



**All Payer Claims Database (APCD) -  
Special Meeting of the Data Privacy and Security  
Subcommittee**

*November 3, 2016*

# Presentation Overview

---

- Acceptance of Minutes from Previous Meeting
- Subcommittee Chairperson's Update
- Discussion on current data release policy and APCD enabling legislation
- Difference between De-identified versus Limited Data Set (LDS)
- What is the case for using LDS: Pros and Cons
- How are other APCDs using LDS: Best Practices
- Next steps
- Adjournment

# APCD Enabling Legislation and APCD Privacy Policy and Procedure

- CGS § 38a-1091 and § 38a-1090 provide for the disclosure of de-identified data by the APCD to state agencies, insurers, employers, health care providers, consumers of health care service, or researchers for the review of such data as it relates to health care utilization, costs or quality of health care services pursuant to 45 CFR 164.514
- APCD Privacy Policy and Procedure approved by Board of Directors on 2/18/2016 sets forth the policy and procedure for the release of data by the APCD
  - Data may only be released when release is consistent with APCD legislation and the Policy, and for legal and public purposes

# De-identified Data vs. Limited Data Set

De-identified data refers to healthcare information from which all 18 identifiers listed in 45 CFR 164.514(b)(2) have been removed.

Limited Data Sets exclude 16 of the listed identifiers but may include city, state, zip code, elements of date or other numbers, characteristics or codes not listed as direct identifiers. The two rows highlighted in yellow illustrate the difference.

#	Deidentified Data	Limited Data Set
1	Names	Names
2	State only, allowed 3-digit Zip if >20,000 eligibles	Postal Address information, other than town or city, State, and Zip code
3	No dates, just Year	Actual dates of events
4	Telephone #	Telephone #
5	Fax #	Fax #
6	Electronic Mail Address	Electronic Mail Address
7	Social Security Number	Social Security Number
8	Medical Record #	Medical Record #
9	Health Plan Beneficiary #	Health Plan Beneficiary #
10	Account #	Account #
11	Certificate/License #	Certificate/License #
12	Vehicle Identifiers, serial number, inc. license plate	Vehicle Identifiers, serial number, inc. license plate
13	Device identifiers and serial #	Device identifiers and serial #
14	Web Universal Resource Locators (URLs)	Web Universal Resource Locators (URLs)
15	Internet Protocol (IP) address #	Internet Protocol (IP) address #
16	Biometric identifiers, inc. finger or voice prints	Biometric identifiers, inc. finger or voice prints
17	Full face photographic images and any comparable images	Full face photographic images and any comparable images
18	Any other unique identifying number, characteristic, or code	Any other unique identifying number, characteristic, or code

Source: HIPAA's Safe Harbor - <http://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/>

# Various Data Set Samples

## Fully Identifiable Data

Person ID	Name	Gender	DOB	Address	Town	Zip Code	Service Date	Procedure	Diagnosis	Place of Service	Allowed Amount	Paid Amount	Copay
999-99-1234	Roger Smith	Male	1/1/1960	355 Main Street	Windsor	06095	1/7/2016	Office Vis.	Diabetes	Office	\$ 190	\$ 170	\$ 20
999-99-1234	Roger Smith	Male	1/1/1960	355 Main Street	Windsor	06095	1/29/2016	Office Vis.	Pneumonia	Office	\$ 250	\$ 220	\$ 30
999-99-1234	Roger Smith	Male	1/1/1960	355 Main Street	Windsor	06095	2/15/2016	CT Scan	Diabetes	Outpatient	\$ 750	\$ 500	\$ 250
999-99-1234	Roger Smith	Male	1/1/1960	355 Main Street	Windsor	06095	3/18/2016	EKG	Chest Pain	ER	\$ 950	\$ 750	\$ 200
999-99-1234	Roger Smith	Male	1/1/1960	355 Main Street	Windsor	06095	7/18/2016	Hospital	Stent	Hospital	\$ 9,500	\$ 9,000	\$ 500
999-99-1234	Roger Smith	Male	1/1/1960	355 Main Street	Windsor	06095	7/29/2016	Hospital	Infection	Hospital	\$ 7,500	\$ 7,000	\$ 500

## Limited Data Set

Person ID	Name	Gender	DOB	Address	Town	Zip Code	Service Date	Procedure	Diagnosis	Place of Service	Allowed Amount	Paid Amount	Copay
xqyrt2styiz3		Male	1/1/1960		Windsor	06095	1/7/2016	Office Vis.	Diabetes	Office	\$ 190	\$ 170	\$ 20
xqyrt2styiz3		Male	1/1/1960		Windsor	06095	1/29/2016	Office Vis.	Pneumonia	Office	\$ 250	\$ 220	\$ 30
xqyrt2styiz3		Male	1/1/1960		Windsor	06095	2/15/2016	CT Scan	Diabetes	Outpatient	\$ 750	\$ 500	\$ 250
xqyrt2styiz3		Male	1/1/1960		Windsor	06095	3/18/2016	EKG	Chest Pain	ER	\$ 950	\$ 750	\$ 200
xqyrt2styiz3		Male	1/1/1960		Windsor	06095	7/18/2016	Hospital	Stent	Hospital	\$ 9,500	\$ 9,000	\$ 500
xqyrt2styiz3		Male	1/1/1960		Windsor	06095	7/29/2016	Hospital	Infection	Hospital	\$ 7,500	\$ 7,000	\$ 500

## Deidentified Data Set

Person ID	Name	Gender	DOB	Address	Town	Zip Code	Service Date	Procedure	Diagnosis	Place of Service	Allowed Amount	Paid Amount	Copay
xqyrt2styiz3		Male				060	2016	Office Vis.	Diabetes	Office	\$ 190	\$ 170	\$ 20
xqyrt2styiz3		Male				060	2016	Office Vis.	Pneumonia	Office	\$ 250	\$ 220	\$ 30
xqyrt2styiz3		Male				060	2016	CT Scan	Diabetes	Outpatient	\$ 750	\$ 500	\$ 250
xqyrt2styiz3		Male				060	2016	EKG	Chest Pain	ER	\$ 950	\$ 750	\$ 200
xqyrt2styiz3		Male				060	2016	Hospital	Stent	Hospital	\$ 9,500	\$ 9,000	\$ 500
xqyrt2styiz3		Male				060	2016	Hospital	Infection	Hospital	\$ 7,500	\$ 7,000	\$ 500

# LDS vs. De-identified Data - Pros and Cons

---

## PROS

- A. Date (in LDS) enables understanding of utilization pattern quarterly or seasonally
- B. Patient safety - complications or unanticipated effects due to treatments
- C. Hospital readmission - has a person who was released from hospital readmitted within 30 days?
- D. Have patients developed severe complications after the administration of certain drugs?
- E. Have patients recovered from certain treatments? 'Pre-post' effect
- F. Develop episodic view of costs and utilization applying clinical groupers - DRG, APDRG, ETG, CRG, etc.
- G. Develop HEDIS metrics for analyzing population health
- H. Dates enable Medication adherence studies
- I. Evaluation studies for population (or patient) intervention studies can be performed using LDS

# LDS vs. De-identified Data - Pros and Cons

---

## CONS

- A. Dates may enable patient re-identification, particularly those with rare diseases or surgeries
- B. If a zip code area has smaller population, risk of identification is higher for those with rare disease and/or surgeries
- C. Date of Birth (DOB) may add increased possibility for patient re-identification
- D. Requestors with limited data infrastructure and limited experience handling sensitive data may pose risk of data breach and re-identification
- E. Inadequate Data Use Agreement (DUA) may pose risk of data breach and re-identification

# Data Uses in Colorado

---

1. Testing Bundled Care Analytics to Improve Care and Lower Costs
2. Analyzing Inflammatory Bowel Disease Treatments and Medications to Identify Best Practices
3. Developing Bundled Payments for Palliative Care in Colorado
4. Assessing the Impact of Cost-Sharing Policy Changes on the Use of Preventive Services
5. Analyzing ED visits for non-traumatic dental services in Colorado
6. Identifying Opportunities to Reduce Use of Potentially Harmful Medications During and Post Surgery
7. Analyzing Disease-specific Treatment Cost and Utilization
8. Developing Bundled Payments for Orthopedic Care
9. Developing Innovative 21st Century Ambulatory Health Care Model
10. Evaluating Cost of Care to Move Towards Value-Based Economic Model
11. Analyzing Health Plan Rates in Rural Areas

# Data Uses in Maine

---

1. Primary care Practice Reports - Testing rates for preventative and chronic care, and NQF endorsed measures - Total Cost of Care, and Resource Use Indices
2. Conduct volume and service use analyses for hospitals in Maine; statistics on utilization patterns and market share; conduct community needs assessments including patient volume, patient origin, service lines, and utilization for capacity planning
3. Study medication adherence rates for chronic diseases, for example, diabetes, hypertension, etc.
4. Study well child visits as a state initiative linked to their Raising Readers program, which is an early childhood literacy program in Maine
5. SIM grant deliverable for Year 3; to support a formal analysis of MaineCare claims cost for diabetes/pre-diabetes/no-diabetes for coverage to receive the National Diabetes Prevention program

# Data Uses in Massachusetts

---

1. The Effects of Network Structure on Provider Learning, Health Outcomes, and Costs of Care
2. Narrow Networks in the Health Insurance Market
3. Measuring the Clinical and Economic Outcomes Associated with Delivery Systems
4. Massachusetts Healthcare Costs and Utilization Estimates
5. Improving Healthcare Outcomes and Lowering Costs by Improving Prescribing Effectiveness through Data Analysis
6. Hospital Risk Adjusted Mortality and Complications in Massachusetts
7. A Study on High Deductible Health Plans in Massachusetts
8. Patterns in ED Use With and Without Ambulance Use, and Subsequent Healthcare Utilization
9. Oral health services and oral-systemic health connections
10. Exploring Factors Affecting Medication Adherence Trajectories
11. Assessing Susceptibility to Air Pollution Using Hybrid Data Mining and Epidemiological Techniques

# Measures needed for securing LDS data release

Any change to data disclosure types from De-identified Data to Limited Data Sets will require risk mitigation

## Discussion:

1. Releasing Actual Dates -
  - a. Date of Birth -restrict release of DOB based on research requirements, ability to maintain data security, past experience working with limited data sets, and other requirements TBD.
  - b. Date of Service - need for healthcare research; may be allowed based on research requirements; will be restricted for rare surgeries and/or providers/hospitals; will need past experience and capability to secure data and other requirements TBD.
2. Releasing Zip Code - various public policy research and state agency research needs data at zip code level. If a particular zip code has very few people with rare condition(s) or surgery, restrict data release. Other restrictions may need to be identified.