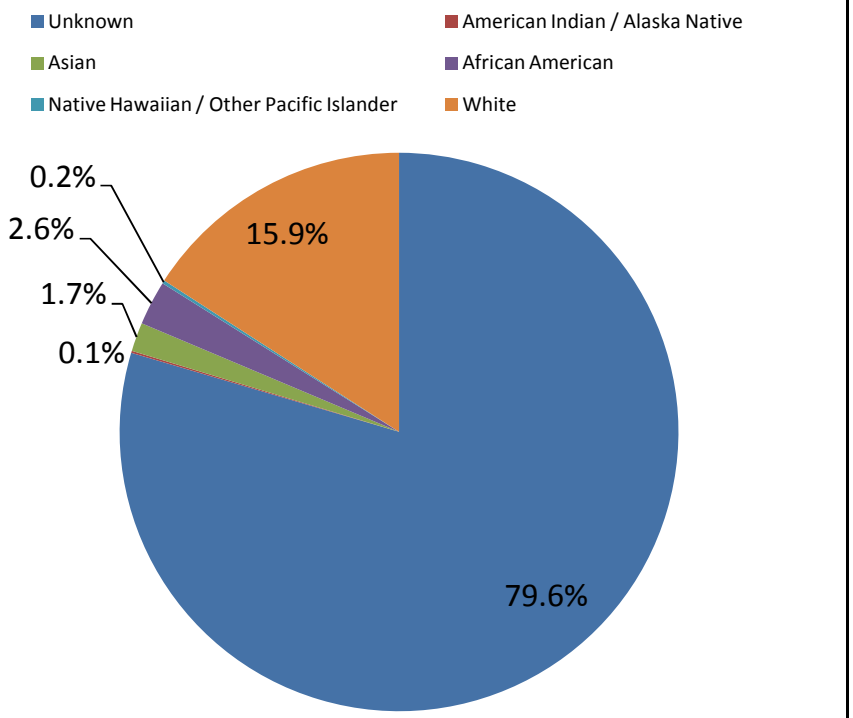


Figure 4. Distribution of Sex in MSDD Demographic Table in Extract 1 (Unique Individuals = 83,003,100)

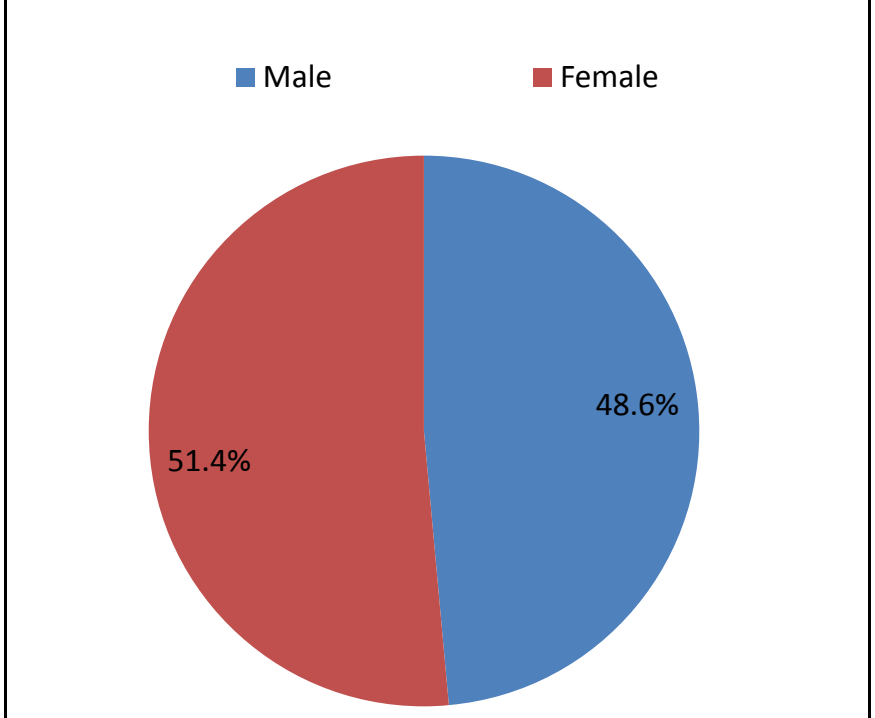


Figure 5. Distribution of Hispanic in Demographic Table in Extract 1 (Unique Individuals = 83,003,100)

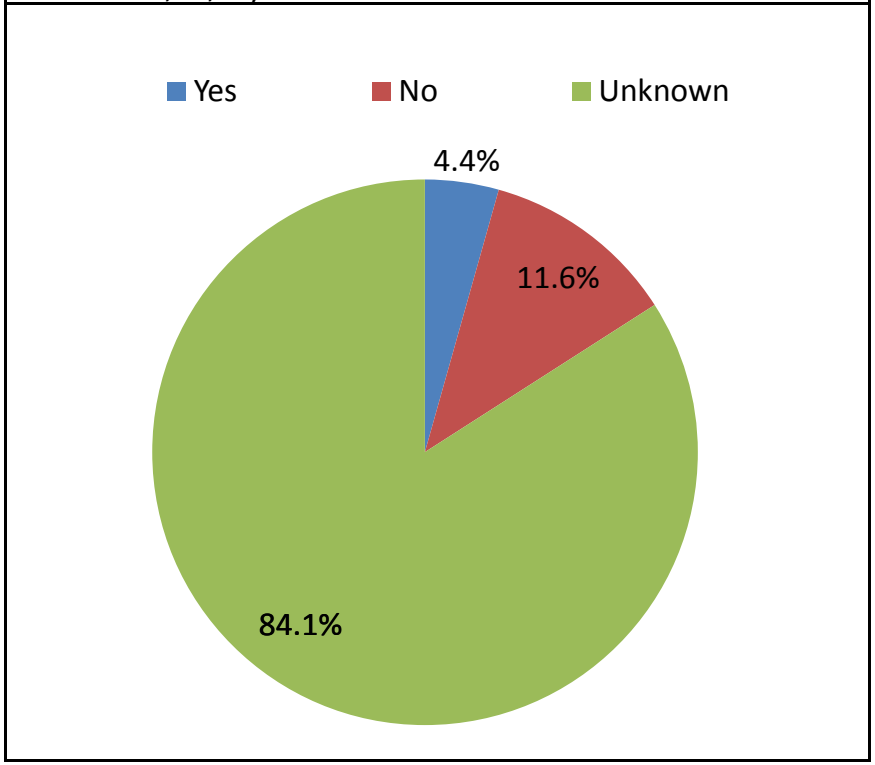


Figure 6. Distribution of Age Group in MSDD Demographic Table in Extract 1 (Unique Individuals = 83,003,100)

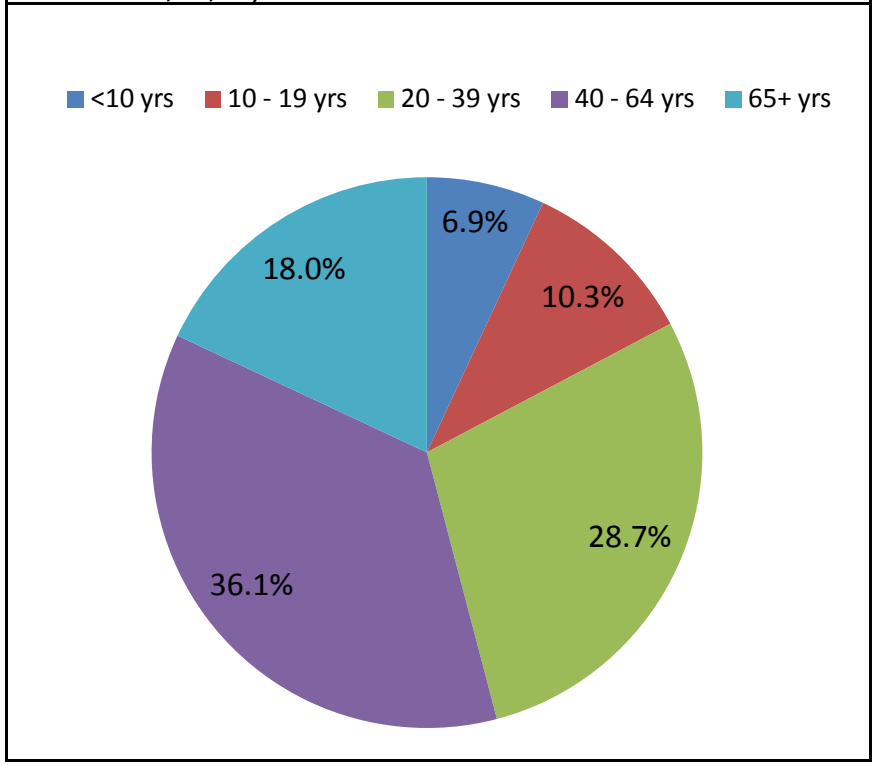


Figure 7. Distribution of Age Group in Demographic Table in Extract 1 (Unique Individuals = 83,003,100), by Data Partner

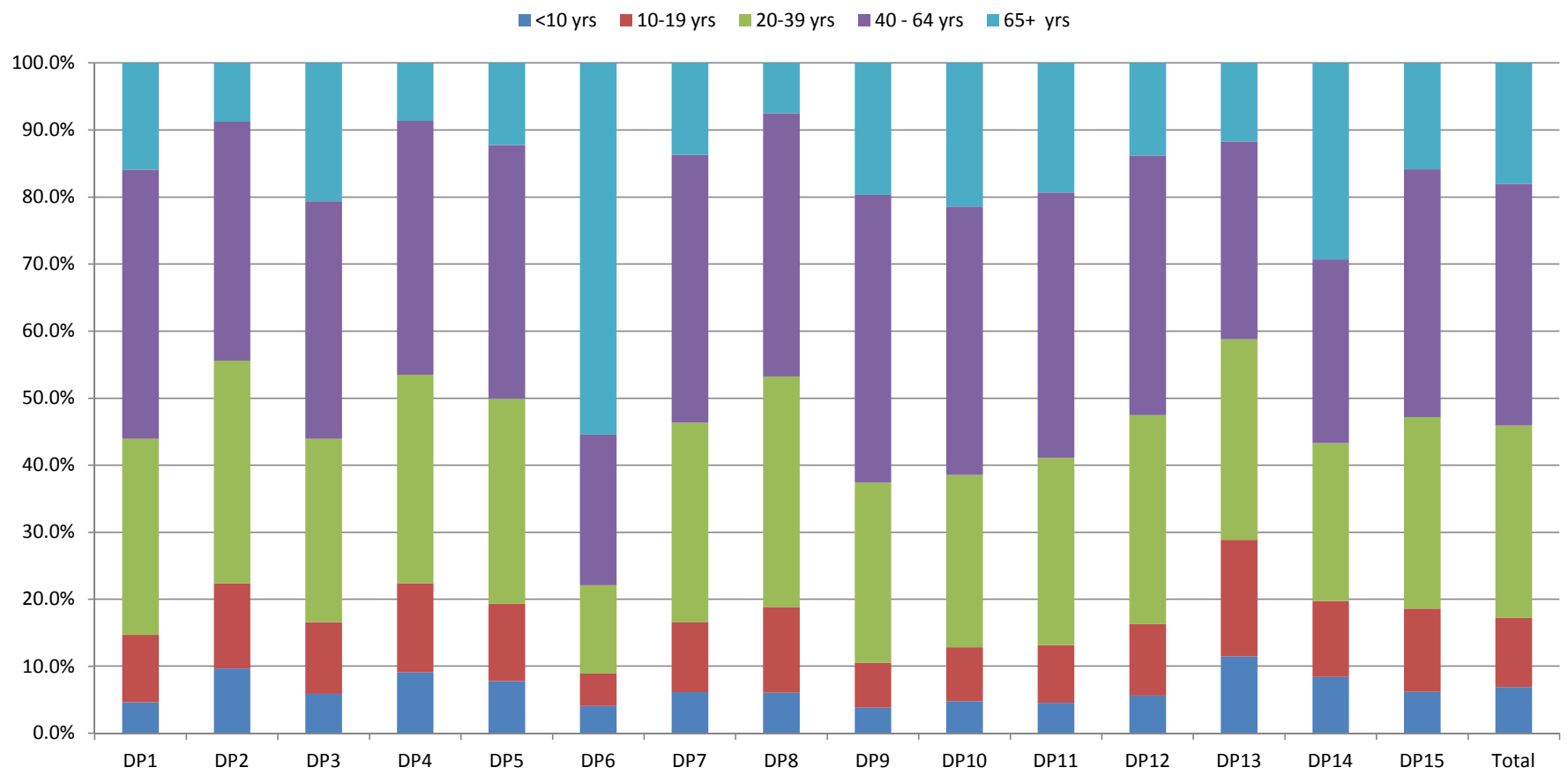


Figure 8. Percentage of Persons in Demographic Table, Aged 0-4 Weeks in Extract 1 (Unique Individuals = 31,028), by Data Partner

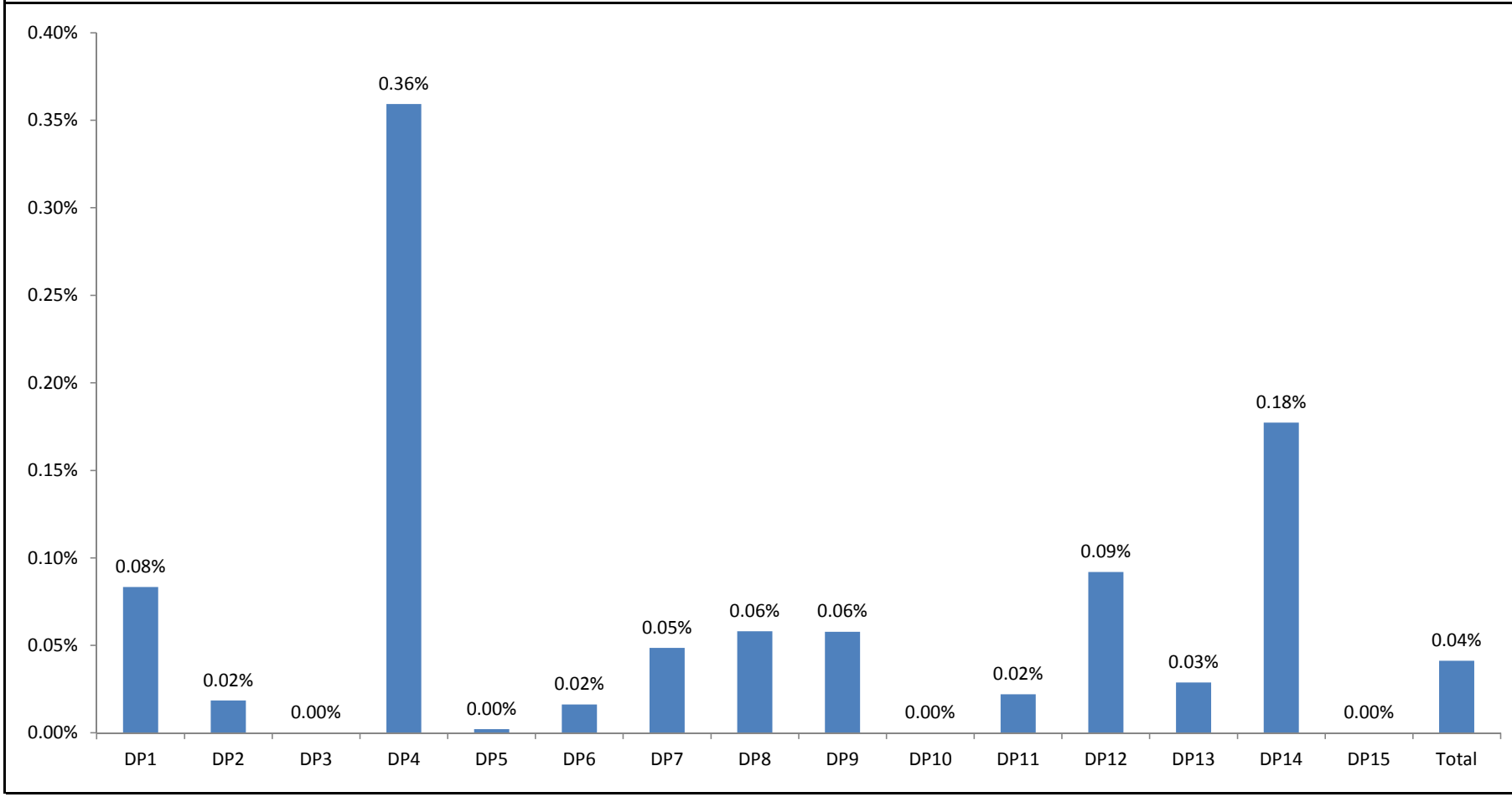


Figure 9. Percentage of Persons in Demographic Table, Aged 5-52 Weeks in Extract 1 (Unique Individuals = 236,309), by Data Partner

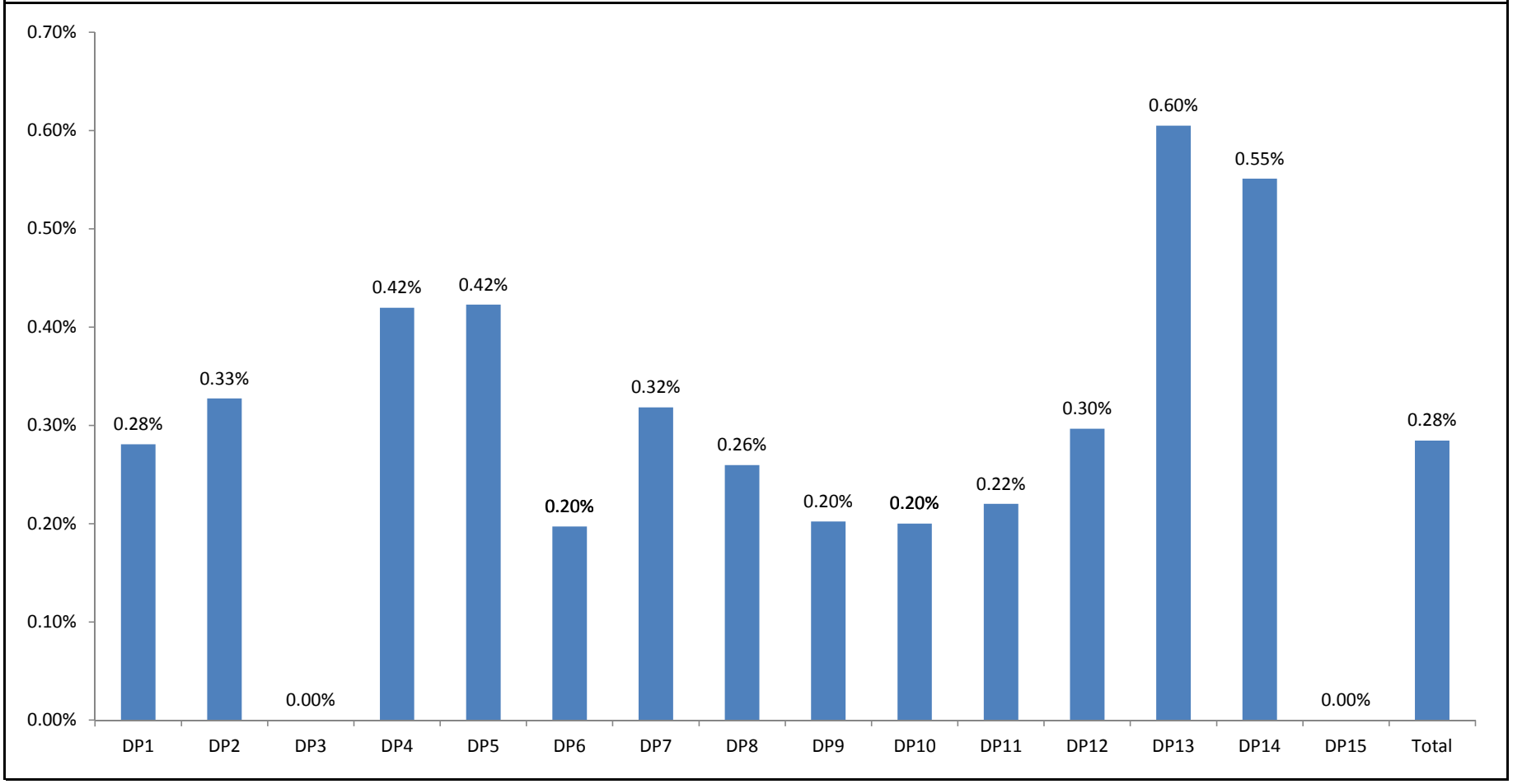


Figure 10. Percentage of Persons in Demographic Table, Aged 1-4 Years in Extract 1 (Unique Individuals = 1,951,582), by Data Partner

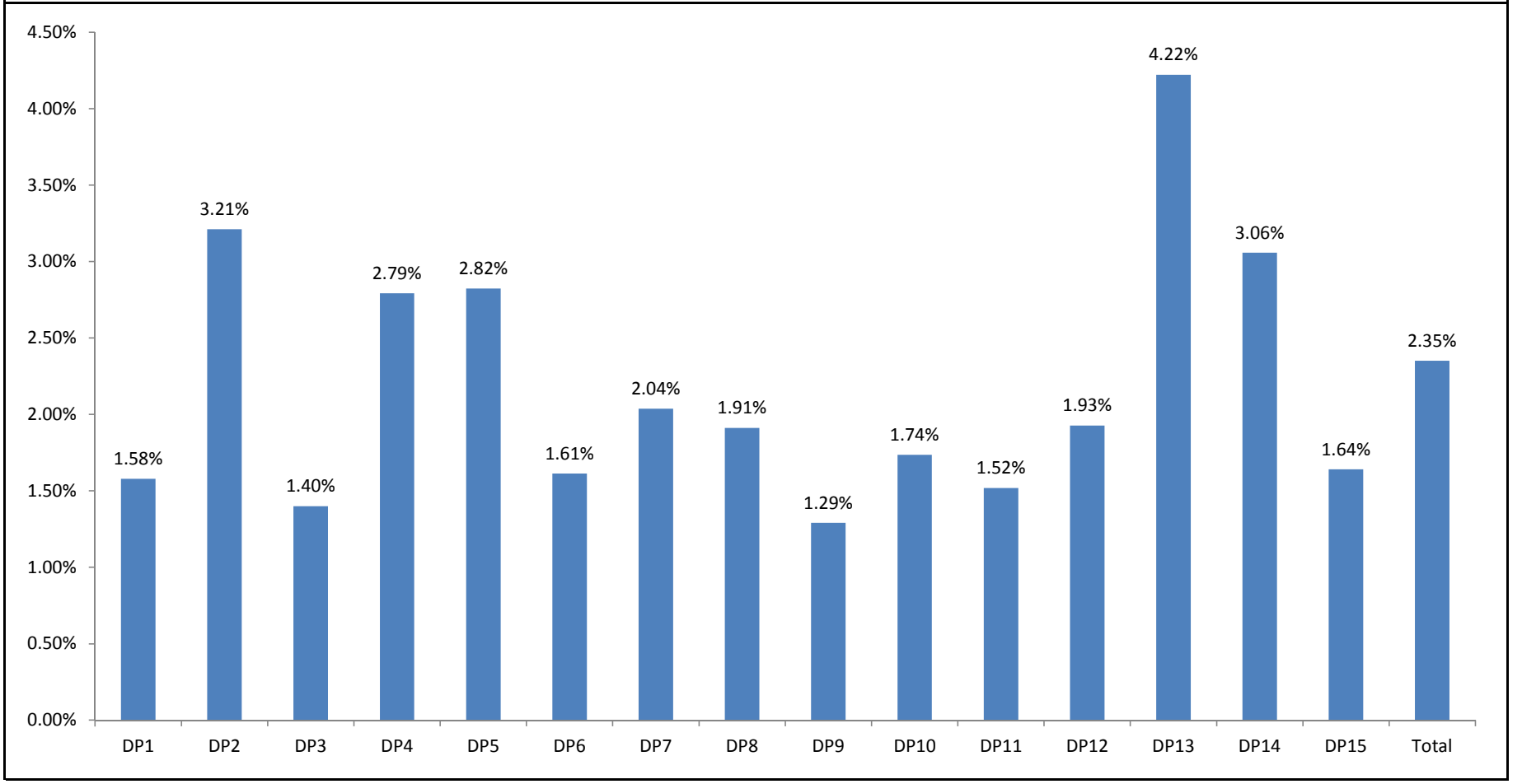


Figure 11. Percentage of Persons in Demographic Table, Aged 5-9 Years in Extract 1 (Unique Individuals = 3,518,665), by Data Partner

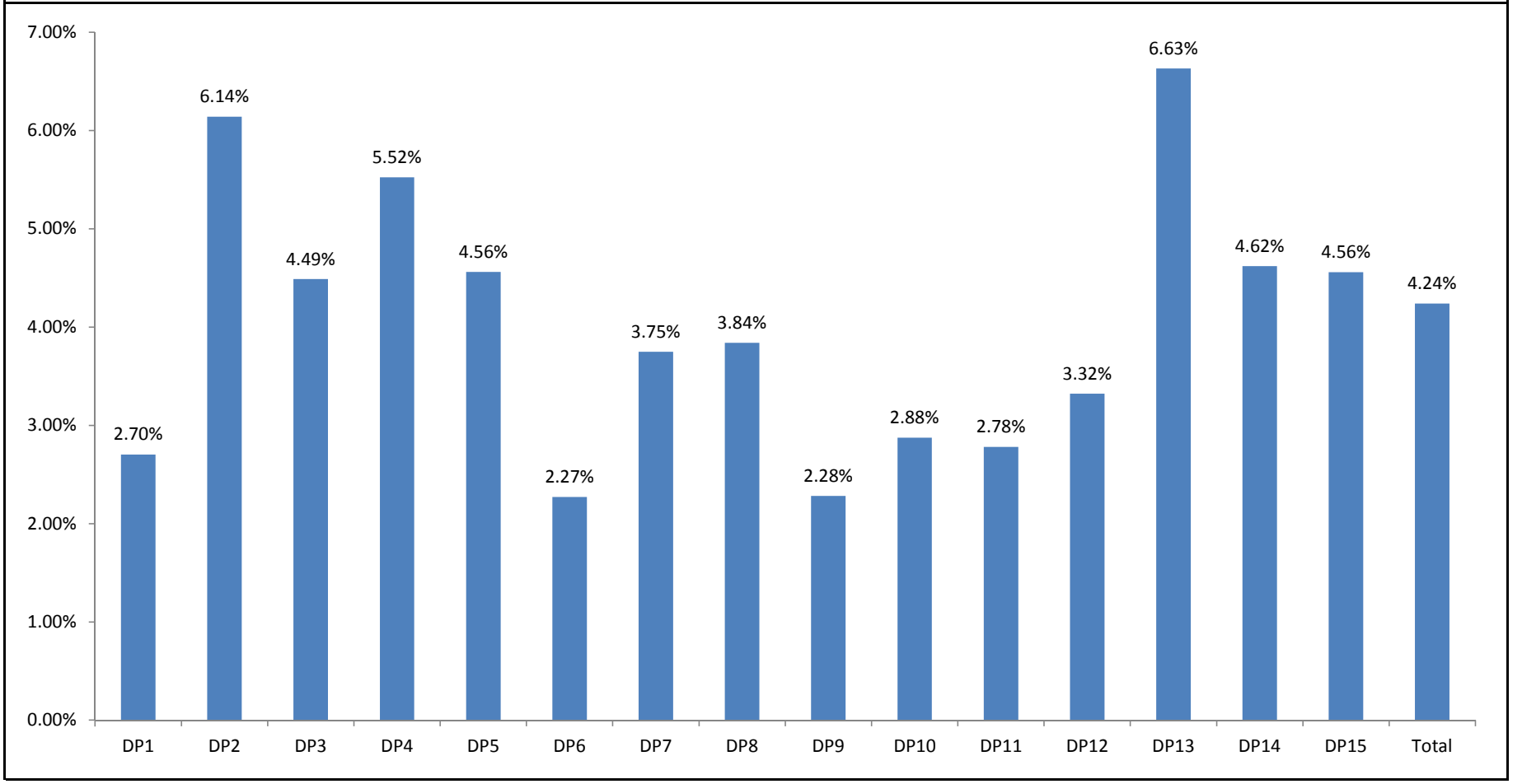


Figure 12. Percentage of Persons in Demographic Table, Aged 10-19 Years in Extract 1 (Unique Individuals = 8,587,642), by Data Partner

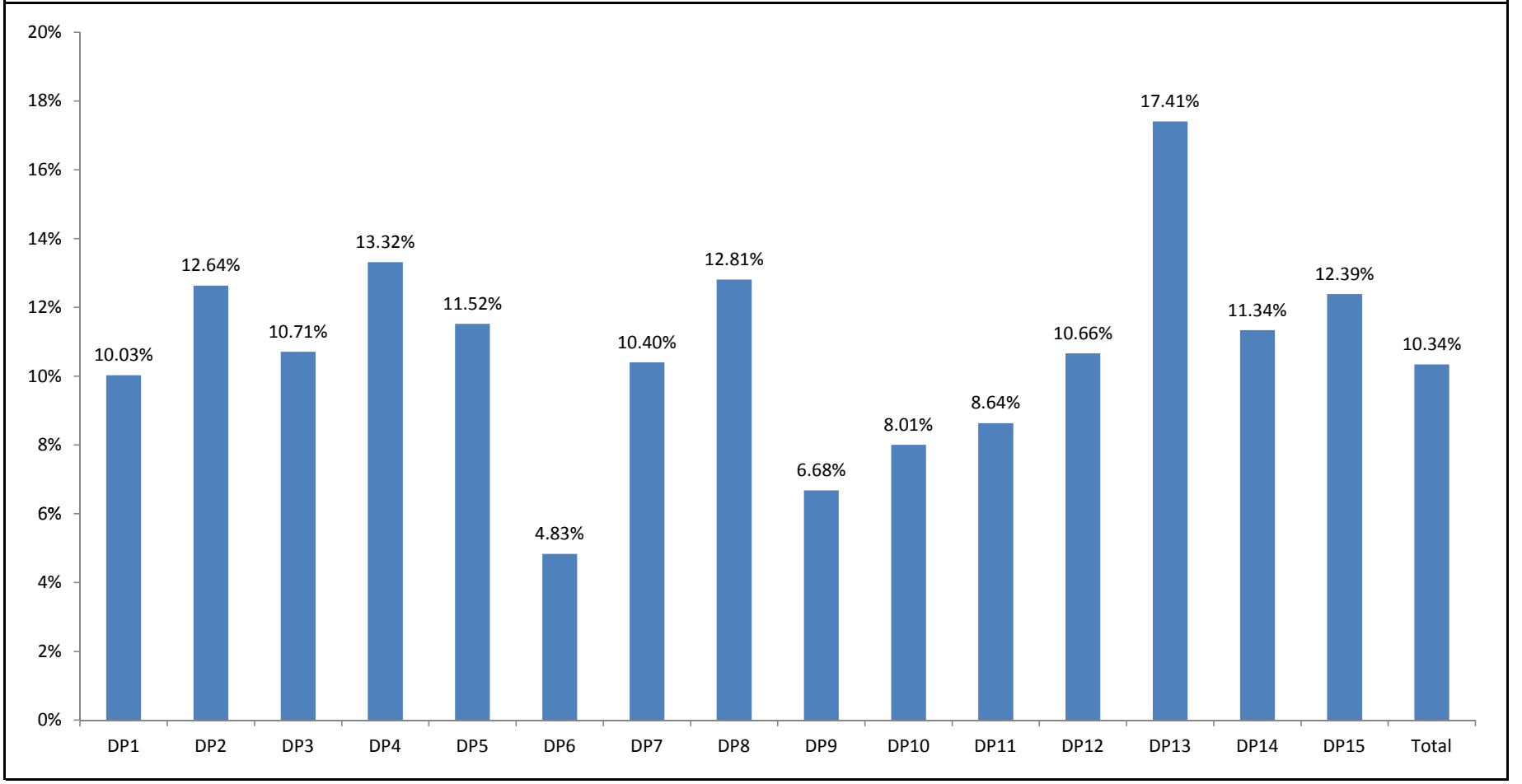


Figure 13. Percentage of Persons in Demographic Table, Aged 20-39 Years in Extract 1 (Unique Individuals = 23,794,453), by Data Partner

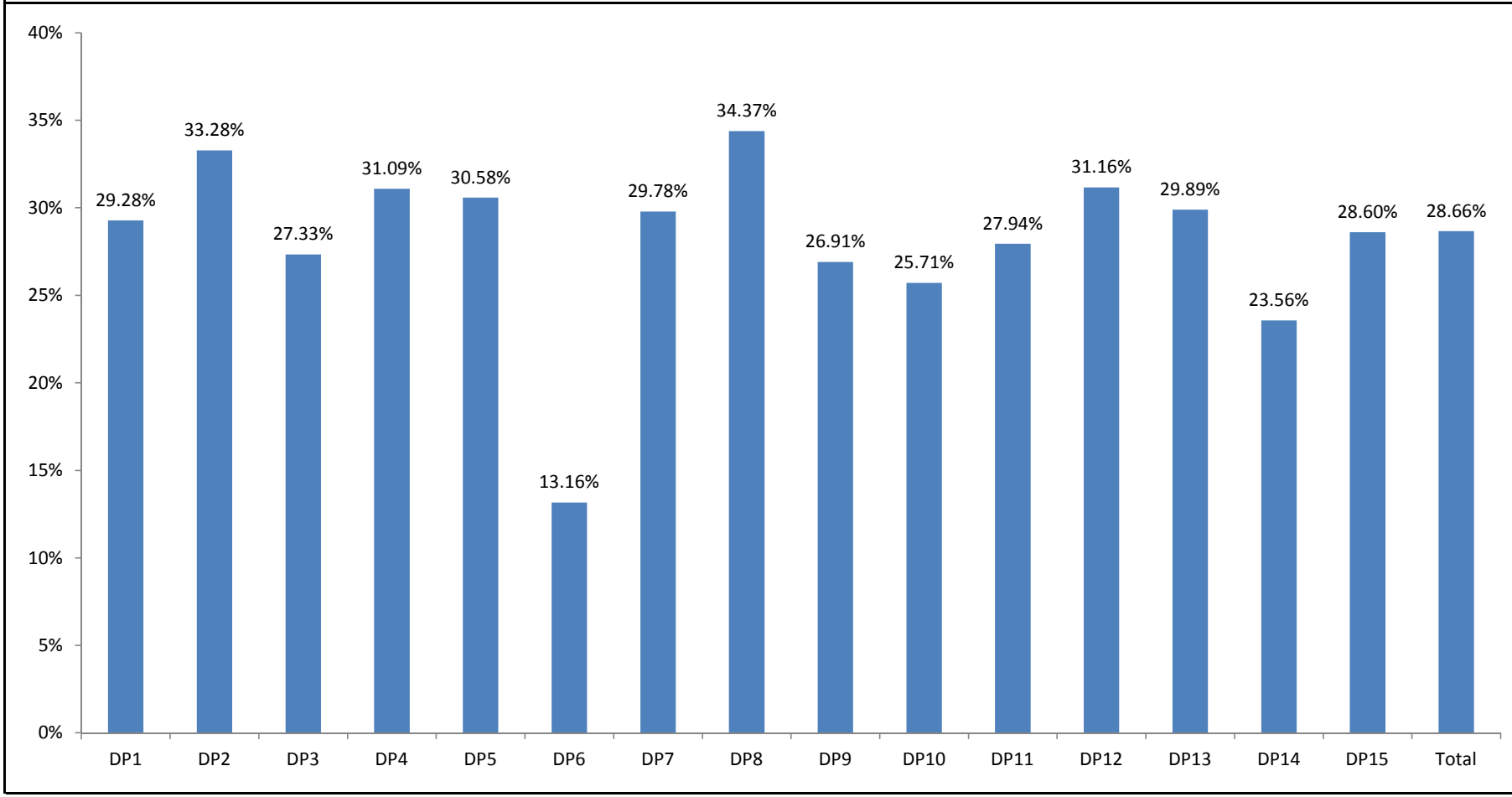


Figure 14. Percentage of Persons in Demographic Table, Aged 40-64 Years in Extract 1 (Unique Individuals = 29,934,809), by Data Partner

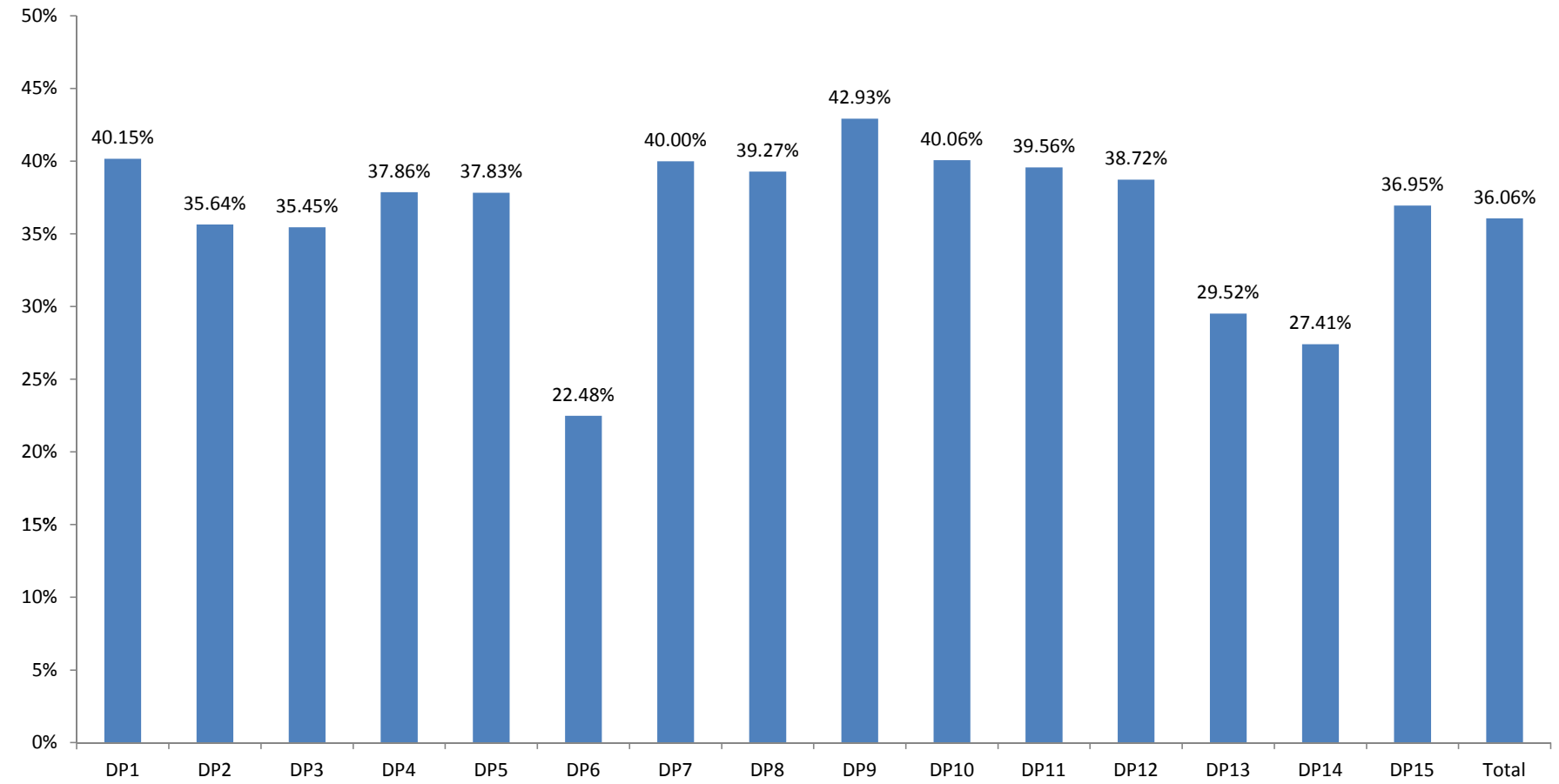


Figure 15. Percentage of Persons in Demographic Table, Aged 65-74 Years in Extract 1 (Unique Individuals = 7,288,220), by Data Partner

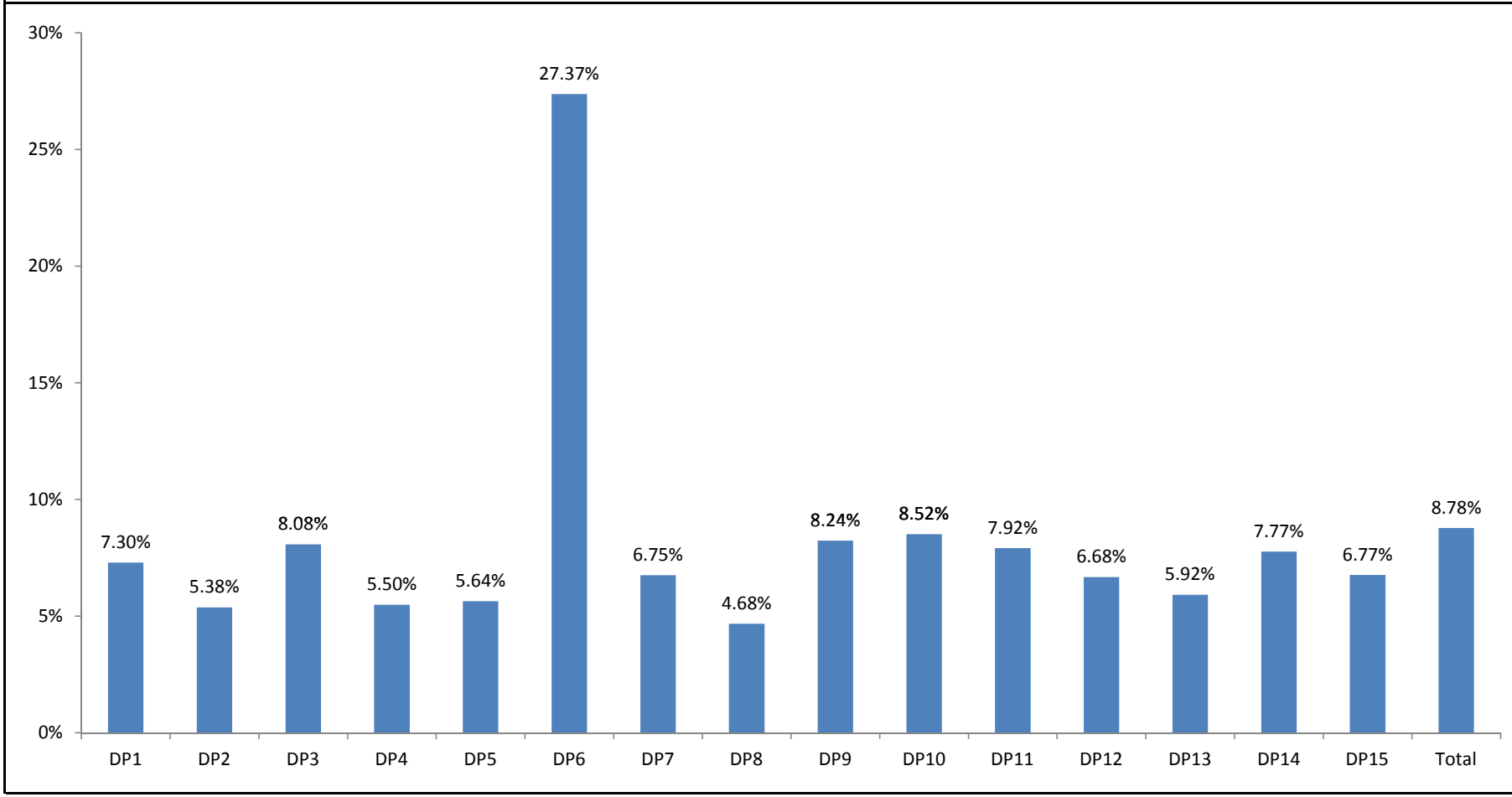


Figure 16. Percentage of Persons in Demographic Table, Aged 75-84 Years in Extract 1 (Unique Individuals = 4,244,913), by Data Partner

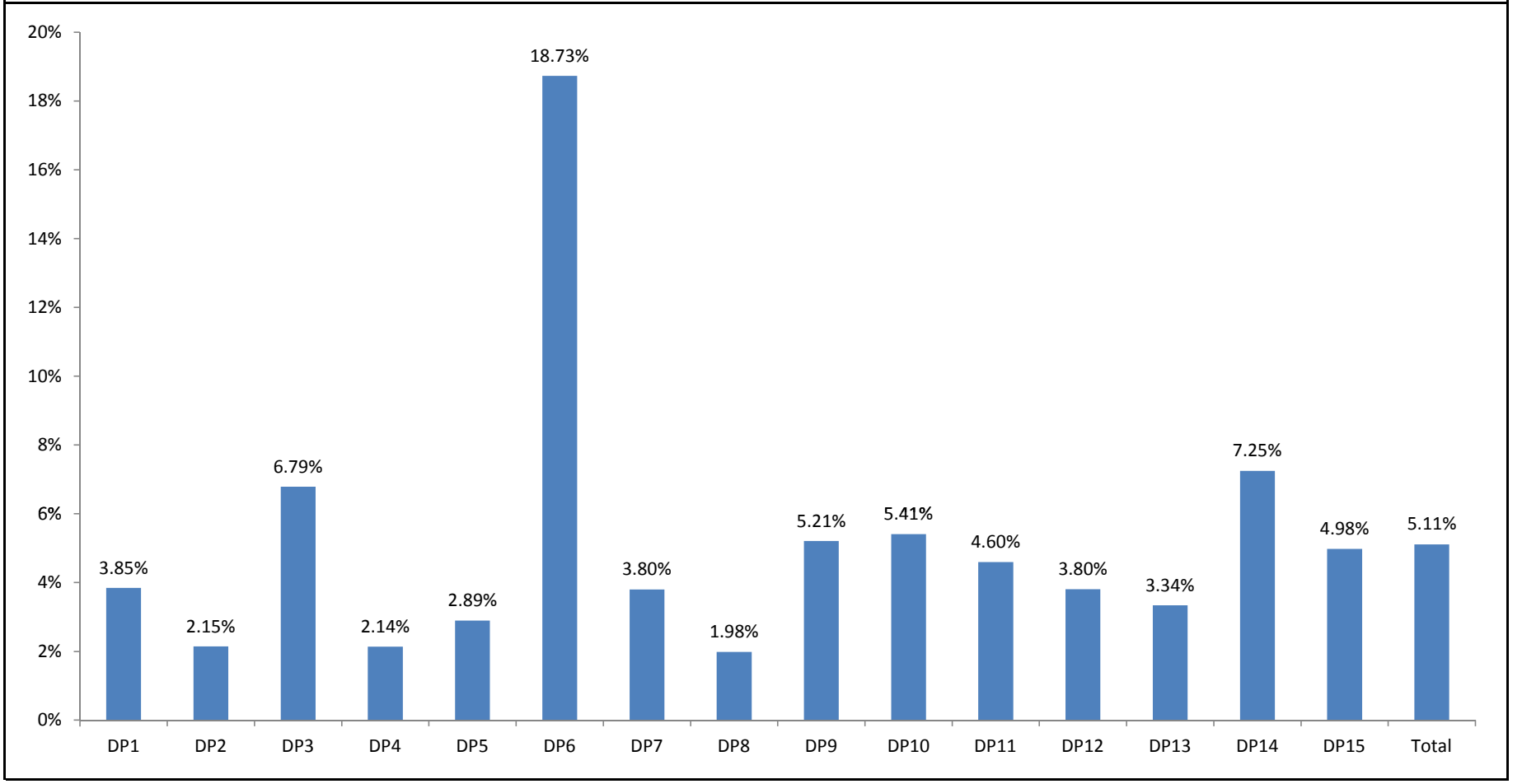


Figure 17. Percentage of Persons in Demographic Table, Aged 85+ Years in Extract 1 (Unique Individuals = 3,432,937), by Data Partner

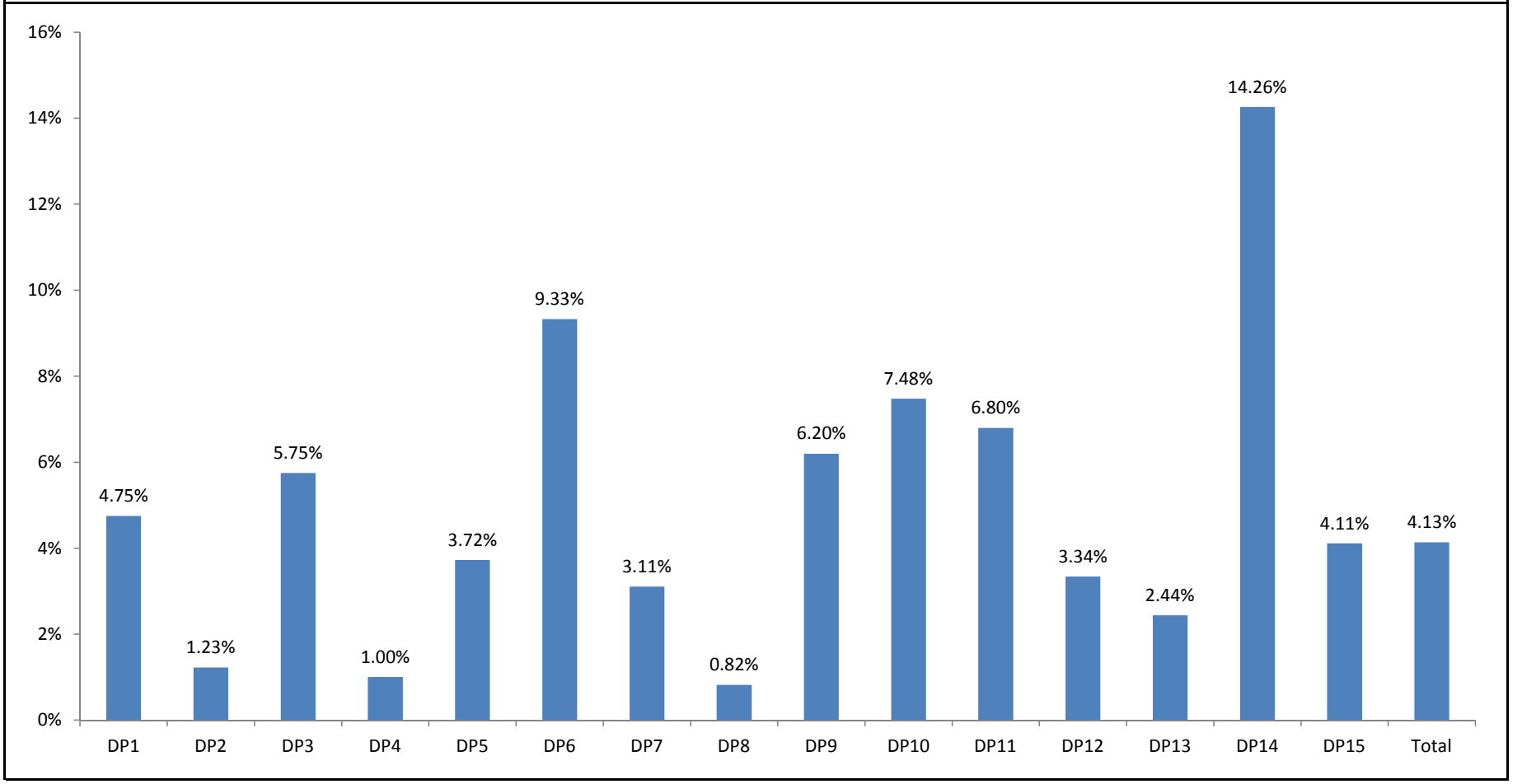


Table 3. Snapshot of the Mini-Sentinel Distributed Database Dispensing Table in Extract 1 (Unique patients = 47,026,303)

	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10	DP11	DP12	DP13	DP14	DP15	Total
Dispensings per Patient																
2000	12.51	---	13.63	11.26	10.71	---	11.40	9.94	9.93	10.05	11.67	9.00	12.70	12.06	12.00	10.38
2001	13.13	---	14.82	12.25	11.18	---	11.70	10.75	10.24	10.20	13.10	9.43	13.24	12.52	12.00	10.81
2002	13.29	---	15.44	12.86	11.65	---	11.98	11.33	10.66	10.14	13.49	9.45	13.50	13.16	14.69	10.99
2003	13.49	---	15.92	12.64	11.96	---	12.05	11.05	10.75	10.15	13.34	9.51	12.62	12.63	14.90	10.98
2004	14.30	13.44	16.25	13.16	12.27	---	12.21	11.14	10.87	10.16	13.56	9.44	13.44	13.41	14.63	12.33
2005	14.65	13.38	16.33	13.33	12.61	---	12.11	10.83	10.79	10.24	13.43	9.71	13.84	13.42	14.49	12.37
2006	15.04	13.62	17.64	13.86	12.86	---	12.55	11.44	10.79	10.24	13.83	9.86	14.48	13.58	16.83	12.64
2007	15.34	14.07	18.55	13.40	13.26	21.18	12.71	11.52	10.79	10.34	14.10	9.87	15.83	14.80	17.16	14.92
2008	15.32	14.14	19.04	13.57	13.21	32.49	12.90	11.37	10.65	10.46	14.20	10.08	16.03	14.77	17.02	18.23
2009	15.12	14.25	15.39	13.57	12.22	28.47	13.10	11.35	10.56	10.57	14.24	10.23	16.13	15.49	4.33	16.43
2010	10.98	2.73	---	---	---	---	8.02	6.20	8.19	---	11.14	8.47	10.26	12.53	---	6.25
Distribution of:																
Days Supplied																
MISSING	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
< 0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0	0.0%	6.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.0%	2.1%
> 0 - 33	75.0%	85.4%	72.7%	92.0%	82.3%	89.7%	45.4%	84.3%	72.7%	51.8%	70.3%	49.1%	92.0%	81.7%	89.5%	76.4%
34 - 66	5.5%	1.5%	12.9%	1.0%	3.7%	1.6%	45.5%	4.1%	5.0%	12.1%	4.2%	9.6%	1.7%	4.7%	2.9%	5.1%
67 - 99	17.2%	7.0%	8.0%	7.1%	13.6%	8.6%	8.4%	11.2%	21.9%	5.4%	23.2%	8.1%	6.2%	11.7%	7.1%	8.5%
100 - 499	2.3%	0.1%	6.2%	0.0%	0.3%	0.0%	0.8%	0.3%	0.4%	30.7%	1.8%	33.1%	0.0%	1.8%	0.5%	7.9%
>= 500	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Rx Amount																
MISSING	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	15.1%	0.0%	0.0%	0.2%
< 0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0	0.0%	6.7%	0.2%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.3%
> 0 - 40	56.5%	63.8%	53.6%	66.8%	57.5%	66.3%	37.0%	58.5%	50.1%	34.5%	45.5%	34.5%	52.8%	59.8%	60.9%	55.5%
> 40 - 100	31.9%	23.0%	39.2%	25.6%	33.0%	27.5%	43.8%	32.2%	36.3%	49.4%	39.8%	47.0%	24.7%	31.3%	29.9%	32.1%
> 100 - 999	11.4%	6.4%	6.9%	7.4%	9.2%	6.2%	18.9%	9.1%	13.3%	15.9%	14.6%	18.1%	7.3%	8.8%	8.8%	9.8%
>= 1000	0.1%	0.2%	0.2%	0.2%	0.2%	0.0%	0.3%	0.2%	0.3%	0.3%	0.2%	0.4%	0.1%	0.1%	0.3%	0.2%

Figure 18. Total Number of Dispensings in MSDD Dispensing Table, by Year in Extract 1 (Unique patients = 47,026,303)

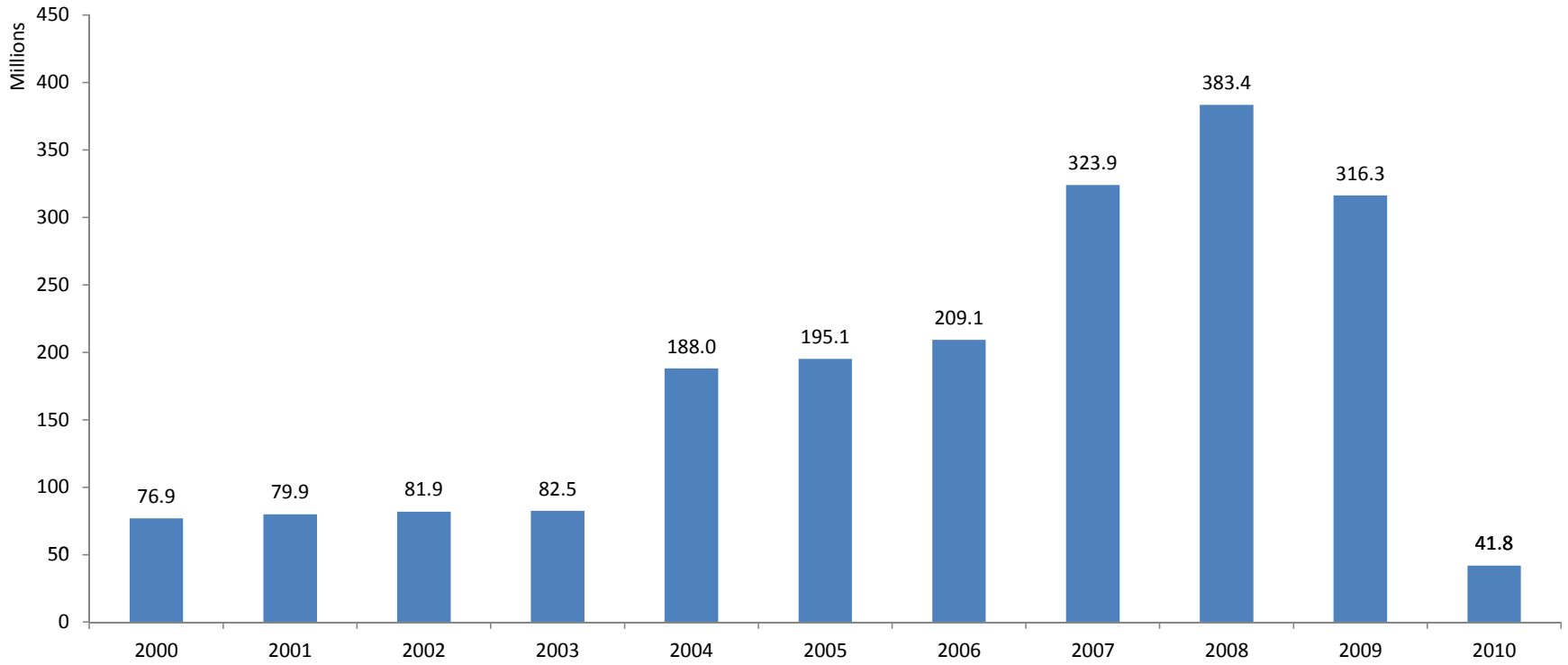


Figure 19. Average Number of Dispensings per Patient in MSDD Dispensing Table in Extract 1 (Unique patients = 47,026,303)

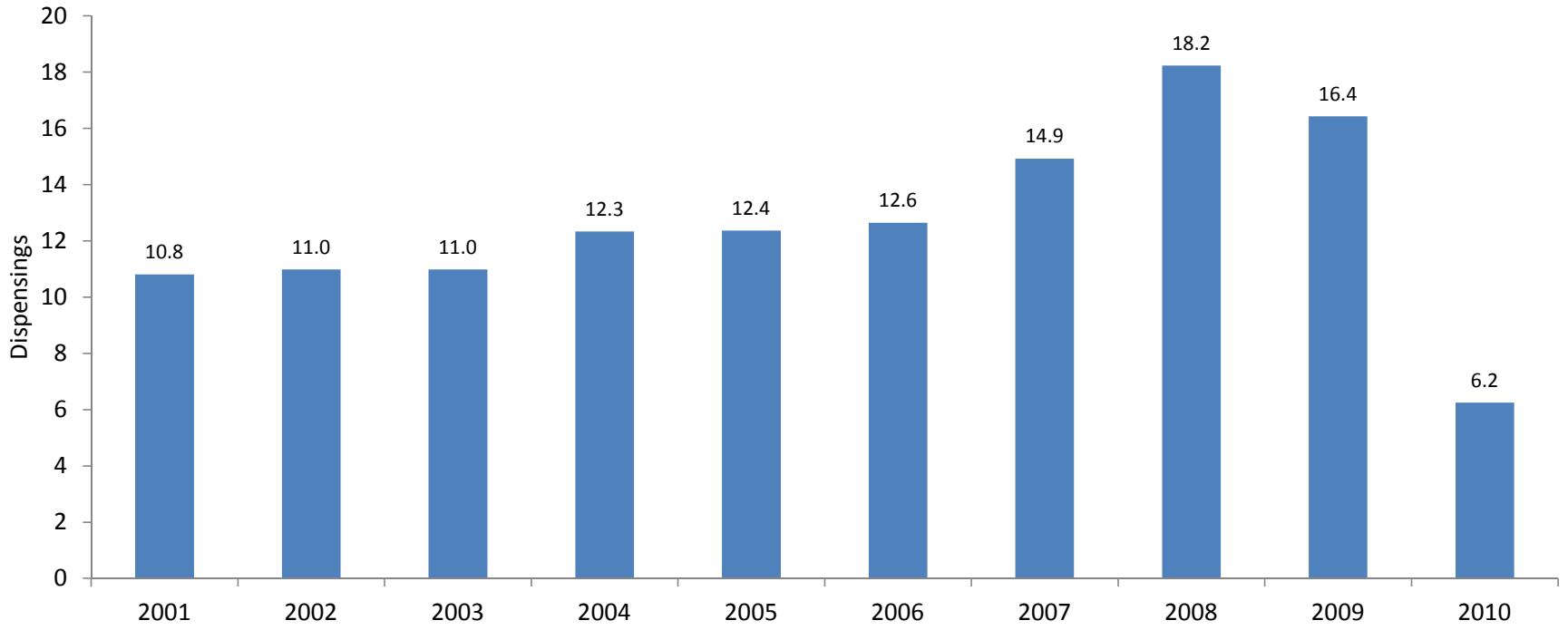


Figure 20. Average Number of Dispensings per Patient in MSDD Dispensing Table in Extract 1 (Unique patients = 47,026,303), by Data Partner and Year

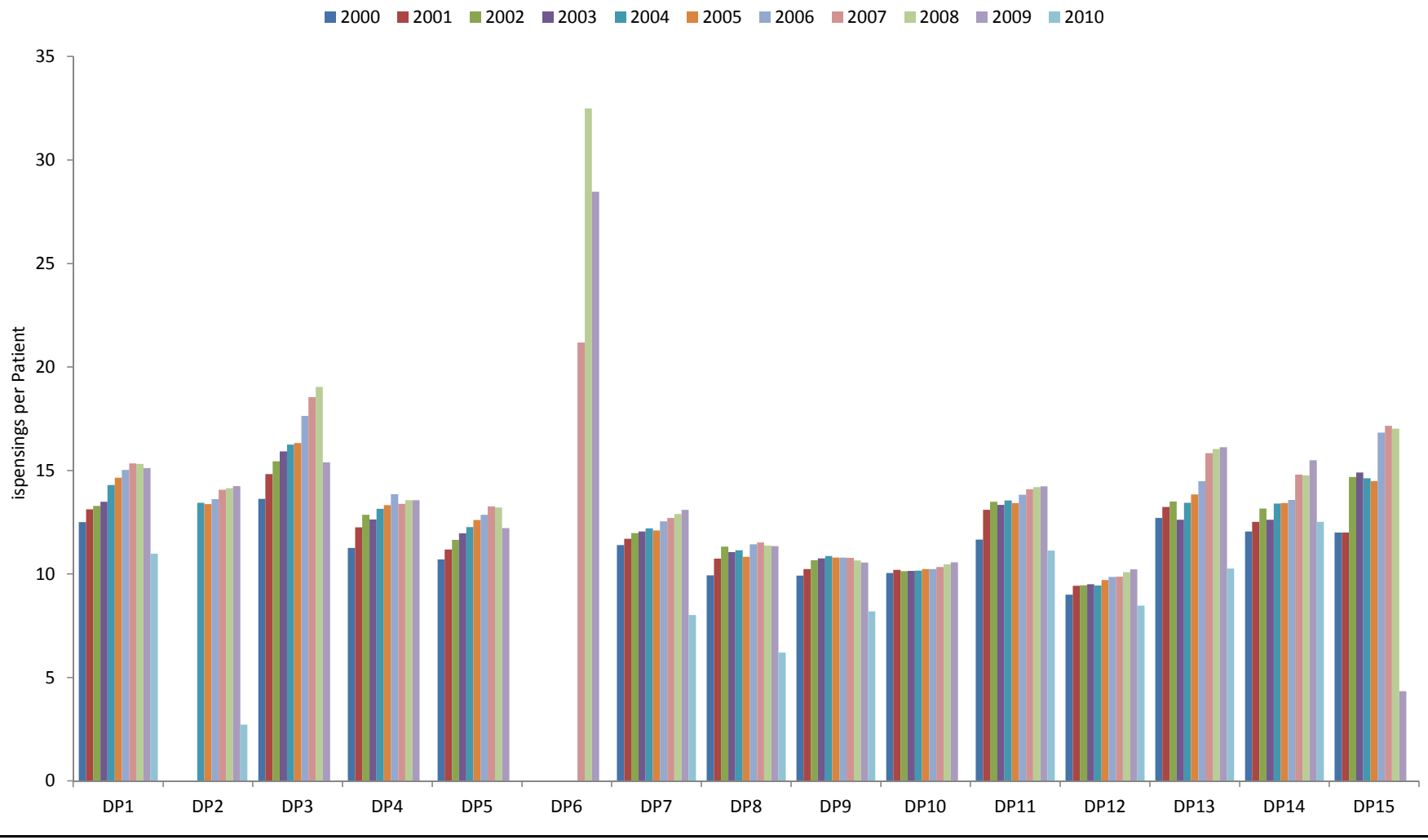


Figure 21. Distribution of Days Supplied in MSDD Dispensing Table in Extract 1 (Unique patients = 47,026,303)

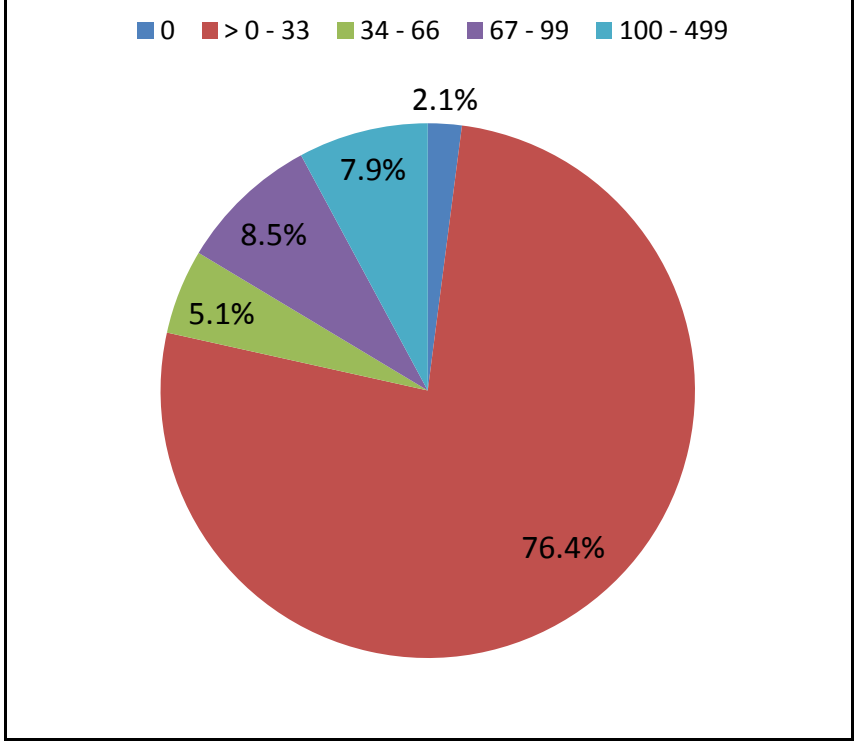


Figure 22. Distribution of Rx Amount in MSDD Dispensing Table in Extract 1 (Unique patients = 47,026,303)

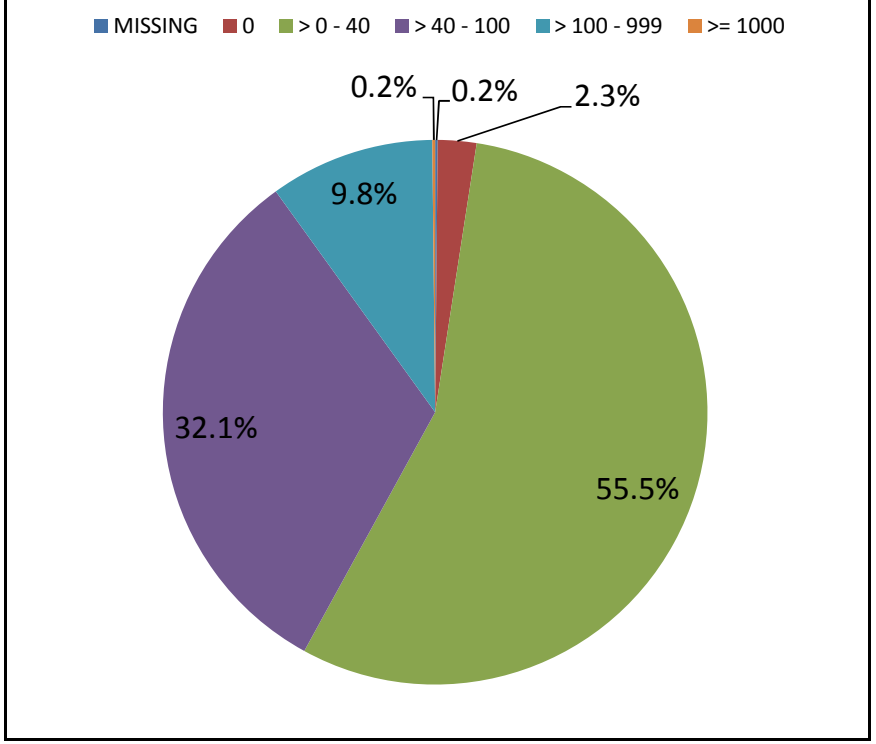


Table 4. Snapshot of the Mini-Sentinel Distributed Database Encounter Table in Extract 1 (Unique patients = 49,872,825)

	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10	DP11	DP12	DP13	DP14	DP15	Total
Encounters per Patient (All Years)	39.8	26.5	13.8	34.5	32.3	70.0	34.3	17.1	36.5	33.9	56.2	30.9	27.8	94.5	44.7	33.7
2000	5.49	-	4.74	5.11	6.77	-	4.52	3.49	8.07	-	5.84	4.99	-	6.14	10.90	5.40
2001	5.56	-	4.81	5.38	6.92	-	4.36	3.38	8.12	-	6.03	5.08	-	6.29	9.98	5.49
2002	5.85	-	4.77	5.56	7.08	-	5.13	3.27	8.29	-	6.24	5.17	-	6.42	10.43	5.67
2003	5.65	-	4.77	5.59	7.31	-	5.09	3.33	8.18	-	6.36	5.34	-	6.68	9.92	5.75
2004	5.94	6.09	4.76	5.69	7.57	-	5.09	3.19	8.17	-	6.41	5.89	-	6.83	7.36	6.06
2005	5.99	6.02	4.65	5.78	7.76	-	5.03	3.18	9.34	-	6.39	5.94	-	6.97	8.87	6.05
2006	6.06	6.11	4.77	5.97	7.95	-	5.03	3.14	5.99	-	6.31	5.56	-	7.93	8.44	6.03
2007	6.16	6.27	4.83	5.80	8.13	21.26	5.19	3.21	3.87	-	6.71	5.28	-	8.50	9.25	8.20
2008	6.28	6.37	4.29	5.94	8.26	34.26	5.22	3.45	3.97	-	7.13	6.51	-	8.73	10.40	11.55
2009	6.22	6.43	2.09	6.11	7.75	31.26	5.31	4.37	4.20	-	7.49	5.21	-	8.16	3.43	10.88
2010	4.50	2.67	-	-	-	6.44	-	2.77	3.70	-	5.38	4.70	-	6.84	-	3.84
Encounter Type																
Inpatient Hospital Stay	1.4%	1.2%	4.3%	0.8%	1.1%	7.2%	1.7%	3.6%	1.1%	1.4%	0.7%	2.6%	---	2.0%	4.4%	2.9%
Non-Acute Institutional Stay	0.5%	0.9%	0.0%	0.2%	0.1%	37.3%	0.4%	0.1%	0.0%	0.1%	0.1%	1.1%	---	0.0%	0.4%	9.8%
Emergency Department	3.1%	3.8%	5.4%	2.7%	1.9%	3.9%	2.4%	2.6%	0.7%	3.6%	1.7%	4.7%	---	3.3%	2.0%	3.6%
Ambulatory Visit	73.5%	77.3%	90.0%	78.2%	96.9%	37.9%	60.0%	83.8%	84.7%	81.3%	47.4%	91.6%	---	76.8%	89.6%	69.2%
Other Ambulatory Visit	21.5%	16.7%	0.3%	18.2%	0.0%	13.8%	35.5%	9.8%	13.4%	13.6%	50.1%	0.0%	---	17.9%	3.6%	14.5%

Figure 23. Average Number of Encounters per Patient in MSDD Encounter Table in Extract 1 (Unique patients = 49,872,825)

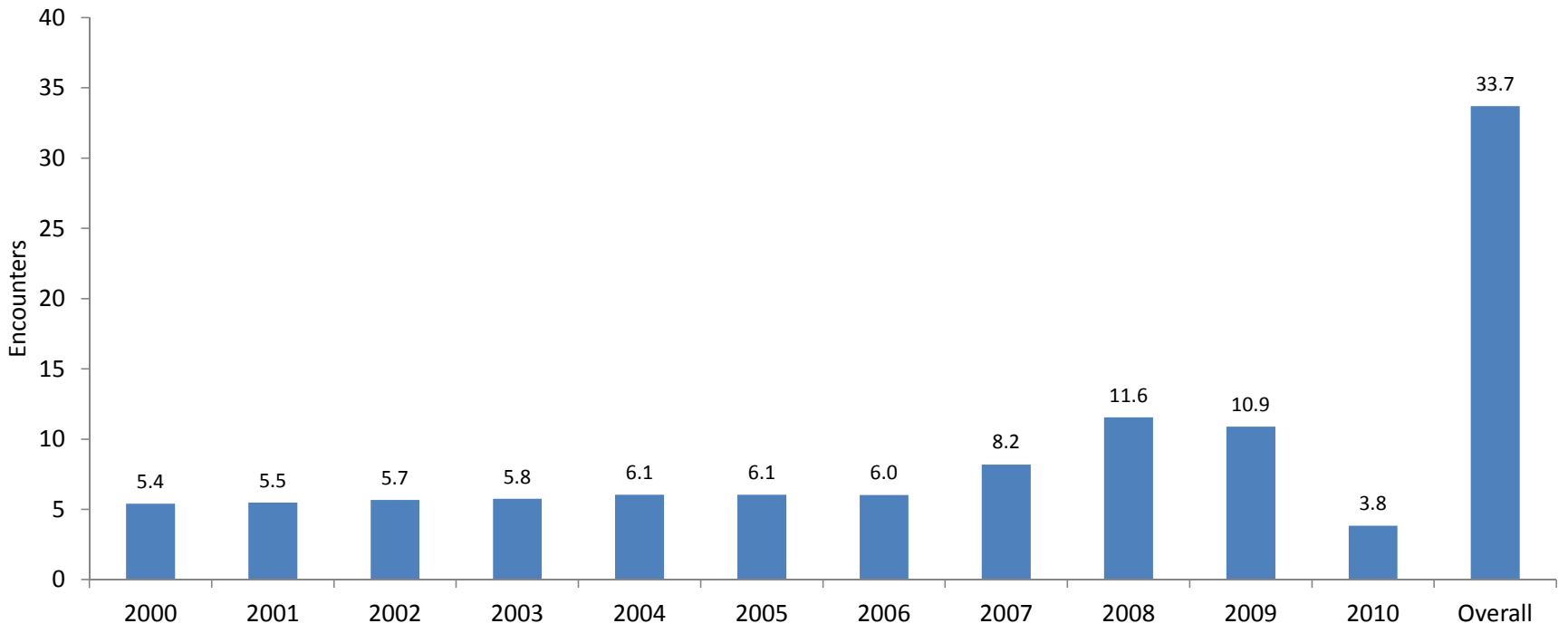


Figure 24. Distribution of Encounter Type in MSDD Encounter Table in Extract 1 (Unique patients = 49,872,825)

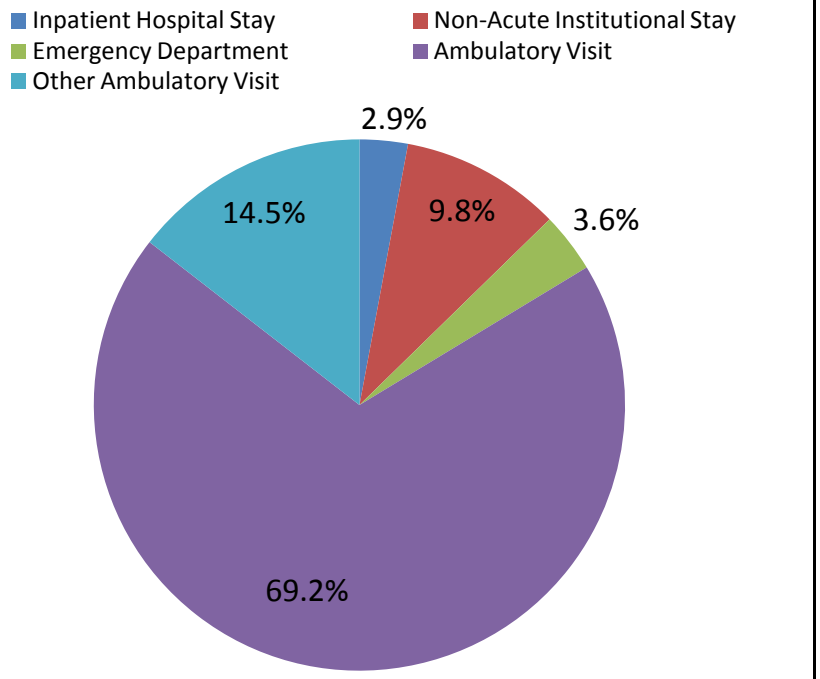
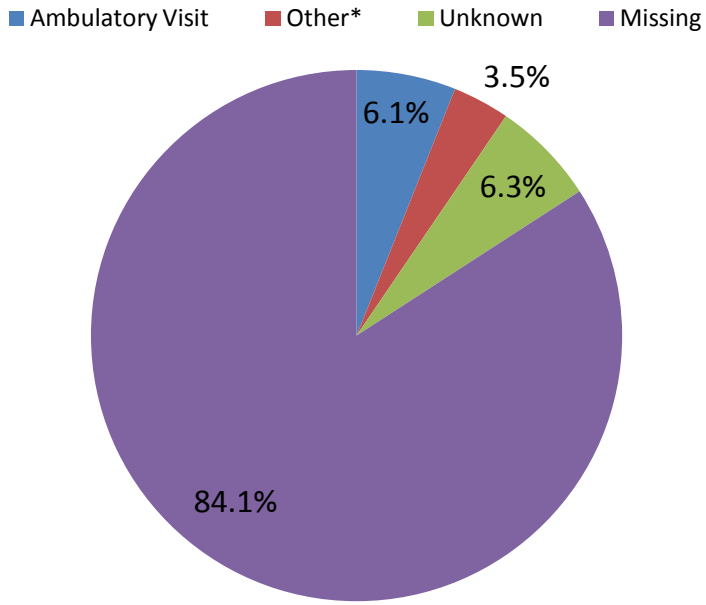


Figure 25. Distribution of Admission Source in MSDD Encounter Table in Extract 1 (Unique patients = 49,872,825)



*includes ambulatory visit, emergency department, home health, hospice, home/self care, other acute inpatient hospital, nursing home (includes ICF), other, residential facility, rehabilitation facility, skilled nursing facility

Figure 26. Distribution of Discharge Disposition in MSDD Encounter Table in Extract 1 (Unique patients = 49,872,825)

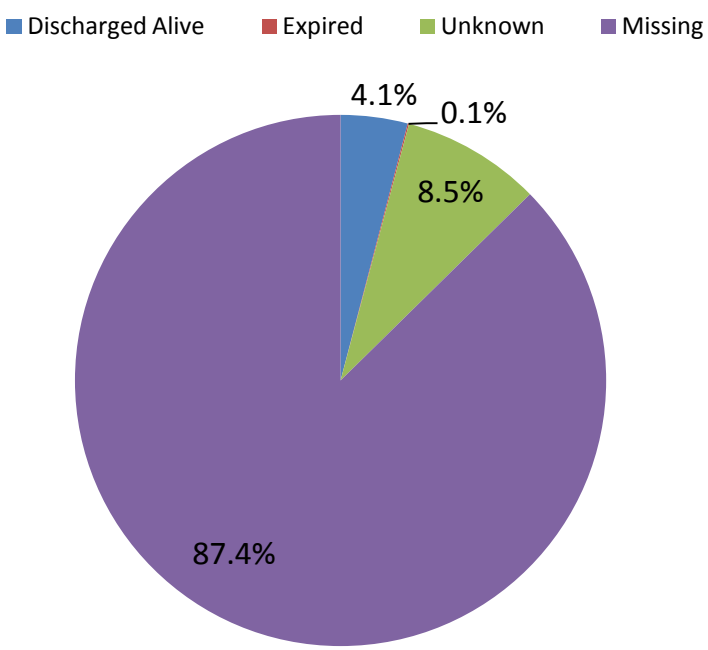
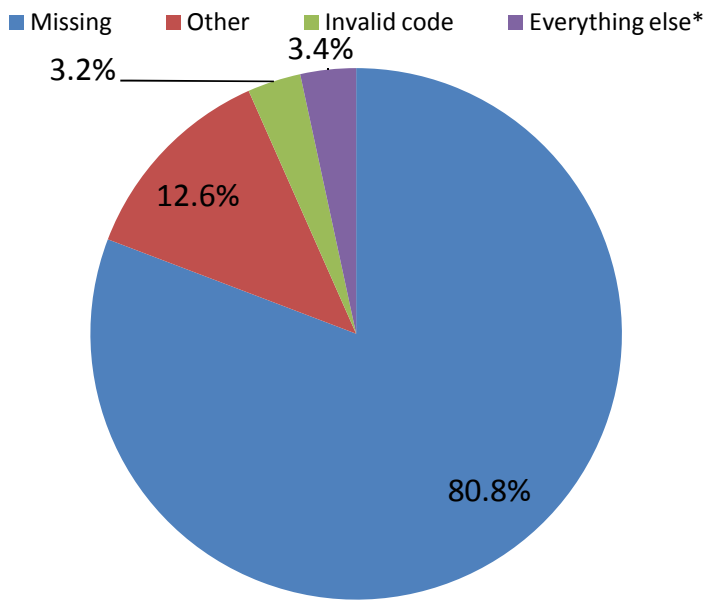
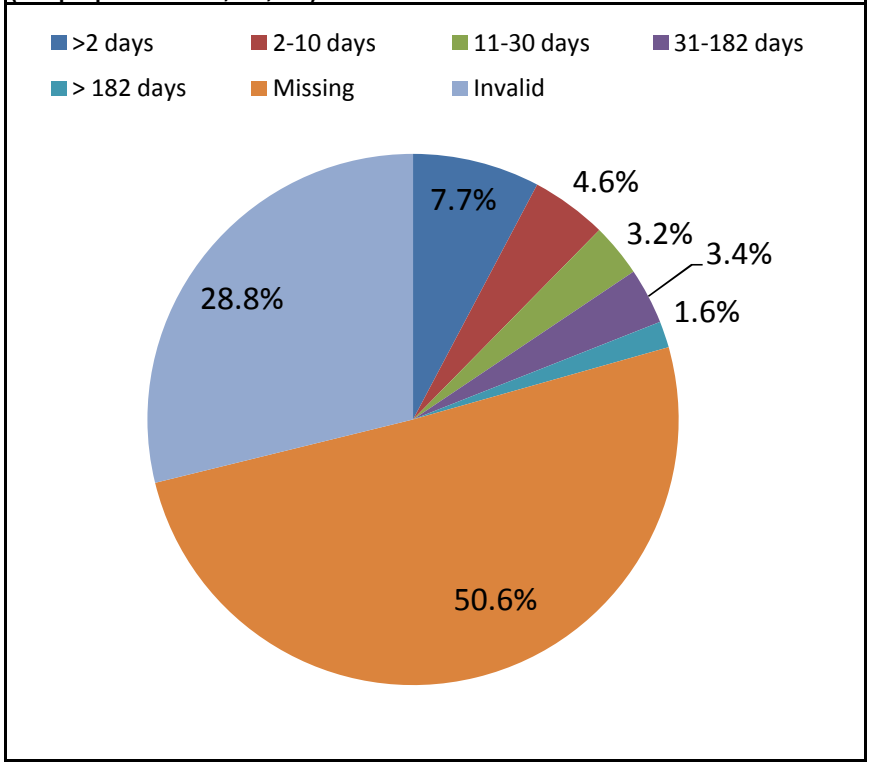


Figure 27. Distribution of Discharge Status in MSDD Encounter Table in Extract 1 (Unique patients = 49,872,825)



*includes adult foster home, assisted living facility, against medical advice, absent without leave, expired, home health, hospice, home/self care, other acute inpatient hospital, nursing home (includes ICF), residential facility, rehabilitation facility, still in hospital, skilled nursing facility

Figure 28. Distribution of Length of Stay in MSDD Encounter Table in Extract 1 (Unique patients = 49,872,825)



Mini-Sentinel

Table 5. Snapshot of the Mini-Sentinel Distributed Database Diagnosis Table in Extract 1 (Unique patients = 49,170,130)

	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10	DP11	DP12	DP13	DP14	DP15	Total
Dx per Encounter (Overall)	1.8	1.6	2.5	2.0	2.0	3.1	2.4	2.8	2.9	3.4	2.0	2.3	2.0	2.2	1.7	2.2
By Year																
2000	1.4	---	2.6	1.7	1.7	---	1.8	1.7	0.6	1.8	0.9	0.2	---	1.9	1.3	1.3
2001	1.5	---	2.7	1.8	1.8	---	1.7	1.7	0.7	2.3	0.9	1.0	---	1.9	1.4	1.6
2002	1.5	---	2.7	1.8	1.8	---	1.7	1.8	0.7	2.5	0.9	1.2	---	2.0	1.5	1.7
2003	1.4	---	2.8	1.9	1.8	---	1.7	1.9	0.8	2.8	1.0	1.2	---	2.0	1.5	1.8
2004	1.5	1.4	2.8	2.0	1.9	---	1.8	2.1	0.8	3.0	1.0	1.1	---	2.1	1.6	1.6
2005	1.6	1.4	2.9	2.1	1.9	---	1.3	2.3	0.9	3.0	1.0	1.2	---	2.1	1.6	1.6
2006	1.6	1.5	3.0	2.2	1.9	---	1.4	2.7	1.4	2.3	1.0	1.5	---	1.9	2.2	1.7
2007	1.7	1.5	3.0	2.2	2.0	2.0	1.5	3.2	1.9	1.8	1.1	1.9	---	2.0	2.1	1.8
2008	1.7	1.5	2.9	2.2	2.0	1.8	1.6	4.7	1.9	1.8	1.1	1.7	---	2.0	2.1	1.8
2009	1.8	1.8	0.0	2.3	2.1	2.1	1.6	6.6	1.8	2.0	1.1	2.3	---	2.3	2.2	2.0
2010	1.8	1.8	---	---	---	5.7	---	5.4	1.6	2.1	1.2	2.4	---	2.3	---	2.0
By Encounter Type																
Inpatient Stay	7.8	7.3	7.8	12.9	9.5	5.1	6.8	12.3	7.9	7.0	7.6	3.9	---	10.6	3.4	5.7
Non-Acute Institutional Stay	2.5	2.1	---	17.3	11.3	0.5	5.0	16.5	5.6	2.3	0.9	1.9	---	---	2.2	0.6
Emergency Department	2.9	3.4	---	10.5	7.9	0.9	---	0.0	0.0	1.1	0.0	0.5	---	0.1	0.1	0.2
Ambulatory Visit	1.8	1.5	2.5	1.9	1.8	2.5	1.9	2.6	1.1	2.4	1.7	1.3	---	2.0	1.6	1.8
Other Ambulatory Visit	0.6	1.2	2.1	1.7	---	2.3	0.5	2.8	0.4	1.3	0.2	3.1	---	1.3	2.4	1.3
Distribution of Encounter Type																
Inpatient Stay	6.9%	5.9%	12.0%	4.9%	5.4%	18.9%	7.4%	14.7%	8.0%	4.1%	5.0%	7.0%	---	10.4%	8.5%	9.6%
Non-Acute Institutional Stay	0.8%	1.3%	0.0%	1.5%	0.7%	9.2%	1.4%	0.3%	0.2%	0.1%	0.1%	1.4%	---	0.0%	0.6%	3.2%
Emergency Department	3.6%	4.1%	6.3%	6.2%	2.5%	6.7%	4.8%	4.5%	0.1%	3.6%	4.3%	8.1%	---	3.5%	2.6%	5.1%
Ambulatory Visit	81.2%	75.3%	81.4%	72.1%	91.5%	49.3%	74.2%	71.5%	86.8%	84.2%	81.0%	83.5%	---	74.7%	83.3%	71.1%
Other Ambulatory Visit	7.5%	13.5%	0.3%	15.3%	0.0%	16.0%	12.2%	8.9%	4.9%	8.0%	9.6%	0.0%	---	11.4%	5.0%	11.0%
Code Type																
09	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	98.0%	100.0%	100.0%	100.0%
10	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
11	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Missing	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Other	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.0%	0.0%	0.0%	0.0%

Figure 29. Number Diagnoses per Encounter, by Year, in MSDD Diagnosis Table in Extract 1 (Unique patients = 49,170,130)

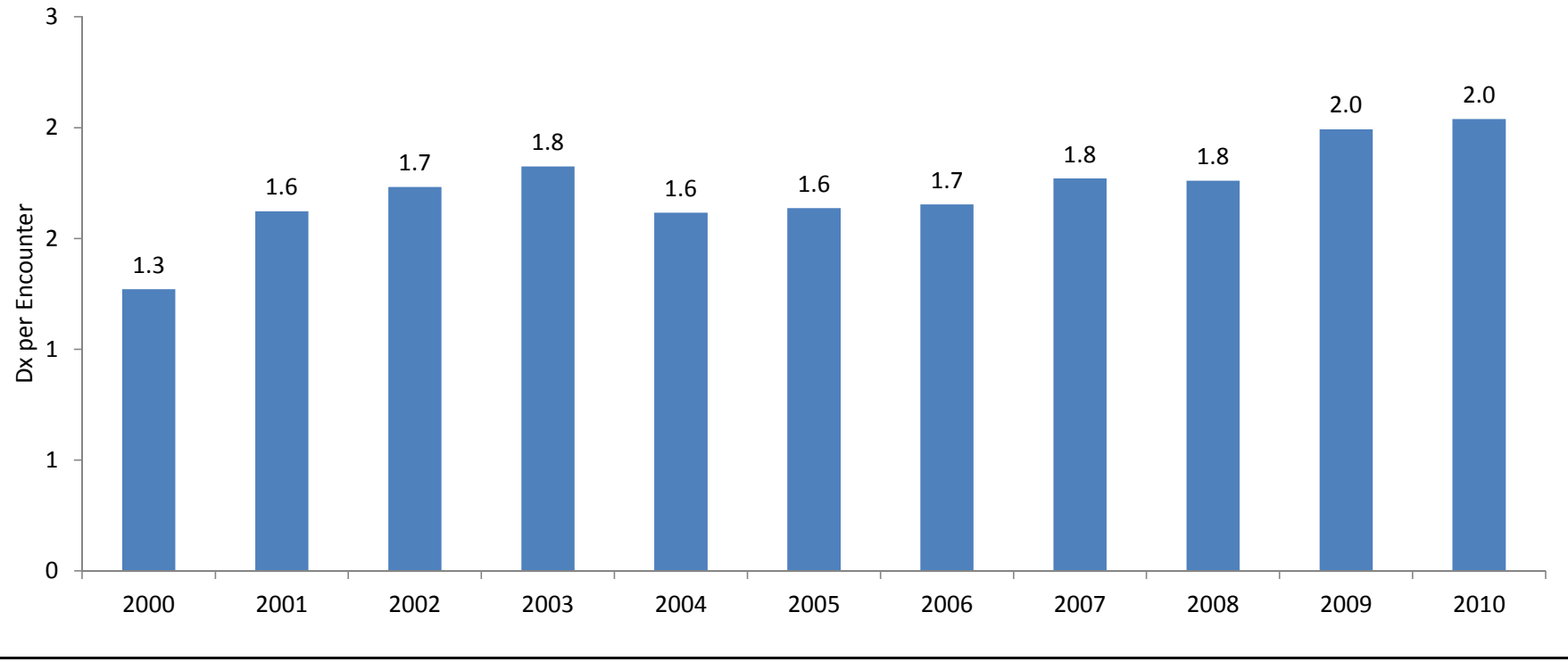


Figure 30. Distribution of Encounter Type in MSDD Diagnosis Table in Extract 1 (Unique patients = 49,170,130)

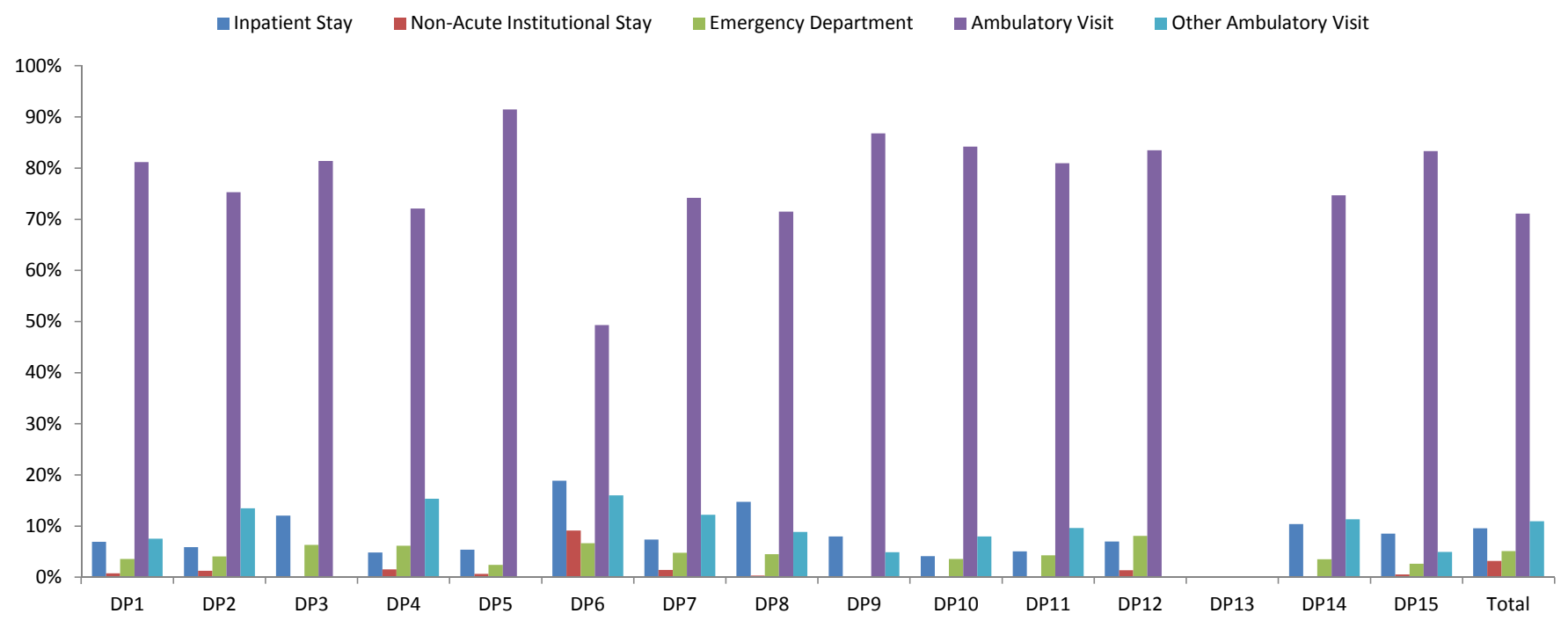


Table 6. Snapshot of the Mini-Sentinel Distributed Database Procedure Table in Extract 1 (Unique patients = 47,357,192)

	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10	DP11	DP12	DP13	DP14	DP15	Total
Px per Encounter (Overall)	2.2	2.7	2.3	2.3	3.0	6.1	3.2	2.7	6.5	2.4	2.3	2.0	2.9	1.9	2.6	3.2
By Year																
2000	2.6	---	2.1	2.1	2.7	---	0.9	1.9	1.5	0.9	1.9	0.0	---	1.4	2.1	1.2
2001	2.4	---	2.2	2.2	2.8	---	1.7	2.0	1.7	1.2	1.9	0.3	---	1.4	2.1	1.4
2002	2.2	---	2.2	2.2	2.9	---	1.8	2.1	1.7	1.0	2.0	0.4	---	1.4	2.4	1.4
2003	2.1	---	2.3	2.3	2.9	---	1.8	2.2	1.8	0.8	2.0	0.4	---	1.4	2.5	1.3
2004	2.1	2.4	2.3	2.4	3.0	---	1.9	2.4	1.5	0.8	2.0	0.3	---	1.4	2.4	1.9
2005	2.1	2.5	2.3	2.4	3.0	---	2.2	2.5	1.0	0.8	2.0	0.6	---	1.5	2.2	2.0
2006	2.1	2.5	2.4	2.4	3.0	---	2.1	2.5	1.6	0.7	2.0	0.9	---	1.4	3.4	2.1
2007	2.1	2.6	2.4	2.4	3.0	4.0	2.1	2.9	2.4	0.8	1.8	0.8	---	1.3	3.1	2.7
2008	2.1	2.6	2.4	2.5	3.0	3.6	2.1	4.3	2.3	0.7	1.7	0.5	---	1.3	2.9	2.8
2009	2.1	2.6	0.9	2.5	3.0	4.0	2.1	6.3	2.1	0.7	1.6	0.7	---	1.5	2.8	3.0
2010	2.1	2.6	---	---	---	3.2	---	4.7	1.8	0.7	1.6	0.7	---	1.4	---	1.5
By Encounter Type																
Inpatient Stay	12.2	18.8	2.0	11.8	17.7	14.0	17.9	11.1	2.8	2.7	13.8	1.0	---	3.6	5.8	12.2
Non-Acute Institutional Stay	1.7	1.7	---	14.6	23.2	0.7	2.7	4.0	0.0	4.3	0.6	0.0	---	---	2.8	0.8
Emergency Department	4.1	10.5	---	11.6	29.7	2.8	---	0.1	0.0	1.7	0.1	0.1	---	0.0	0.2	0.5
Ambulatory Visit	1.6	2.2	2.3	2.1	2.6	3.9	2.0	2.3	1.8	0.6	1.7	0.4	---	1.0	2.3	1.9
Other Ambulatory Visit	2.7	2.1	2.1	2.2	---	4.8	0.6	3.0	0.0	1.1	1.6	108.6	---	2.2	2.1	2.5
Code Type																
ICD-9-CM	1.2%	1.7%	3.8%	0.6%	1.4%	3.3%	1.1%	1.3%	1.5%	0.0%	1.2%	6.7%	0.5%	5.3%	0.5%	2.3%
ICD-10-CM	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ICD-11-CM	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
CPT-4 (HCPCS Level I)	84.4%	73.5%	34.3%	88.1%	73.9%	59.6%	72.4%	83.1%	91.9%	84.9%	79.6%	76.4%	84.8%	94.7%	73.0%	69.7%
HCPCS (HCPCS Level II)	5.3%	6.9%	0.0%	5.4%	9.4%	9.4%	4.9%	5.2%	6.6%	1.0%	3.5%	2.0%	13.8%	0.0%	9.0%	7.3%
HCPCS Level III	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
CPT Category II	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.4%	0.1%	0.0%	0.0%	8.4%	0.0%	0.0%	0.0%	0.2%
CPT Category III	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Revenue	9.1%	17.9%	61.8%	2.8%	15.3%	27.2%	18.9%	9.9%	0.0%	3.5%	5.9%	0.0%	0.0%	0.0%	16.9%	19.1%
Local homegrown	0.0%	0.0%	0.0%	3.2%	0.0%	0.6%	2.5%	0.0%	0.0%	0.0%	9.8%	6.5%	0.9%	0.0%	0.0%	0.9%
Missing	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Invalid Type	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	10.6%	0.0%	0.0%	0.0%	0.0%	0.6%	0.4%

Figure 31. Number Procedures per Encounter, by Year, in MSDD Procedure Table in Extract 1 (Unique patients = 47,357,192)

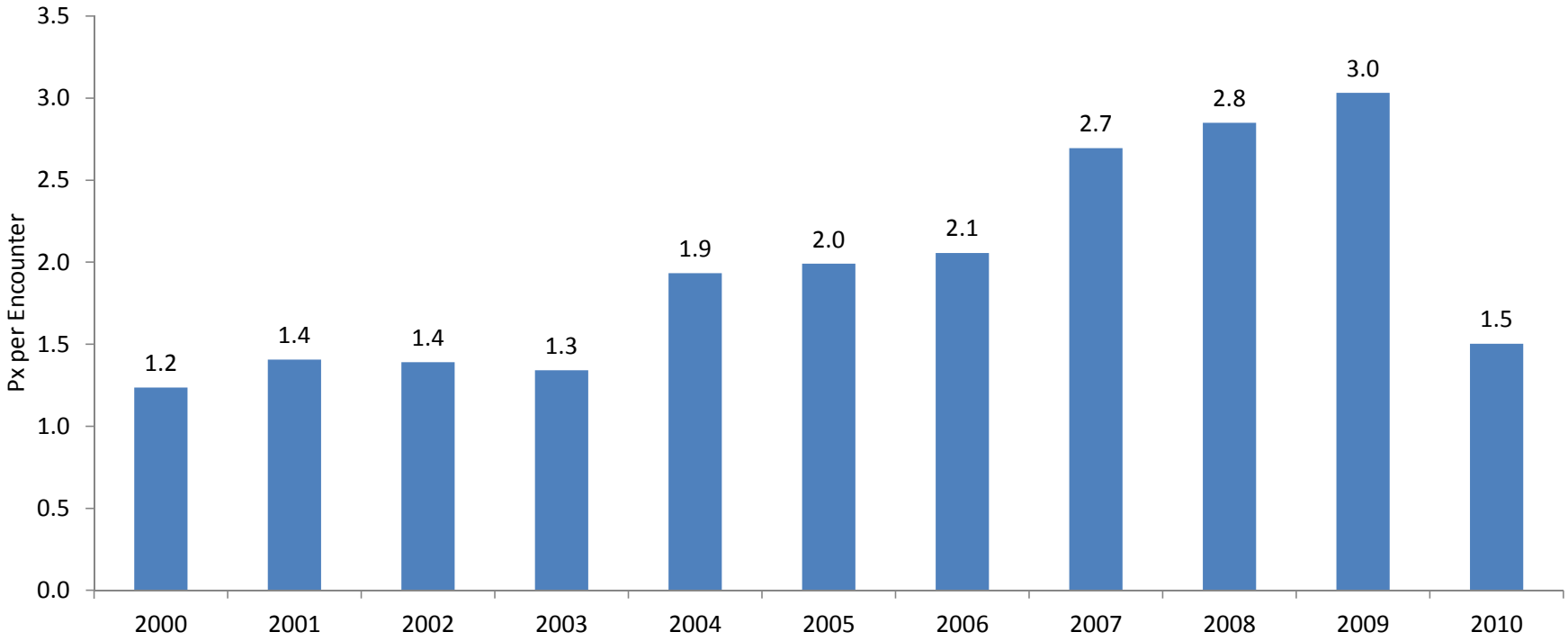


Figure 32. Distribution of Code Type in MSDD Procedure Table in Extract 1 (Unique patients = 47,357,192)

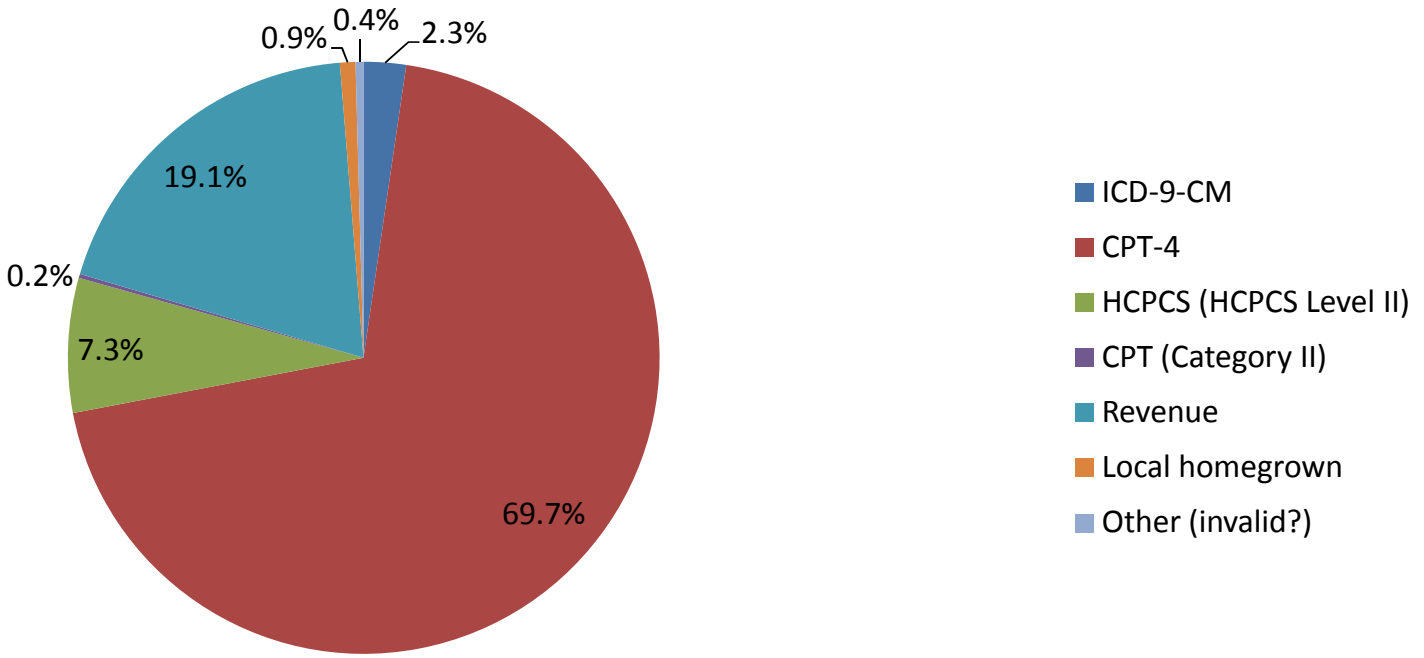


Table 7. Snapshot of the Mini-Sentinel Distributed Database Death Table in Extract 1 (Unique patients = 4,266,538)

	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10	DP11	DP12	DP13	DP14	DP15	Total
Source																
State Death Files	89.5%	0.0%	---	93.9%	87.2%	0.0%	72.3%	85.9%	0.0%	30.8%	59.8%	73.9%	---	42.1%	0.0%	37.1%
National Death Index	0.0%	0.0%	---	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	---	0.0%	0.0%	0.0%
Tumor data	4.3%	0.0%	---	0.5%	0.0%	0.0%	8.6%	0.0%	0.0%	0.0%	31.0%	0.0%	---	0.0%	0.0%	0.7%
Other, locally defined	6.2%	100.0%	---	5.6%	12.8%	100.0%	19.1%	14.0%	100.0%	69.2%	9.1%	26.1%	---	57.9%	0.0%	62.1%
Missing	0.0%	0.0%	---	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	---	0.0%	100.0%	0.1%
Else (Invalid?)	0.0%	0.0%	---	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	---	0.0%	0.0%	0.0%
Confidence																
Excellent	65.0%	100.0%	---	88.4%	87.2%	100.0%	76.0%	0.0%	100.0%	18.6%	100.0%	47.5%	---	100.0%	0.0%	55.3%
Fair	22.1%	0.0%	---	11.6%	12.8%	0.0%	22.9%	100.0%	0.0%	28.6%	0.0%	10.7%	---	0.0%	0.0%	14.8%
Poor	13.0%	0.0%	---	0.0%	0.0%	0.0%	1.1%	0.0%	0.0%	52.7%	0.0%	41.8%	---	0.0%	0.0%	29.7%
Missing	0.0%	0.0%	---	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	---	0.0%	100.0%	0.1%
Else (Invalid?)	0.0%	0.0%	---	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	---	0.0%	0.0%	0.0%

Figure 33. Distribution of Source in MSDD Cause of Death Table in Extract 1 (Unique patients = 4,266,538)

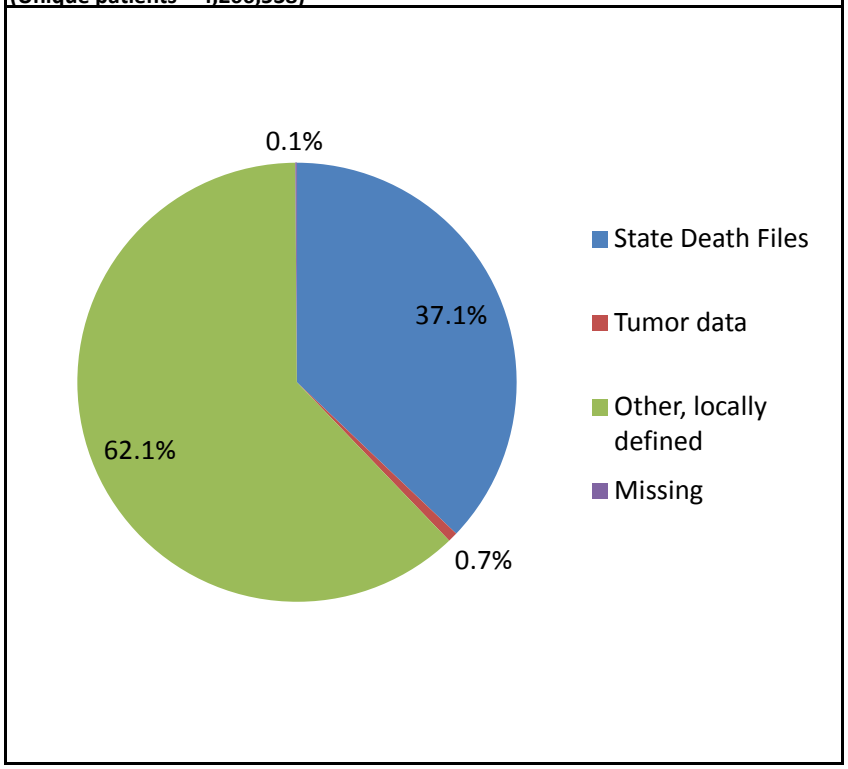


Figure 34. Distribution of Confidence in MSDD Cause of Death Table in Extract 1 (Unique patients = 4,266,538)

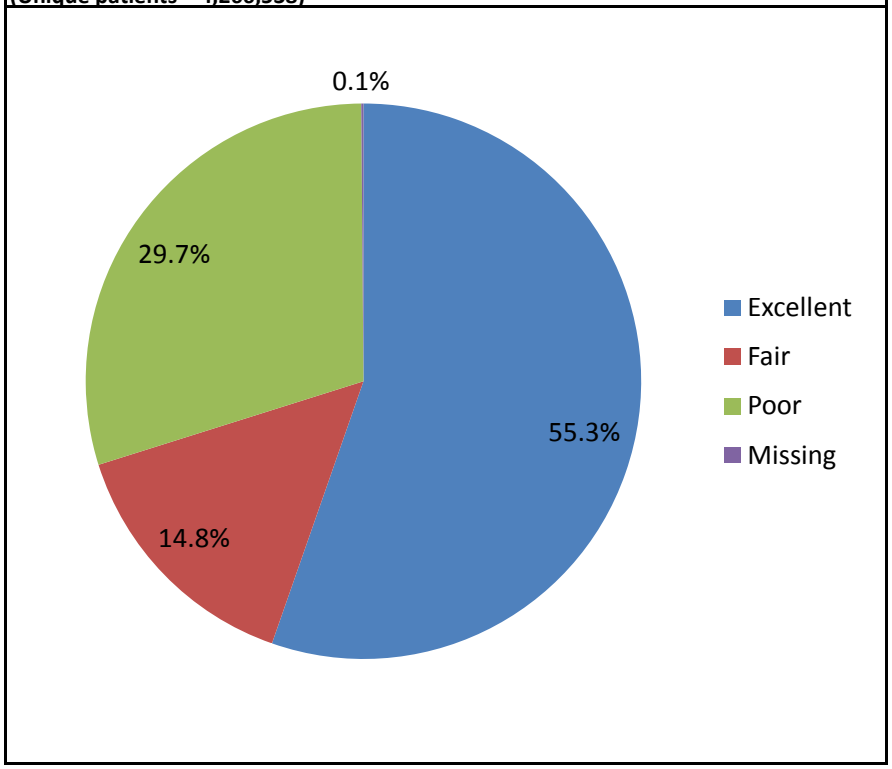


Table 8. Snapshot of the Mini-Sentinel Distributed Database Cause of Death Table in Extract 1 (Unique patients = 1,649,592)

	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10	DP11	DP12	DP13	DP14	DP15	Total
Cause Type																
Immediate/Primary	24.3%	---	---	0.0%	27.3%	---	34.2%	100.0%	---	22.0%	20.0%	24.8%	---	---	---	23.9%
Underlying	24.3%	---	---	28.5%	0.0%	---	33.6%	0.0%	---	30.3%	11.3%	28.5%	---	---	---	26.8%
Contributory	50.0%	---	---	0.0%	68.6%	---	28.3%	0.0%	---	28.3%	43.3%	46.7%	---	---	---	41.0%
Other	0.0%	---	---	71.5%	4.0%	---	3.4%	0.0%	---	19.4%	25.5%	0.0%	---	---	---	8.2%
Missing	0.0%	---	---	0.0%	0.0%	---	0.6%	0.0%	---	0.0%	0.0%	0.0%	---	---	---	0.0%
Else (Invalid?)	1.4%	---	---	0.0%	0.0%	---	0.0%	0.0%	---	0.0%	0.0%	0.0%	---	---	---	0.1%
Source																
State Death Files	98.6%	---	---	99.9%	96.0%	---	99.4%	85.9%	---	30.3%	100.0%	100.0%	---	---	---	79.9%
National Death Index	0.0%	---	---	0.0%	0.0%	---	0.0%	0.0%	---	0.0%	0.0%	0.0%	---	---	---	0.0%
Tumor data	0.0%	---	---	0.1%	0.0%	---	0.0%	0.0%	---	0.0%	0.0%	0.0%	---	---	---	0.0%
Other, locally defined	1.4%	---	---	0.0%	4.0%	---	0.6%	14.0%	---	69.7%	0.0%	0.0%	---	---	---	20.1%
Missing	0.0%	---	---	0.0%	0.0%	---	0.0%	0.0%	---	0.0%	0.0%	0.0%	---	---	---	0.0%
Else (Invalid?)	0.0%	---	---	0.0%	0.0%	---	0.0%	0.0%	---	0.0%	0.0%	0.0%	---	---	---	0.0%
Confidence																
Excellent	74.2%	---	---	93.7%	96.0%	---	94.1%	0.0%	---	29.7%	100.0%	39.9%	---	---	---	45.7%
Fair	24.4%	---	---	6.3%	4.0%	---	4.4%	100.0%	---	65.2%	0.0%	7.1%	---	---	---	24.4%
Poor	0.0%	---	---	0.0%	0.0%	---	1.4%	0.0%	---	5.2%	0.0%	53.0%	---	---	---	29.9%
Missing	0.0%	---	---	0.0%	0.0%	---	0.0%	0.0%	---	0.0%	0.0%	0.0%	---	---	---	0.0%
Else (Invalid?)	1.4%	---	---	0.0%	0.0%	---	0.0%	0.0%	---	0.0%	0.0%	0.0%	---	---	---	0.1%

Figure 35. Distribution of Cause Type in MSDD Cause of Death Table in Extract 1 (Unique patients = 1,649,592)

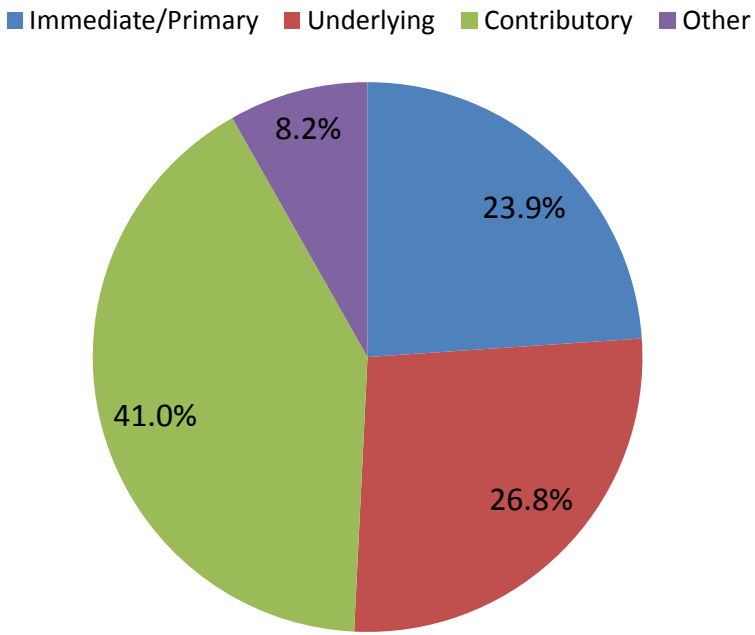


Figure 36. Distribution of Source in MSDD Cause of Death Table in Extract 1 (Unique patients = 1,649,592)

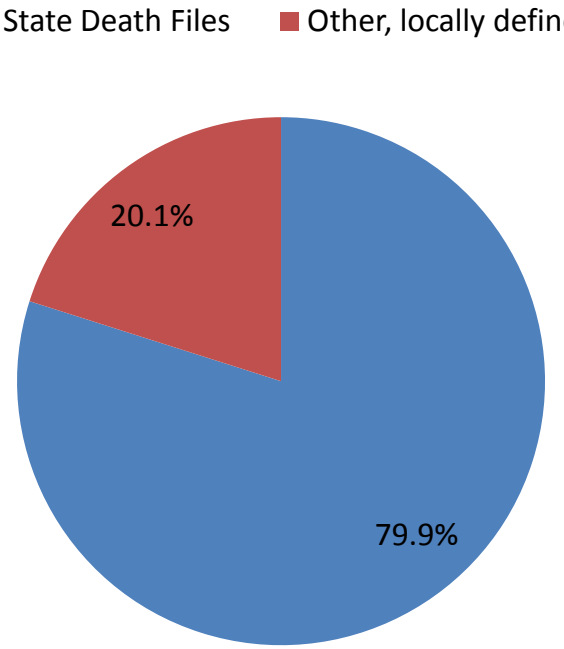


Figure 37. Distribution of Confidence in MSDD Cause of Death Table in Extract 1 (Unique patients = 1,649,592)

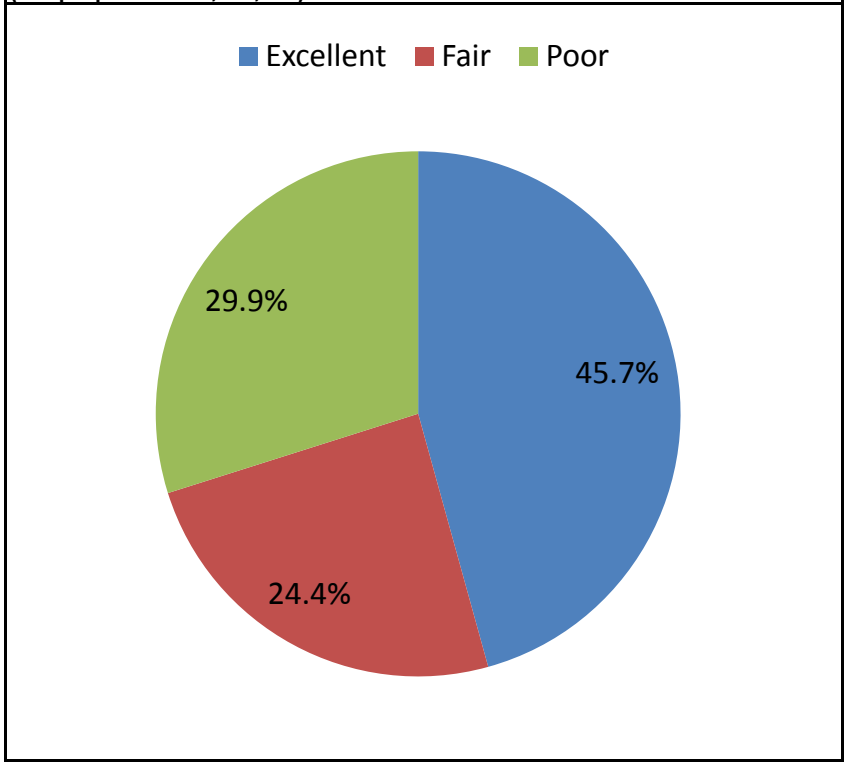


Table 9. Snapshot of the Mini-Sentinel Distributed Database Patient Match in Extract 1

	DP1	DP2	DP3	DP4	DP5	DP6	DP7	DP8	DP9	DP10	DP11	DP12	DP13	DP14	DP15
% of Patients Matched with Enrollment															
Demographic	48.3%	100.0%	41.8%	100.0%	54.6%	---	71.8%	91.2%	72.5%	54.8%	81.1%	98.4%	91.1%	77.5%	97.0%
Dispensing	92.3%	100.0%	99.8%	99.8%	---	97.9%	98.7%	88.6%	96.5%	95.4%	95.4%	96.2%	98.7%	99.4%	99.5%
Encounter	89.5%	100.0%	33.5%	100.0%	99.9%	---	96.7%	98.2%	80.0%	90.0%	91.9%	90.5%	98.4%	90.6%	99.3%
Diagnosis	89.8%	100.0%	33.6%	100.0%	99.9%	---	97.1%	98.2%	82.0%	90.8%	93.4%	91.7%	98.4%	90.6%	99.4%
Procedure	89.5%	100.0%	33.4%	100.0%	99.9%	---	97.0%	98.2%	79.8%	91.6%	92.3%	96.0%	98.4%	90.6%	99.4%
Death	53.3%	99.9%	---	100.0%	55.2%	---	84.2%	94.5%	91.3%	43.4%	98.8%	99.2%	---	64.6%	100.0%
Cause of Death	54.1%	---	---	100.0%	55.2%	---	100.0%	94.5%	---	47.7%	99.7%	100.0%	---	---	---