

MA APCD CY 2024: De-Identification Summary

Introduction

The Center for Health Information and Analysis (CHIA)'s enabling statute ([M.G.L. Ch 12C¹](#)) permits the disclosure of data from the commercial payers, third-party administrators, and public programs included in the Massachusetts All-Payer Claims Database (MA APCD). While CHIA is not a covered entity under the Health Insurance Portability and Accountability Act of 1996 (HIPAA), HIPAA provides requirements for the de-identification of health information.

To make the most robust set of health care data available to the widest audience, CHIA contracted with Onpoint Health Data to assist in the creation of an expert de-identified version of the MA APCD (as defined by the [HIPAA Privacy Rule²](#)). This document outlines the de-identification pathways defined in the HIPAA Privacy Rule and summarizes all the data masking, generalizations, and transformations required to produce a de-identified version of the MA APCD.

De-Identification Concepts and Methods

The HIPAA Privacy Rule is intended to safeguard identifiable health information by allowing only specific uses and disclosures. Recognizing the value of health data even when it does not identify individuals, the Rule permits the creation and disclosure of **de-identified data** that adhere to certain de-identification standards and implementation guidelines.

To achieve de-identification, the Privacy Rule accepts two different methods:

1. **Expert Determination:** A qualified expert applies statistical principles to determine that the risk of re-identifying an individual is very small.
2. **Safe Harbor:** The disclosing entity removes specific identifiers and certifies that it has no knowledge that the remaining data could be used to identify someone, either in isolation or when combined with other information.

The Safe Harbor method of deidentification requires the removal of 18 categories of identifiers. Removed identifiers include names, addresses, five-digit ZIP codes, birth dates, admission dates, discharge dates, and dates of service). Removal of these key data elements often makes the resulting data less useful in research, quality improvement and public health analysis.

De-Identification in the MA APCD

To make the most robust set of health care data readily available to the widest audience, CHIA and Onpoint employed the Expert Determination method.

¹ <https://malegislature.gov/Laws/GeneralLaws/PartI/TitleII/Chapter12C>

² <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>

There are several differences between CHIA's Expert Determination De-Identification and HIPAA Safe Harbor. Firstly, the CHIA model is less restrictive than Safe Harbor with respect to:

- certain ZIP Codes in Massachusetts, which are reported at the 5-digit level (or as an amalgamation of multiple 5-digit ZIP Codes)
- admission date, discharge date, and date of service
- birth month/year

Secondly, the CHIA model is more restrictive with respect to individual geographic and demographic details, as well as on certain codes corresponding to notable publicly recognizable or reportable events. The following types of information are removed:

- diagnosis codes indicating birth, death, trauma-related events, or public emergencies
- marital status
- family size
- student status
- race
- ethnicity
- ZIP Codes for states outside of Massachusetts
- U.S. state codes outside of New England and New York

Additionally, the CHIA model generalizes certain terms to mitigate re-identification risks:

- language spoken, which is limited to English, Spanish, and Other
- gender

Finally, the CHIA model applies risk mitigation strategies that are not required by Safe Harbor:

- Recipients of CHIA's de-identified data must sign a [Data Use Agreement](#).³

The appendix of attenuated data elements provides more information on specific value transformations for de-identified data elements.

³ <https://www.chiamass.gov/application-documents>

Appendix of Attenuated Data Elements in the De-Identified MA APCD

Admission Code

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
Medical Claims (MC)	Admission Source	MC021	admissionsourceDE

In this element, CHIA sets the value of 8, indicative of court or law enforcement, to 9, "information not available." Other values and other elements associated with the claim are not affected.

Condition Codes

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
in file	element name	element source	de-identified column name
Medical Claims (MC)	Condition Codes	MC179–190	conditioncode1–12DE

In this element, CHIA sets values in the table below to null. Other values and other elements associated with the claim are not affected.

Removed Values

Code	Description
01	Military service related - Medical condition incurred during military service.
05	Lien has been filed - Provider has filed legal claim for recovery of funds potentially due a patient as a result of legal action initiated by or on behalf of the patient.
18	Maiden name retained - A dependent spouse entitled to benefits who does not use her husband's last name.
19	Child retains mother's name - A patient who is a dependent child entitled to CHAMPVA benefits that does not have father's last name.
31	Patient is student (full time day) - Patient declares that he or she is enrolled as a full time day student.
32	Patient is student (cooperative/work study program)
33	Patient is student (full time night) - Patient declares that he or she is enrolled as a full time night student.
34	Patient is student (part time) - Patient declares that he or she is enrolled as a part time student.

Code	Description
45	Gender incongruence (eff. 7/2023)
81	C-sections/inductions < 39 weeks - medical necessity (eff. 10/1/13)
82	C-sections/inductions < 39 weeks - elective (eff. 10/2013)
83	C-sections/inductions 39 weeks or greater (eff. 10/2013)
A7	Induced abortion to avoid danger to woman's life
A8	Induced abortion - victim of rape/incest. Special program indicator code
AA	Abortion performed due to rape (eff. 10/1/2002)
AB	Abortion performed due to incest (eff. 10/1/2002)
AG	Abortion performed due to social economic reasons (eff. 10/1/2002)

Date of Birth

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
Member Eligibility (ME), Medical Claims (MC), Dental Claims (DC), Pharmacy Claims (PC)	Member Birth (Year)	ME014 MC013 DC013 PC013	memberdateofbirthye arcleaned
ME MC DC PC	Member Birth (Month)	ME014 MC013 DC013 PC013	memberbirthmonthDE
ME	Member Age at Enrollment	derived	memberageyearsasofs ubDE
MC, DC, PC	Member Age at Service	derived	memberageatservicei nyearsDE

In these elements, CHIA sets any age value greater than 90 to “90,” changes the corresponding birth month value to “99,” and changes the corresponding birth year value to “999.” Other values and other elements associated with the claim are not affected.

Discharge Codes

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
Medical Claims (MC)	Discharge Status	MC023	dischargestatusDE

In this element, CHIA sets values in the table below either to null or a value indicative of unknown status. Other values and other elements associated with the claim are not affected.

Transformed Values

CODE	DESCRIPTION
20	Expired (did not recover - Christian Science patient).
21	Discharged/transferred to Court/Law Enforcement
40	Expired at home (hospice claims only)
41	Expired in a medical facility such as hospital, SNF, ICF, or freestanding hospice. (Hospice claims only)
42	Expired - place unknown (Hospice claims only)
69	Discharged/transferred to a designated disaster alternative care site
87	Discharged/transferred to court/law enforcement with a planned acute care hospital inpatient readmission

Gender

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
Member Eligibility (ME), Medical Claims (MC), Dental Claims (DC), Pharmacy Claims (PC)	Member Gender	ME013 MC012 DC012 PC012	membergenderDE

In this element, CHIA sets all values other than “F” (female) or “M” (male) to “U” (unknown). Other values and other elements associated with the claim are not affected.

ICD Codes (Diagnosis, Occurrence Span, Present on Admission, Other)

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
Medical Claims (MC)	Admitting Diagnosis	MC039	admittingdiagnosisDE
MC	E-Code	MC040	ecodeDE
MC	Principal Diagnosis	MC041	principaldiagnosisDE
MC	Discharge Diagnosis	MC136	dischargedisagnosisDE
MC	Other Diagnosis 1–24	MC042–MC053, MC142–MC153	otherdiagnosis1-24DE
MC	Occurrence Span Codes	MC225 MC228 MC231 MC234 MC237	occurrencecodeDE1–5
MC	Present on Admission Code	MC154–MC178	presentonadmissionDE1–25
Dental Claims (DC)	Diagnosis Code	DC061	diagnosisDE
Pharmacy Claims (PC)	Diagnosis Codes	PC114	diagnosisDE

CHIA removed (i.e., set to null) or generalized International Classification of Diseases (ICD) codes according to the following table. Other values and other elements associated with the claim are not affected.

Removed Values

ICD-9	ICD-10	Description
764–779	P	Other conditions originating in the perinatal period (including birth trauma)
V3	Z38	Liveborn infants according to type of birth
798	R99	Unknown cause of death
799	—	Unknown cause of death
799.9	—	Unknown cause of death
E99	Y36	Operations of war
E000.1	Y37	Military operations
E970–E978	Y35	Legal intervention
E979	Y38	Terrorism
E96	X92–Y09	Homicide and injury purposely inflicted by other persons

ICD-9	ICD-10	Description
E95	X71–X83	Suicide and self-inflicted injury
E928.0	X52	Prolonged stay in weightlessness
E910	W65–W74	Accidental drowning and submersion
E913, 994.7	T71	Accidental mechanical suffocation / traumatic asphyxiation
E80-E84	V	Vehicle accident

Transformed Values

ICD-9	ICD-10	Description
[V85.41–V85.45] → V85.41	[Z68.41–Z68.45] → Z68.41	Body Mass Index 40 and over

Language

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
Member Eligibility (ME)	Member Language Preference	ME033	memberlanguagepreferenceDE

In this element, CHIA sets any value other than English or Spanish (including “unknown,” null, and “not used”) to Other. Other values and other elements associated with the claim are not affected.

MassHealth Aid Category Codes

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
MassHealth Enhanced Eligibility (MHEE)	Aid Category	n/a	n/a (not released in de-identified data)
MHEE	Aid Category Disability Indicator	De-Identified Aid Category	aidcatdisabilityindicatorDE
MHEE	Aid Category Premium Assistance Indicator	De-Identified Aid Category	aidcatpremiumassistanceindicatorDE
MHEE	Aid Category Coverage Type	De-Identified Aid Category	aidcatcoveragetypeDE
MHEE	Aid Category Population	De-Identified Aid Category	aidcatpopulationDE
MHEE	Aid Category Temporary Coverage Type	De-Identified Aid Category	aidcattemporarycoveragetypeDE

In these elements, CHIA transforms values according to the table below. Other values and other elements associated with the claim are not affected.

Transformed Values

Aid Category	De-Identified Aid Category	De-Identified Aid Category Disability Indicator	De-Identified Aid Category Premium Assistance Indicator	De-Identified Aid Category Coverage Type	De-Identified Aid Category Population	De-Identified Aid Category Temporary Coverage Type
00	40	N	N	Standard	MassHealth	N
66	37	N	N	Limited	MassHealth	N
67	37	N	N	Limited	MassHealth	N
68	37	N	N	Limited	MassHealth	N
69	38	Y	N	Limited	MassHealth	N
AL	61	N	N	Basic	MassHealth	N
AR	61	N	N	Basic	MassHealth	N
AS	38	Y	N	Limited	MassHealth	N
AT	61	N	N	Basic	MassHealth	N
AU	61	N	N	Basic	MassHealth	N
B1	40	N	N	Standard	MassHealth	N
B2	40	N	N	Standard	MassHealth	N
B4	40	N	N	Standard	MassHealth	N
C1	40	N	N	Standard	MassHealth	N
C2	D1	N	N	CarePlus	MassHealth	N
C3	37	N	N	Limited	MassHealth	N
C4	37	N	N	Limited	MassHealth	N
C5	37	N	N	Limited	MassHealth	N
TT	61	N	N	Basic	MassHealth	N
TU	37	N	N	Limited	MassHealth	N
TV	61	N	N	Basic	MassHealth	N
TW	38	Y	N	Limited	MassHealth	N

Occurrence Codes

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
Medical Claims (MC)	Occurrence Codes 1–5	MC215	occurrencecodeDE1–5
		MC217	
		MC219	
		MC221	
		MC223	

In these elements, CHIA removes (i.e., sets to null) the values in the table below. Other values and other elements associated with the claim are not affected.

Removed Values

Code	Description
1	Accident/Medical Coverage
2	No-Fault Insurance Involved - Including Auto Accident/Other
3	Accident/Tort Liability
4	Accident/Employment Related
5	Accident/No Medical or Liability Coverage
6	Crime Victim
19	Date of Retirement - Spouse
55	Date of Death
A1	Birthdate - Insured A
B1	Birthdate - Insured B
C1	Birthdate - Insured C
DR	Reserved for disaster-related Occurrence Code

State

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
Geography (GEO)	Member State (or Province)	ME016	memberstateorprovinceDE
GEO	Subscriber State (or Province)	ME109	subscriberstateorprovinceDE

In these elements, CHIA reports values within New England and New York (CT, MA, ME, NH, NY, RI, and VT) but sets all other values to “XX.” Other values and other elements associated with the claim are not affected.

ZIP Codes

Affected Data Elements

In File	Element Name	Element Source	De-Identified Column Name
Geography (GEO)	Member ZIP Code	ME017	memberzipcodeDE
GEO	Subscriber ZIP Code	ME110	subscriberzipcodeDE
GEO	Member County	derived	membercountyDE
GEO	Subscriber County	derived	subscribercountyDE
Member Eligibility (ME)	Employer ZIP Code	ME078	employerzipcodeDE

In these elements, CHIA:

sets non-Massachusetts ZIP Codes to “99999” and corresponding county values to “xx.”

sets some Massachusetts ZIP Codes to nearby alternatives, according to the following table, and updates corresponding county values accordingly.

Other ZIP Codes and other elements associated with the claim are not affected.

Transformed Values

Identifiable ZIP	De-Identified ZIP	Identifiable ZIP	De-Identified ZIP	Identifiable ZIP	De-Identified ZIP	Identifiable ZIP	De-Identified ZIP
01003	01002	01199	01151	01367	01339	01704	01701
01004	01002	01202	01201	01368	01331	01705	01701
01012	01096	01203	01201	01378	01364	01718	01720
01014	01013	01222	01257	01379	01344	01731	01730
01021	01020	01224	01237	01380	01376	01745	01721
01026	01098	01227	01226	01434	01432	01784	01748
01029	01230	01229	01226	01438	01436	01805	01803
01032	01096	01242	01240	01441	01440	01806	01801
01037	01082	01243	01011	01467	01464	01807	01801
01041	01040	01244	01236	01470	01450	01812	01810
01059	01002	01245	01230	01472	01450	01813	01801
01061	01060	01252	01230	01477	01475	01815	01801
01066	01063	01253	01230	01508	01507	01822	01824
01070	01098	01254	01266	01509	01507	01831	01830
01074	01068	01255	01230	01517	01516	01842	01840
01079	01069	01256	01220	01525	01524	01853	01852
01084	01098	01258	01257	01526	01590	01865	01821
01086	01085	01259	01257	01531	01068	01866	01821
01090	01089	01260	01240	01538	01569	01885	01921
01093	01072	01263	01262	01546	01545	01888	01801
01094	01092	01264	01238	01561	01523	01899	01810
01097	01085	01270	01235	01580	01569	01903	01901
01102	01104	01302	01301	01582	01581	01910	01905
01111	01109	01338	01330	01586	01527	01931	01930
01115	01103	01343	01339	01601	01608	01936	01982
01116	01106	01346	01339	01614	01608	01937	01929
01138	01108	01347	01344	01615	01605	01961	01960
01139	01109	01349	01344	01653	01605	01965	01915
01144	01103	01350	01247	01655	01605	01971	01970
01152	01151	01355	01366	01703	01701	02018	02043
02020	02050	02216	02215	02366	02364	02663	02660
02027	02026	02217	02116	02381	02360	02669	02668
02031	02048	02218	02118	02447	02445	02671	02639

Identifiable ZIP	De-Identified ZIP	Identifiable ZIP	De-Identified ZIP	Identifiable ZIP	De-Identified ZIP	Identifiable ZIP	De-Identified ZIP
02040	02066	02222	02114	02455	02451	02672	02670
02041	02050	02228	02128	02456	02472	02712	02766
02047	02045	02238	02139	02471	02472	02713	02740
02051	02050	02241	02110	02475	02476	02714	02748
02055	02066	02266	02111	02477	02472	02722	02720
02059	02050	02269	02169	02479	02478	02741	02740
02060	02066	02284	02111	02495	02458	02742	02740
02065	02050	02293	02111	02534	02532	02761	02760
02070	02093	02298	02111	02541	02540	02768	02767
02071	02081	02303	02302	02542	02649	02783	02379
02112	02109	02304	02301	02543	02740	02791	02790
02123	02115	02305	02301	02552	02535		
02133	02108	02325	02324	02553	02540		
02137	02136	02327	02359	02561	02559		
02153	02155	02331	02332	02564	02554		
02156	02155	02334	02356	02565	02563		
02163	02134	02337	02333	02573	02568		
02185	02184	02340	02339	02574	02540		
02187	02186	02344	02346	02584	02554		
02196	02111	02345	02360	02634	02632		
02201	02199	02349	02347	02637	02635		
02203	02199	02350	02341	02643	02653		
02204	02150	02355	02330	02647	02646		
02205	02111	02357	02356	02650	02633		
02206	02150	02358	02359	02659	02633		
02211	02109	02361	02360	02661	02645		
02212	02109	02362	02360	02662	02653		



CENTER FOR HEALTH INFORMATION AND ANALYSIS

501 Boylston Street, Boston, MA 02116
617.701.8100 | www.chiamass.gov
