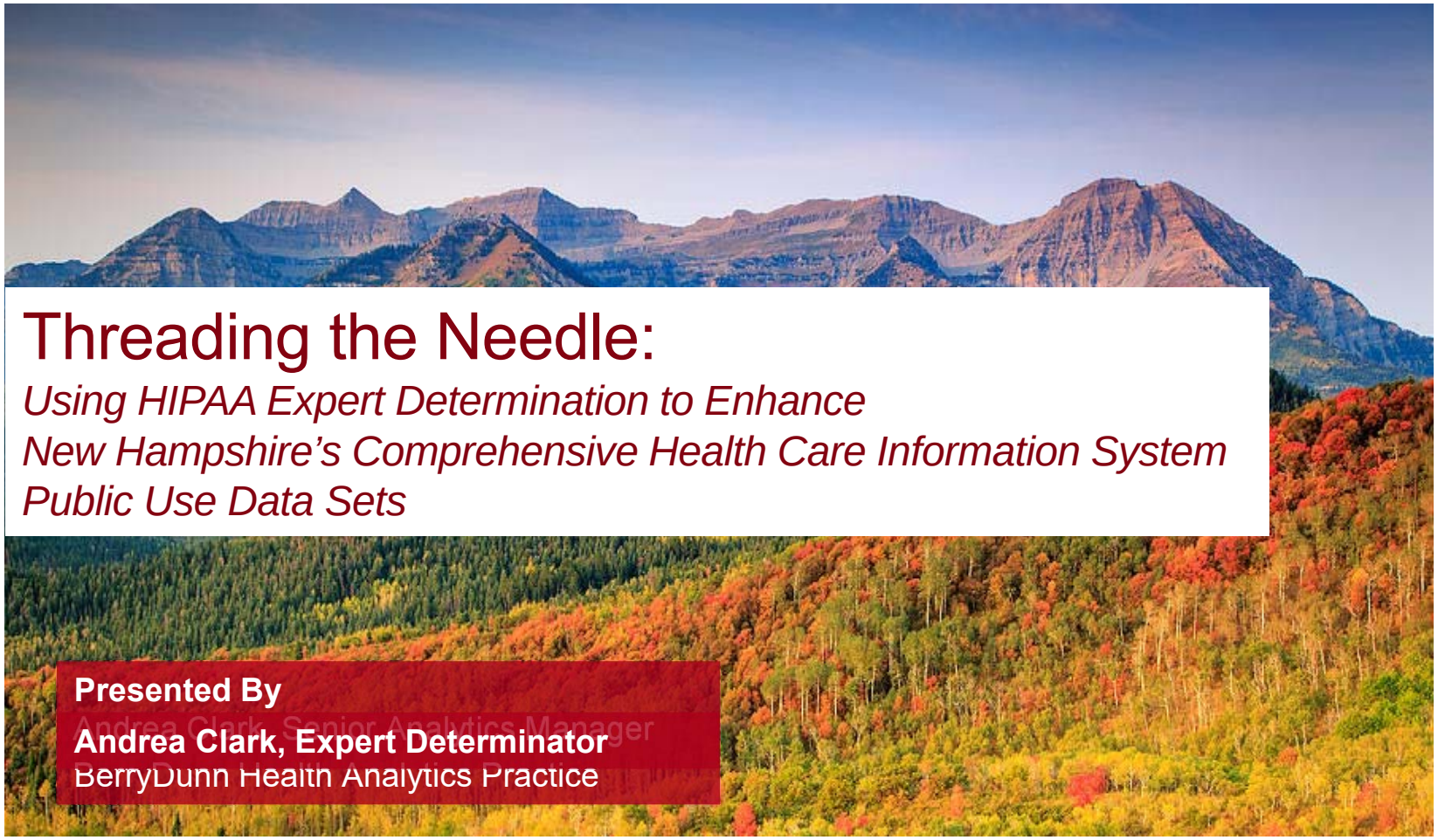
BerryDunn



Threading the Needle:

Using HIPAA Expert Determination to Enhance New Hampshire's Comprehensive Health Care Information System Public Use Data Sets

Presented By

Andrea Clark, Expert Determinator

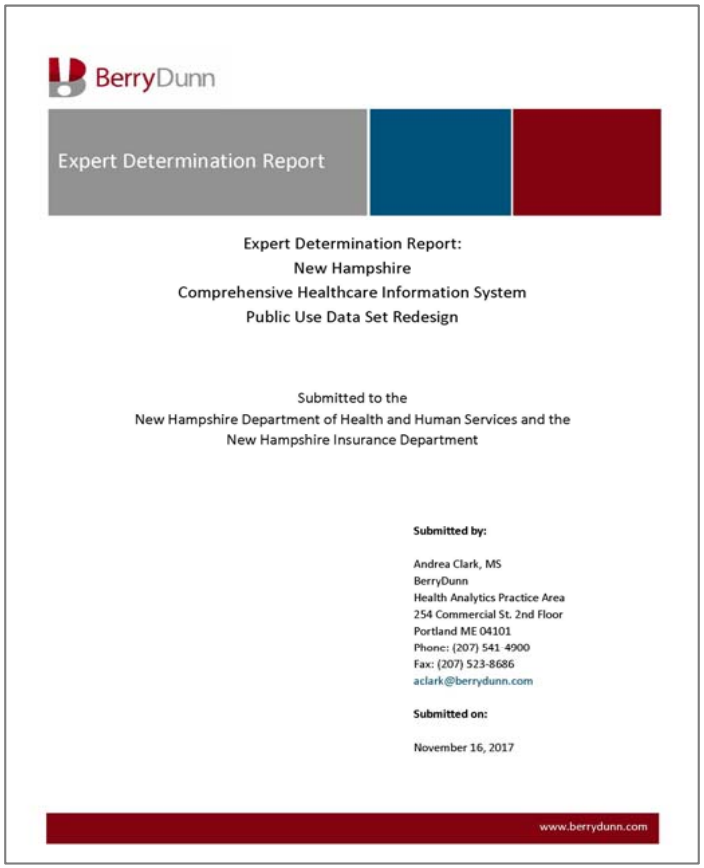
BerryDunn Health Analytics Practice

National Association of Health Data Organizations
Health Care Data Summit
Park City, Utah
October 11, 2018

Public Use Files Enhancement Project

Background

- New Hampshire has been a pioneer in APCDs, particularly with respect to public access to data
- New Hampshire is the only state to provide access to claim-level APCD data for no charge
  - Limited Data Sets
  - Public Use Sets
- This transparency requires trade-offs
- BerryDunn chosen through competitive bid process to propose enhancements to public use data


BerryDunn

Expert Determination Report

Expert Determination Report:
New Hampshire
Comprehensive Healthcare Information System
Public Use Data Set Redesign

Submitted to the
New Hampshire Department of Health and Human Services and the
New Hampshire Insurance Department

Submitted by:

Andrea Clark, MS
BerryDunn
Health Analytics Practice Area
254 Commercial St. 2nd Floor
Portland ME 04101
Phone: (207) 541 4900
Fax: (207) 523-8686
aclark@berrydunn.com

Submitted on:

November 16, 2017

www.berrydunn.com

HIPAA De-Identification Methods (2012 Privacy Rule)

Safe Harbor Method

- ✓ 18 forbidden identifiers
- ✓ Cookbook
- ✓ No individual member concept
- ✓ Used almost universally for all-payer claim database (APCD) public release

Expert Determination Method

- ✓ Who is an “expert”?
  - o No specific certification
  - o Knowledge of and experience with data and privacy issues
  - o Knowledge of statistical methods
- ✓ Re-identification risk is “very small” as determined by expert
  - o Standard not quantified by federal rules
- ✓ *Flexible*

Conclusions From De-id. Literature

Acceptable risk for personal information in a public use environment:

Pr(ID)≤ 0.05, where Pr(ID) is the probability of re-identification for the most vulnerable record

Equivalence classes of 1/0.05, or 20 individuals or larger, are required (i.e., k=20, where k is the number of individuals in a class)

Conclusions From De-id. Literature (cont'd)

Generalizations	
Age	10-year interval, 80+
DaysinHospitalY2	Days to 2 weeks; >2 weeks in year 2
DaysinHospitalY3	Days to 2 weeks; >2 weeks in year 3
Specialty	12 generalized specialty groupings
POS	8 generalized POS
CPTCode	17 procedure code groupings
LOS	Days up to 6 days; 1-2 weeks, 2-4 weeks, 4-8 weeks, 8-12 weeks, 12-26 weeks, 26+ weeks
DSFC	4 weeks
Diagnosis	45 primary condition groups

Source: El Emam, K, *et al.* "De-Identification Methods for Open Health Data: The Case of the Heritage Health Prize Claims Dataset." *Journal of Medical Internet Research*, 2012 Jan-Feb; 14(1): e33. DOI: 10.2196/jmir.2001. Accessed 4 August 2017.

Conclusions From De-id. Literature (cont'd)

Aggregated data are not technically personal information

Care should still be taken to minimize uniqueness/ re-identification risk:

CMS cell suppression standard	N ≥ 11 for units, days' supply, and users. Any dollar amount is reportable

Methodological Touchstones

- ✓ Equivalence classes of 20 required for person-level data
  - Sampling reduces required equivalence class size in proportion to the size of the sample
- ✓ CMS cell reporting standards applied to aggregated files
  - N≥11 for units, days' supply, and users
  - Any dollar amount is reportable
- ✓ Achieve adequate equivalence classes and cell sizes through generalization
  - Generalization strategy proceeds such as to preserve information unique to a given file as long as possible

Methodological Touchstones (cont'd)

- ✓ Add information not available on current public use files
- ✓ Minimize data suppression
- ✓ Create annual calendar-year files, unlinkable across years
- ✓ Methodology and results must be continuously re-evaluated as source data evolve
- ✓ Design assumes claim-level public use files will continue to be available in essentially their current form
- ✓ Query tool interface access to the public use data could be added

Given all that…

What types of analyses/users can we best support?

Finance

Price Transparency
Cost Driver Insight

Quality

Provider Profiling

Regulatory/Market

Payer Insight

Our Recommendation

BerryDunn recommended a suite of files to enhance the current public use data

Nine aggregated files:

1	Medical Expense by Payer, Provider, and Service
2	Payer Medical Expense by Demographics, Product Type, Service Type
3	Payer Pharmacy Expense by Demographics, Product Type, Drug Type
4	Medical Expense & Users by Demographics and Procedure Code
5	Medical Expense & Users by Demographics and 3-digit Primary Dx
6	Pharmacy Expense & Users by Demographics and Drug
7-9	Medical, Pharmacy, and Dental Membership

Two de-identified person-level files (managed data):

1	Medical Population Cost by Member
2	Pharmacy Population Cost by Member

Preview of Selected Draft File

Medical Population Cost by Member

service_year	person_key	coverage_class	CCHG_Cat	CCHG_Cat_Desc					Utilization_Category
2016	20000000	MED	119	Other mental health/substance abuse					Clinic/Office
3	2016	20000000	MED	119	Other mental health/substance abuse	OUM	$ 32	$ 24	1
4	2016	20000001	MED	117	Mental retardation/disability congentia anomaly	OUM	$ 30,835	$ 16,425	111
5	2016	20000002	MED	103	Active cancer	Ambulatory Surgery	$ 1,837	$ 1,603	2
6	2016	20000002	MED	103	Active cancer	Clinic/Office	$ 932	$ 861	10
7	2016	20000002	MED	103	Active cancer	Hospital Outpatient	$ 76	$ 76	1
8	2016	20000002	MED	103	Active cancer	OUM	$ 42	$ 23	3
9	2016	20000003	MED	120	Gastrointestinal disorders	Clinic/Office	$ 1,845	$ 774	7
10	2016	20000003	MED	120	Gastrointestinal disorders	Hospital Outpatient	$ 3,672	$ 2,508	29
11	2016	20000004	MED	103	Active cancer	Clinic/Office	$ 210	$ 150	1
12	2016	20000004	MED	103	Active cancer	Home Health/Hospice	$ 4,784	$ 3,525	60
13	2016	20000004	MED	103	Active cancer	Hospital Outpatient	$ 9,677	$ 3,297	34
14	2016	20000004	MED	103	Active cancer	OUM	$ 223	$ 214	1
15	2016	20000005	MED	119	Other mental health/substance abuse	Clinic/Office	$ 6,974	$ 5,872	52
16	2016	20000005	MED	119	Other mental health/substance abuse	Hospital Inpatient	$ 8,053	$ 6,910	6
17	2016	20000005	MED	119	Other mental health/substance abuse	Hospital Inpatient Prof Svcs	$ 341	$ 300	3
18	2016	20000005	MED	119	Other mental health/substance abuse	Hospital Outpatient	$ 77,647	$ 64,805	1585
19	2016	20000005	MED	119	Other mental health/substance abuse	OUM	$ 324	$ 0	2
20	2016	20000006	MED	103	Active cancer	Clinic/Office	$ 436	$ 259	3

Preview of Selected Draft File (cont'd)

Possible Uses . . .

Decile Spending by Utilization Category (000s)						
Deciles of Per-Member Spending	**Amb. Surgery**	**Clinic/Office**	**Hosp. OP**	**Hosp. IP**	**OUM**	**Total, All Util. Categories**
Lowest: Decile #1, Min-10th Percentile						
Decile #2, 11-20th Percentile						
Decile #3, 21-30th Percentile						
Decile #4, 31-40th Percentile						
Decile #5, 41-50th Percentile						
Decile #6, 51-60th Percentile						
Decile #2, 61-70th Percentile						
Decile #2, 71-80th Percentile						
Decile #2, 81-90th Percentile						
Highest: Decile #10, 91-Max						
Total Spending, All Patients						

Per-Patient Chronic Condition Hierarchical Grouping Spending by Utilization Category						
Chronic Condition Hierarchical Groupings	**Amb. Surgery**	**Clinic/ Office**	**Hosp. OP**	**Hosp. IP**	**OUM**	**Total, All Util. Categories**
Major psychosis						
Severe dementia						
Active cancer						
Both CAD & diabetes						
CAD without diabetes						
Diabetes without CAD						
Healthy Male (16-40)						
Healthy Male (41-64)						
Healthy Female (16-40)						
Healthy Female (41-64)						
Etc.						

Questions and Discussion

Thank you!

Andrea Clark, MS

Email me at: aclark@berrydunn.com for a copy of this presentation.

BerryDunn