

PCORnet CDM Forum

Friday, August 19, 2016 from 2–3 PM Eastern

Hosted by Keith Marsolo, PhD

Facilitated by Shelley Rusincovitch and Michelle Smerek



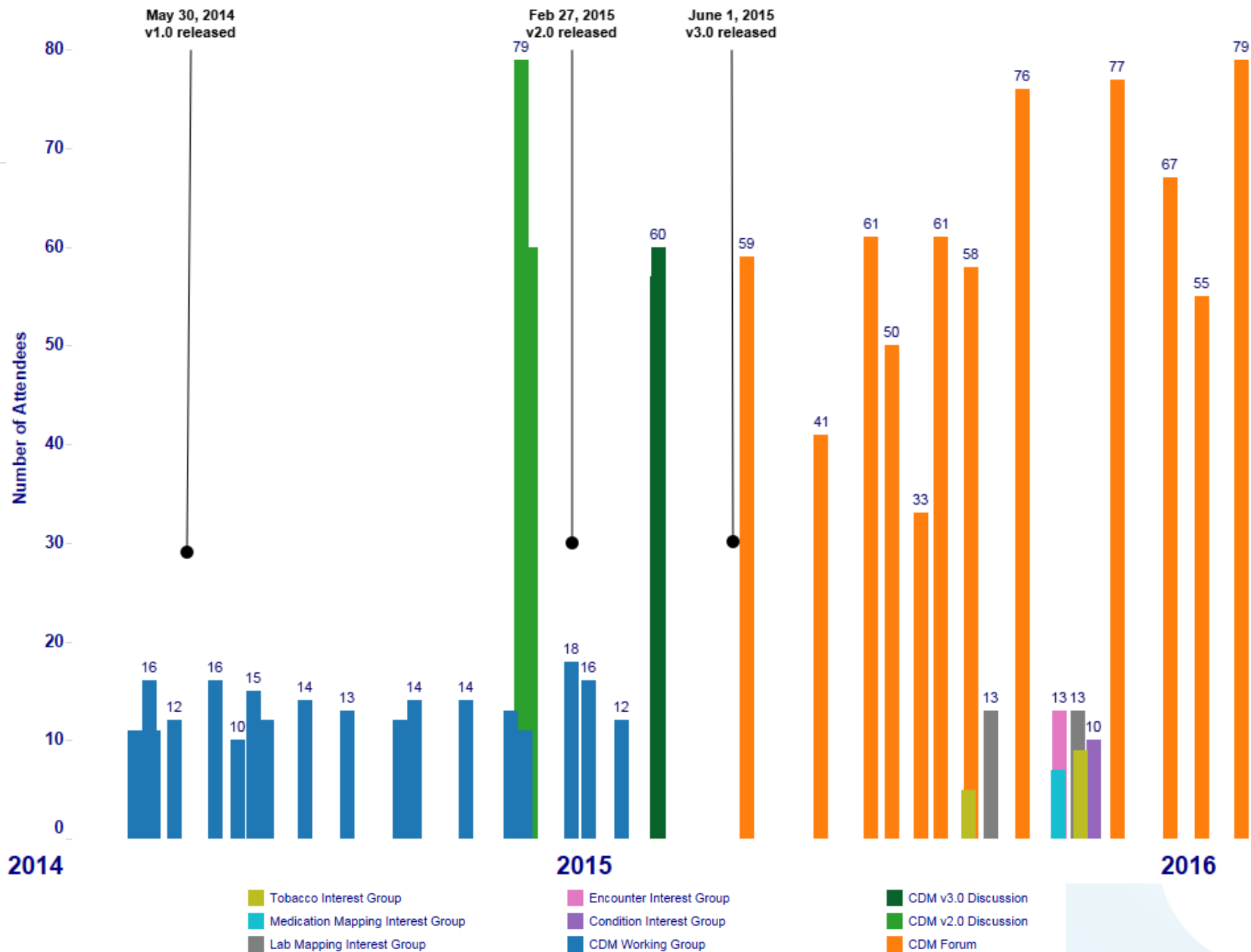
pcornet[®]

The National Patient-Centered
Clinical Research Network

Agenda

- 🌐 Announcements
- 🌐 Challenges in PX_TYPE assignments
- 🌐 Fixing PX Type issues in OneFlorida
- 🌐 PCORnet CDM v3.1 development
- 🌐 Data Committee CDM survey
- 🌐 Wrap up

Announcements



Recap: CDM Forum from June 8

Presentations by Keith Marsolo and Laura Qualls on implementation issues identified through data characterization, and exploratory analysis of variation in procedure data

- 🌐 Slides: <https://github.com/CDMFORUM/CDM-GUIDANCE/wiki/CDM-Forum-Materials>
- 🌐 Recording and meeting summary: <https://pcornet.imeetcentral.com/p/ZgAAAAAAade5K>

Recap: Research Using a DRN Forum from July 11

“Query 1: Description, Distribution and Preliminary Results”



Slides and meeting recording:

https://pcornet.imeetcentral.com/p/ZgAAAAAAAd2_U

Recap: DRN OC Forum from August 1

“DRN OC Updates”



Slides and meeting recording:

<https://pcornet.imeetcentral.com/p/ZgAAAAAAeJIZ>

Recap: Research Using a DRN Forum from August 11

“Obesity Demonstration Projects: Wave 1 Data Characterization Query Description”



Slides and meeting recording:

<https://pcornet.imeetcentral.com/p/ZgAAAAAAeTUg>

Review of Issue Tracker (live)

CDM errata issue tracker:

<https://github.com/CDMFORUM/CDM-ERRATA/issues>

CDM guidance issue tracker:

<https://github.com/CDMFORUM/CDM-GUIDANCE/issues>

Challenges in PX_TYPE assignments

Recap from 6/8/2016 CDM Forum

James Topping
PCORnet DRN OC

PROCEDURES Overview

PX_TYPES

- ICD
 - 09=ICD-9-CM
 - 10=ICD-10-PCS*
 - 11=ICD-11-PCS
- CPT/HCPCS
 - C2=CPT Category II
 - C3=CPT Category III
 - C4=CPT Category I**
 - HC=HCPCS (i.e., HCPCS Level II)
 - H3=HCPCS Level III
- Other terminologies
 - Revenue (RE)
 - NDC (ND)
 - LC (LOINC)
- Flavors of null (NI/UN/OT/Null or Missing)

} HCPCS Level I

Records

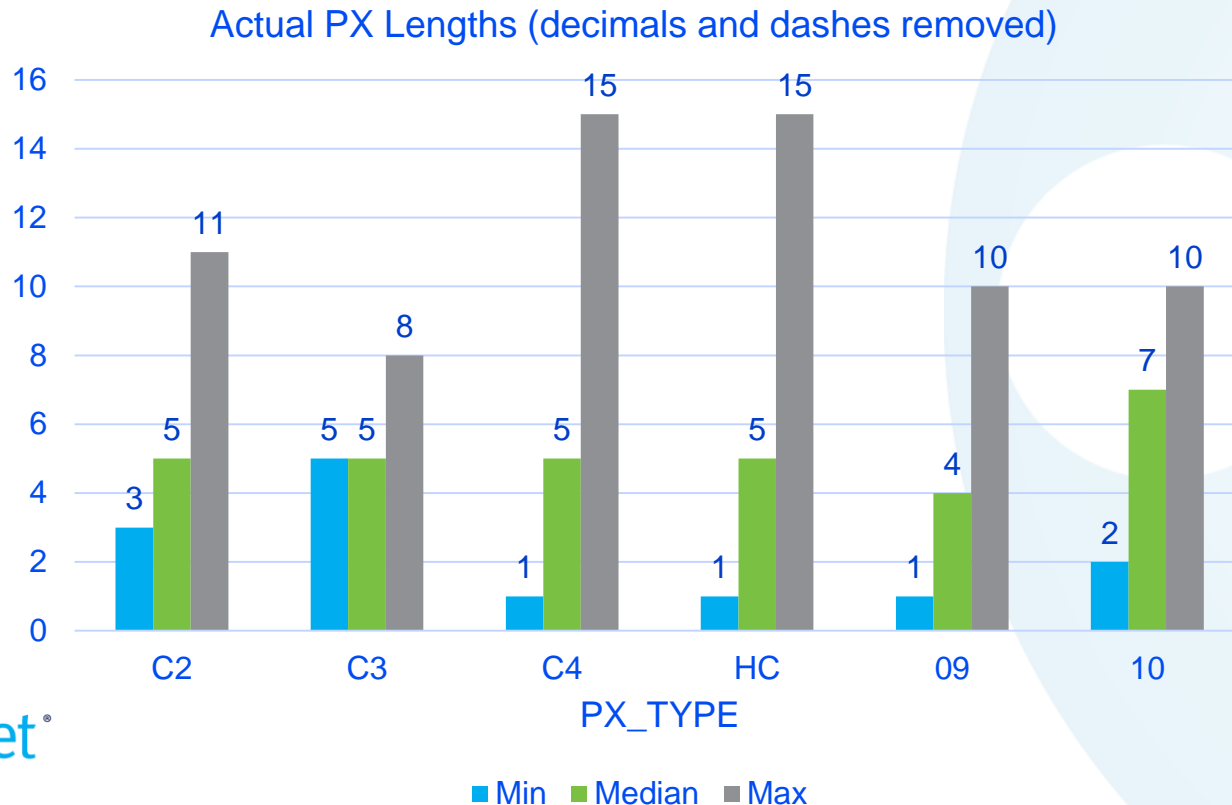


Do codes meet the basic format for the given procedure type?



Basic format

- ICD-9-CM (09) : 3-4 numbers (e.g. 89.39)
- ICD-10-PCS (10): 7 alphanumeric characters (e.g. 09B00ZX).
- CPT/HCPCS (C2/C3/C4/HC/H3): 5 alphanumeric characters (e.g. H2010, 99213); may be longer if modifiers are included

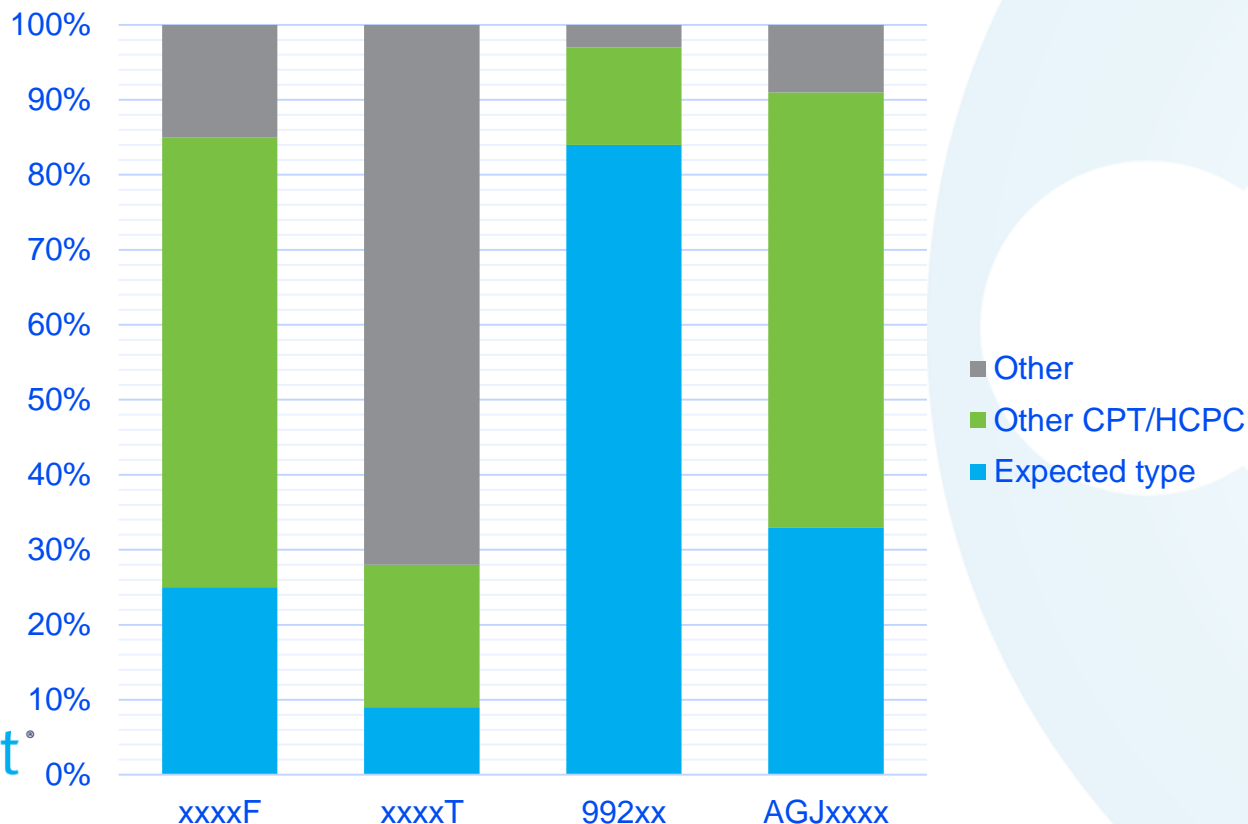


Are selected CPT and HCPCS codes mapped to the correct procedure type?



Sample codes

- xxxxF codes (codes used for tracking performance measurement) should be C2
- xxxxT codes (temporary codes for emerging technologies, services, and procedures) should be C3
- 992xx codes (basic E/M codes) should be C4.
- A/G/Jxxxx (HCPCS Level II codes for various services not covered by CPT) should be HC.



Fixing PX Type issues in OneFlorida

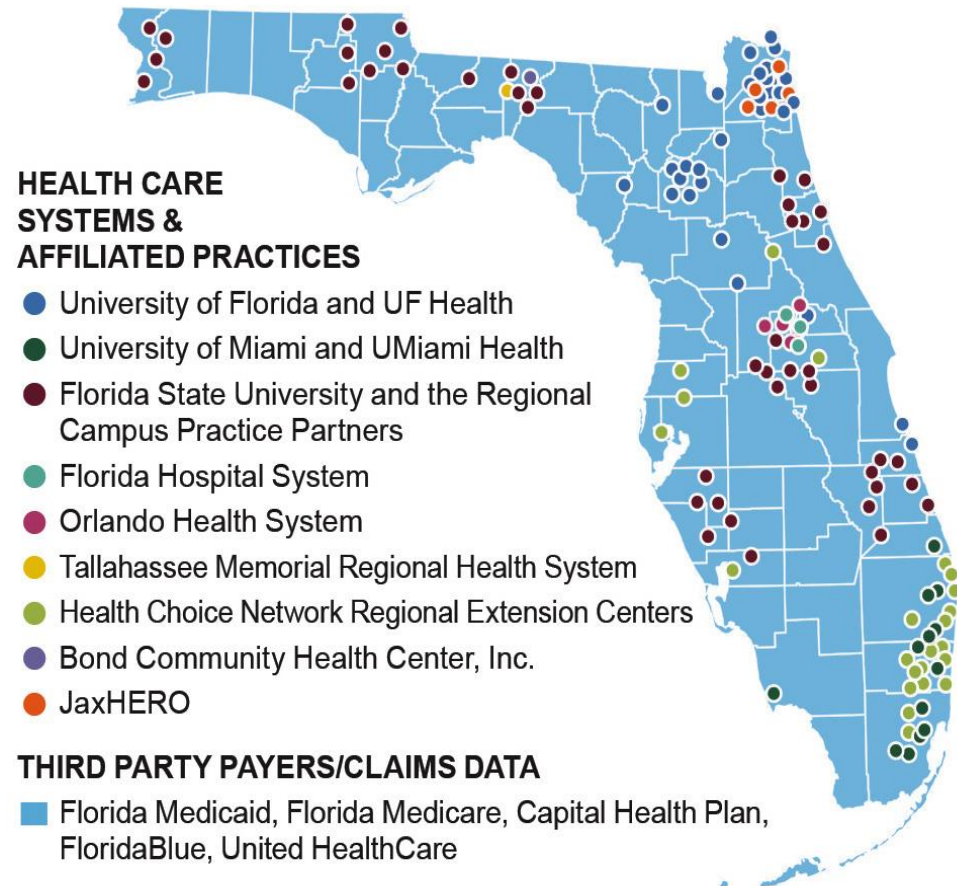
Based on DRN OC feedback and internal QA

*Jiang Bian, PhD
University of Florida
OneFlorida CDRN*

Overview of OneFlorida

- ❁ A centralized DataMart hosted by the University of Florida
- ❁ 9 data partners overall
- ❁ The current OneFL DataMart includes data from 6 out of our 9 data partners
- ❁ We do have some claims data (i.e., Florida Medicaid) already loaded

Figure 1. OneFlorida Clinical Partner Coverage Area



Data Status as of 08/11

- Row counts on EXPECTED tables: DEMOGRAPHIC, DIAGNOSIS, ENCOUNTER, ENROLLMENT, PROCEDURES, VITAL.

	DEMOGRAPHIC	ENROLLMENT	ENCOUNTER	DIAGNOSIS	PROCEDURES	VITAL
1FL	8,233,438	9,122,183	139,709,225	248,334,772	317,557,519	320,280,745

- Row counts on OPTIONAL tables: DISPENSING, LAB_RESULT_CM, CONDITION, PRESCRIBING, DEATH

	DISPENSING	LAB_RESULT_CM	CONDITION	PRO_CM	PRESCRIBING	DEATH	DEATH_CAUSE
1FL	66,746,851	9,465,381	5,164,160	Not Available	78,199,230	37,148	Not Available

The PX_TYPE issue (5/12/2016 DC results)

Table IVC. Missing or Unknown Values

This table includes fields in the DEMOGRAPHIC, ENROLLMENT, ENCOUNTER, DIAGNOSIS, PROCEDURES and VITAL tables which are included in the query results and are not required to be populated (see Table IIC for required fields). VITAL measures are not included because the VITAL table structure (1 record per result) does not support missingness assessment. The table depicts the percentage of records with missing or unknown values and supports Data Checks 3.03 and 3.04.

Table	Field	Encounter Type Constraint	Records with missing, NI, UN, or OT values			Data Check		
			Numerator	Denominator	%	Data Check	Threshold	Source
DEMOGRAPHIC	BIRTH_DATE		962	6,531,625	0.0	3.03	>=5%	DEM_L3_AGEYRSDIST1
DEMOGRAPHIC	SEX		5,022	6,531,625	0.1	3.03	>=5%	DEM_L3_SEXDIST
DEMOGRAPHIC	HISPANIC		669,010	6,531,625	10.2	--	--	DEM_L3_HISPDIST
DEMOGRAPHIC	RACE		2,116,245	6,531,625	32.4	3.04	>=15%	DEM_L3_RACEDIST
ENROLLMENT	ENR_END_DATE		0	7,420,370	.	--	--	ENR_L3_DIST_END
ENCOUNTER	DISCHARGE_DATE	IP, IS, EI	24	2,033,541	0.0	--	--	ENC_L3_ENCTYPE_DDATE_YM
ENCOUNTER	ENC_TYPE		17,649,997	129,462,028	13.6	--	--	ENC_L3_ENCTYPE
ENCOUNTER	PROVIDERID		4,580,622	129,462,028	3.5	--	--	ENC_L3_N
ENCOUNTER	DISCHARGE_DISPOSITION	IP, IS, EI	153,505	2,033,541	7.5	3.04	>=15%	ENC_L3_ENCTYPE_DISDISP
ENCOUNTER	DISCHARGE_STATUS	IP, IS, EI	19,837	2,033,541	1.0	--	--	ENC_L3_ENCTYPE_DISSTAT
ENCOUNTER	DRG	IP, IS, EI	569,827	2,033,541	28.0	--	--	ENC_L3_ENCTYPE_DRG
ENCOUNTER	ADMITTING_SOURCE	IP, IS, EI	518,095	2,033,541	25.5	--	--	ENC_L3_ENCTYPE_ADMRSC
DIAGNOSIS	DX_TYPE		4,681,027	247,509,332	1.9	3.03	>=5%	DIA_L3_DXTYPE_DXSOURCE
DIAGNOSIS	DX_SOURCE		0	247,509,332	.	--	--	DIA_L3_DXSOURCE
DIAGNOSIS	PDX	IP, IS, EI	0	17,180,449	.	3.04	>=15%	DIA_L3_PDX_ENCTYPE
PROCEDURES	PX_TYPE		36,506,754	316,287,433	11.5	3.03	>=5%	PRO_L3_PXTYPE_ENCTYPE
PROCEDURES	PX_SOURCE		0	316,287,433	.	--	--	PRO_L3_PXTYPE_SOURCE
VITAL	VITAL_SOURCE		0	86,859,677	.	3.03	>=5%	VIT_L3_VITAL_SOURCE

PX_TYPE

- ❁ We have 6.6M rows where PX_TYPE is missing, UN, NI, or OT
- ❁ The majority of the issue came from one data partner
- ❁ The PX_TYPE was not identified in their source data; and most of them are erroneously set as ICD9 or ICD10 (>6.3M rows)
- ❁ We researched the possible PX values to determine the PX TYPE
 - ICD-9 codes are three or four digit
 - ICD-10 codes are of length 7 (alphanumeric characters)
 - The majority of the codes in our data was of length 5 (i.e., CPT or HCPCS)

Fixed PX_TYPE in Procedures Data

- We have updated the PX_TYPE for an additional 2,162,343 procedures.
- We also fixed PX_TYPE erroneously set as ICD9 or ICD10 (>6.3M procedures).
- We used the following heuristic:

Pattern Description	Pattern (using SQL Server 'LIKE')	PX_TYPE
XXXXF: Supplemental Tracking Codes	[0-9][0-9][0-9][0-9]F	C2
XXXXX[,-][A-Z]: 5-digit + modifiers	992[0-9][0-9]%	C4
ACJXXXX: HCPCS Level II	[ACJ][0-9][0-9][0-9][0-9]	HC
DXXXX: dental codes	D[0-9][0-9][0-9][0-9]	HC

Remaining issues

- ❖ We fixed 80.6% of the procedures, but the remaining records with erroneous PX_TYPE are still problematic.
- ❖ In a recent study, an investigator was looking for HCV related procedures.
- ❖ We have very little hits based on the CPT codes and PX_TYPE=C4.
 - Wrong PX_TYPE with these CPT codes

Lab	CPT Codes
Platelet Counts	85049
ALT	84460
AST	84450
HCV Ab	86803
HCV RNA	87522, 87520, 87521
HCV Genotype	87902

PCORnet CDM v3.1 Development

*Shelley Rusincovitch
PCORnet DRN OC*

*Keith Marsolo, PhD
ImproveCareNow PPRN; PEDSnet CDRN
PCORnet DRN OC; PCORnet Data Committee member*

Common Data Model v3.1 development

- A minor release (v3.1) will be developed **in parallel** with development of the larger CDM strategy (more on this in the next section!)
- A **minor** release (by definition) does **not include** major new functionality
- We expect that the current CDM errata will be a primary source (<https://github.com/CDMFORUM/CDM-ERRATA>)
- We will build on successes in Phase I, and v3.1 offers an opportunity to further refine processes

Building on success from Phase I

Iterative Design and Modeling for the PCORnet Common Data Model v3.0

Shelley A. Rusincovitch, Jeffrey S. Brown, PhD, Keith Marsolo, PhD, Jenny Ibarra, RN, MSN, Abel N. Kho, MD, MS, Daniella Meeker, PhD, Lesley H. Curtis, PhD

S36: Infrastructure for
Data-driven Translational Science
Thursday, March 24, 2016

AMIA
2016 Joint Summits on Translational Science
March 21–24, San Francisco



pcornet

The National Patient-Centered Clinical Research Network

CDM in PCORnet Phase I

- Using an iterative design process, **we released 3 major versions** in a highly accelerated **16-month development period** (February 2014 through May 2015)
- Four feedback cycles with stakeholders and expert advisors
 - More than 700 discreet comments** received during development and incorporated into model
- The CDM v3.0 is the **final version within PCORnet Phase I**

Details of CDM Development

PCORnet CDM v1.0

- Released on **May 30, 2014**
- **276 discrete comments** received and incorporated in **1 feedback cycle**

PCORnet CDM v2.0

- Released on **February 27, 2015**
- **265 discrete comments** received and incorporated in **2 feedback cycles**
- Two stakeholder sessions with **79 and 60 attendees**, respectively

PCORnet CDM v3.0

- Released on **June 1, 2015**
- **236 discrete comments** received and incorporated in **1 feedback cycle**
- Two stakeholder sessions with **57 and 60 attendees**, respectively

Important dates for v3.1

We
are
here

<i>Date</i>	<i>Topic</i>
August 19	CDM Forum
August 19–September 2	Data Committee CDM survey
September 8	v3.1 interest group meeting (1 of 2)
September 19	v3.1 interest group meeting (2 of 2)
October 3–14	Feedback period on v3.1 draft specification
October 5	CDM Forum: Updates from interest groups
November 2	CDM Forum: Review v3.1 feedback, responses, and incorporation of changes

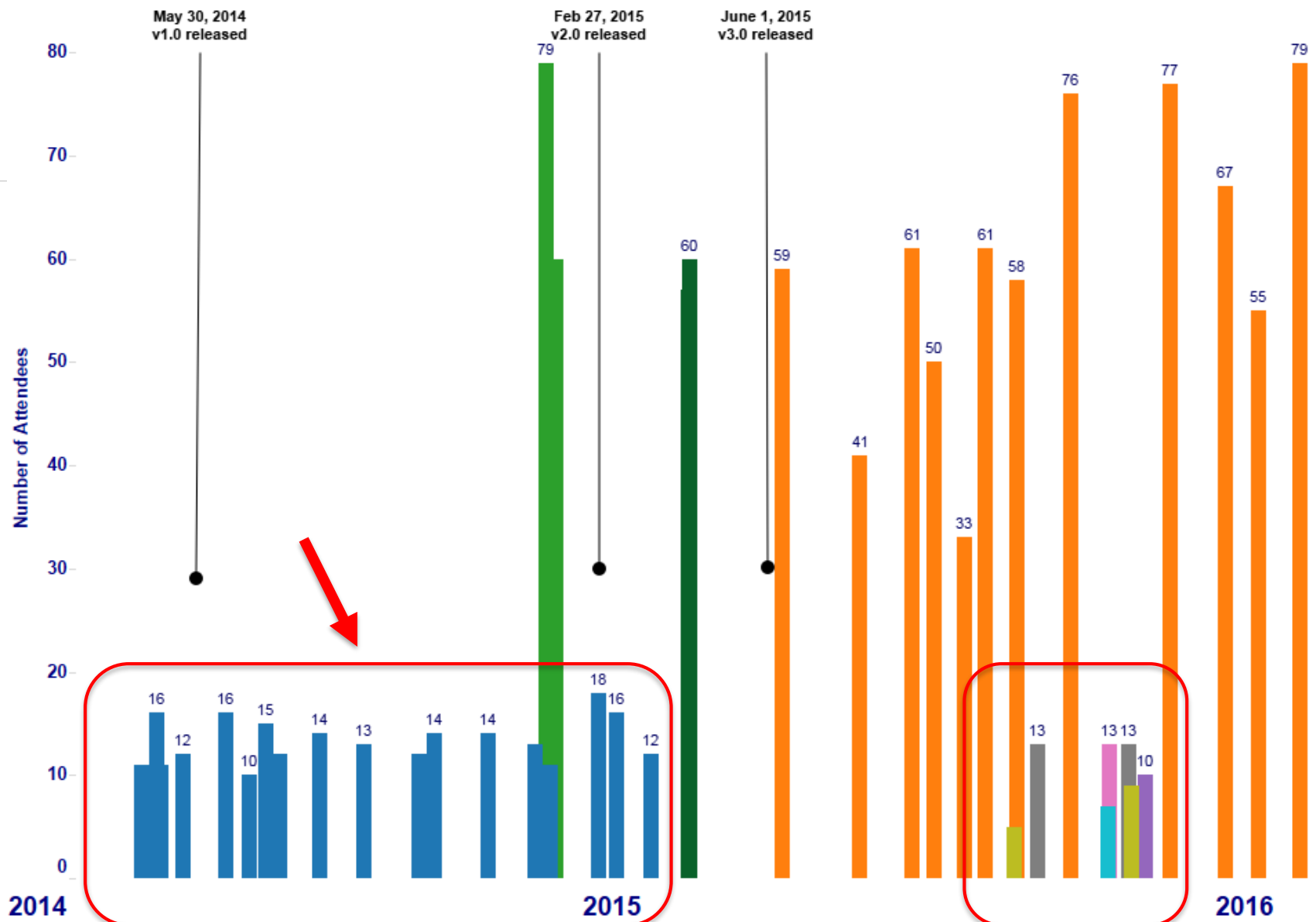
Also being scheduled for late October/early November:

- Presentation to CDRN and PPRN Principal Investigators (PIs)
- Presentation to Executive Committee
- Review by Council

Goal for CDM v3.1: approval & release by the end of 2016! But...

But a CDM release date isn't the same as a network implementation date

- PCORnet does not expect networks datamarts to implement CDM v3.1 in 2016
- Part of v3.1 development will be to assess implementation expectations
- We expect considerations for implementation will include datamart refresh cycles and study participation
- Please keep in mind that v3.1 will be a **minor** release; therefore, it will **not** include new tables



- Tobacco Interest Group
- Medication Mapping Interest Group
- Lab Mapping Interest Group
- Encounter Interest Group
- Condition Interest Group
- CDM Working Group
- CDM v3.0 Discussion
- CDM v2.0 Discussion
- CDM Forum

We are here

Invitation to participate in the v3.1 interest group

- 🌐 This group's expert feedback (drawn from the full PCORnet community) will be used to help shape the draft specification, prior to the full feedback cycle
- 🌐 We will hold 2 sessions in September:
 - Thursday, September 8
 - Monday, September 19
- 🌐 Please join us! To request a calendar invite, please send an e-mail to Ania Berchuck (ania.berchuck@duke.edu)
- 🌐 We will provide updates from the interest group during the full Forum on October 5

PCORnet Data Committee CDM Survey

*Rachel Hess, MD
University of Utah School of Medicine
PaTH CDRN; PCORnet Data Committee member*

The PCORnet Data Committee Charter

I. PURPOSE

The PCORnet Data Committee (PDC) is charged by the PCORnet Council is to oversee PCORnet's data network, which is managed by the PCORnet Coordinating Center. The PDC will provide strategic guidance that ensures the data network is aligned with PCORI's overall goals and that the data network provides the semantic and syntactic interoperability to support PCORnet studies and sustainability. The PDC is also charged with stimulating informatics and research data innovations that advance the goals of PCORnet. (Section 1.1.6 PCORnet Governance Policy).

Learn more about the Data Committee on GitHub:
<https://github.com/PCORnet/DataCommittee/wiki>

The PCORnet Data Committee and the CDM

VI.H: Develop annual updates of the Common Data Model

In collaboration with the Coordinating Center, create the processes and governance that will be used to specify annual updates to information model and domains that are instantiated in the Common Data model, which may include optional modules or new domains. These proposals will be presented to the Council for approval. Included in this process will be the factors used to evaluate whether to include a domain (e.g., effort needed to acquire data, analytic value, ability to standardize, etc.), the updates' cost/sustainability, and methods to gather input and evaluate work from external stakeholders (e.g. NIH/CTSA, FDA, HHS ONC, ODHSI, patients / families, healthcare systems, insurers, medical device, and pharmaceutical companies). The PDC will also be responsible for developing the governance needed to surface and harmonize CDM implementation decisions by individual institutions or networks, with the goal of facilitating high-quality, reproducible research. These efforts will ensure that the structure of the data model continues to serve the purpose of PCORnet and allows the most efficient conduct of research.

Development of a CDM Survey

- The Data Committee would like to understand experiences by networks who have extended the CDM
- The Data Committee will review results from the survey during our face-to-face meeting on September 22
- The survey results will be shared with the full PCORnet community

Please share your thoughts and
experiences with us!

Link to survey:

[https://duke.qualtrics.com/jfe/form/SV_3ruG
C078iEYUsap](https://duke.qualtrics.com/jfe/form/SV_3ruGC078iEYUsap)

Response requested by **Friday, September 2**

Wrap Up

Recap (the shortlist):

- 🌐 **CDM Survey:** https://duke.qualtrics.com/jfe/form/SV_3ruGC078iEYUsap
 - Response requested by Friday, **September 2**
- 🌐 September v3.1 interest group meetings:
 - Thursday, **September 8**
 - Monday, **September 19**
 - To request a calendar invite, please send an e-mail to Ania Berchuck (ania.berchuck@duke.edu)
- 🌐 Feedback cycle for the draft v3.1 specification: **October 3–14**
 - Request will go to PCORnet community, including CDM Forum
- 🌐 Fall CDM Forum meetings: **October 5** and **November 2**
 - Calendar invites coming shortly

References and links

- ❁ DRN OC home page on iMeetCentral:
<https://pcornet.imeetcentral.com/p/aQAAAAAB6T9b>
- ❁ CDM errata issue tracker: <https://github.com/CDMFORUM/CDM-ERRATA/issues>
- ❁ CDM guidance issue tracker:
<https://github.com/CDMFORUM/CDM-GUIDANCE/issues>
- ❁ PCORnet diagnostic query package: <https://github.com/PCORnet-DRN-OC/PCORnet-Diagnostic-Query>
- ❁ PCORnet data characterization query package:
<https://github.com/PCORnet-DRN-OC/PCORnet-Data-Characterization>
- ❁ PCORnet Data Committee on GitHub:
<https://github.com/PCORnet/DataCommittee>

AcademyHealth Concordium next month

PCORnet CDM posters:

- “Practices of Engagement in Developing and Implementing the PCORnet Common Data Model”
- “Representing a Lab Result – Developing a LOINC Reference Table for a Distributed Research Network”



AMIA 2015 poster: Processes from v1.0 and v2.0

Development of a National Distributed Research Network Data Infrastructure: Design of the PCORnet Common Data Model

Jeffrey S. Brown, PhD¹, Shelley A. Rusincovitch², Abel N. Kho, MD, MS³, Keith Marsolo, PhD⁴, and Lesley H. Curtis, PhD⁵

¹Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; ²Duke Translational Medicine Institute, Durham, NC; ³Northwestern University Departments of Medicine and Preventive Medicine, Evanston, IL; ⁴Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; ⁵Duke University School of Medicine, Department of Medicine, Durham, NC

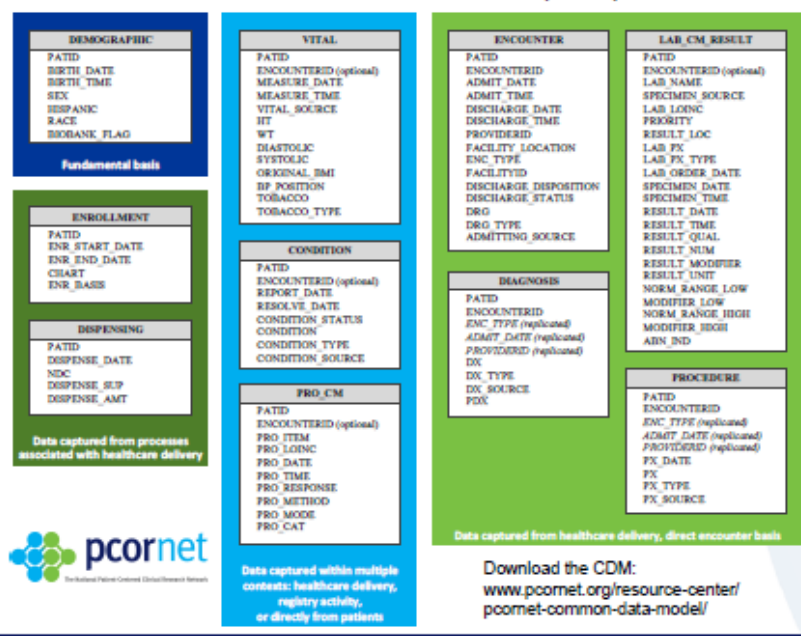
Introduction

- PCORnet is a national Distributed Research Network (DRN) sponsored by the Patient-Centered Outcomes Research Institute (PCORI)
- Formed in late 2013 with 29 networks located across the United States
- Objective is to create a "network of networks" to form a highly representative national infrastructure for research
- Uses the PopMedNet™ software infrastructure
- Architecture will allow queries and analytic programs to be disseminated to individual data stores
- The PCORnet Common Data Model (CDM) has been developed to provide a standard organization and representation

Methods

- Development began with the Guiding Principles, drafted as part of the charter of the PCORnet Data Standards, Security, and Network Infrastructure (DSSNI) Task Force
- The PCORnet CDM is a stand-alone model as a result of the strategic direction laid out in the Guiding Principles
- Heavily derived from the Mini-Sentinel Common Data Model and informed by other initiatives
- Developed in incremental phases, allowing agility in deployment, and new data domains to be incorporated
- Modeling is optimized for analytic functionality and to be intuitive to investigators

PCORnet Common Data Model (CDM) v2.0



Results

PCORnet CDM v1.0

- Released on May 30, 2014
- 276 discrete comments received and incorporated

PCORnet CDM v2.0

- Released on February 27, 2015
- 265 discrete comments received and incorporated in 2 feedback cycles
- Two stakeholder sessions with 79 and 60 attendees, respectively
- Unanimous vote for adoption by the Steering Committee

PCORnet CDM v3.0

- Currently in development

Discussion

- The CDM uses a strong basis of secondary data generated by healthcare delivery processes
- Development takes a pragmatic approach, supporting the current state of these data
- Area of future development is for data domains that support multiple contexts: healthcare delivery, registries and clinical studies, and data generated and received directly from patients
- Participation in review cycles, comment submission, and attendance demonstrates high engagement within the PCORnet community

Acknowledgments

The project described was supported by the PCORnet National Patient-Centered Clinical Research Network. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the Patient-Centered Outcomes Research Institute.

Contact Information

Jeff Brown, PhD, Associate Professor
Harvard Medical School,
Harvard Pilgrim Health Care Institute
jeff.brown@harvardpilgrim.org



AMIA 2015 poster: v2.0 domain selection process

Pragmatic Data Domain Selection for a National Distributed Research Network: The PCORnet Common Data Model Strategy

Shelley A. Rusincovitch¹, Abel N. Kho, MD, MS², Jon E. Puro, MPA:HA³, Daniella Meeker, PhD⁴, Pedro Rivera, MSCS³, Aaron A. Sorensen, MA⁵, Jeffrey S. Brown, PhD⁶, and Lesley H. Curtis, PhD⁷

¹Duke Translational Medicine Institute, Durham, NC; ²Northwestern University Departments of Medicine and Preventive Medicine, Evanston, IL; ³OCHIN, Inc., Portland, OR; ⁴Department of Health, RAND Corporation, Santa Monica, CA; ⁵Temple University School of Medicine, Philadelphia, PA; ⁶Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; ⁷Duke University School of Medicine, Department of Medicine, Durham, NC

Introduction

- The PCORnet Common Data Model (CDM) specifies the data foundation for PCORnet, and is developed with a **phase-based approach**
- Each phase incorporates **new concepts and data tables** to support distributed clinical research (observational and interventional)
- In order to establish priorities for subsequent CDM development, it was necessary to establish a **method of assessing new concepts and making decisions for inclusion** to serve the functional, pragmatic focus of the initiative

Methods

- The assessment was **organized by data domains**; i.e., the high-level concepts of data organization based upon existing data sources, workflows, and processes
- Assessment included best practices established by **existing data models and advice from external experts**
- Close attention to **PCORnet-specific requirements**

Contact Information

Shelley A. Rusincovitch
Project Leader in Applied Informatics
Duke Medicine
shelley.rusincovitch@duke.edu

Data Domain Evaluation Table

The use of "stoplight" colors conveys favorable (green), caution (yellow), or unfavorable (red)

Domain	Effort Needed to Acquire Data	Analytic Utility/Value	Ability to Standardize Across Sites	Availability Across Networks (Anecdotal)
Allergies and/or Contraindications	HIGH	MOD	LOW	MOD
Patient-reported Outcome (PRO) Common Measures	MOD	MOD	HIGH	N/A (prospective)
Condition	LOW	MOD	MOD	HIGH
Death and Death Cause	HIGH	HIGH	HIGH	LOW
Facility	MOD	MOD	LOW	HIGH
Family history	MOD	MOD	LOW	LOW
Inpatient Medication Administration	HIGH	LOW	LOW	LOW
Laboratory Result Common Measures	MOD	HIGH	LOW	HIGH
Medication Reconciliation	MOD	MOD	MOD	HIGH
Outpatient Pharmacy Dispensing	MOD	HIGH	HIGH	LOW
Primary Care Provider (PCP)	MOD	MOD	MOD	MOD
Provider Orders (including Medication Orders)	MOD	MOD	LOW	MOD
Social History & Lifestyle Choices	HIGH	LOW	LOW	MOD
State Vaccine	HIGH	LOW	LOW	MOD
Study Enrollment	LOW	HIGH	MOD	N/A (prospective)
Study Visits	MOD	MOD	MOD	N/A (prospective)

Data Model Landscape Scanning

- Mini-Sentinel Common Data Model, v4.0
- i2b2 Data Repository Cell, v1.7.00
- OMOP Common Data Model, v5.0
- HMORN VDM, v3.2
- ESPnet Data Form, 2013
- National Quality Forum, Quality Data Model, Version 4.1.1.

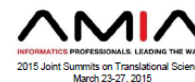
Results

- Assessment resulted in recommendation for **prioritization of data domains**
- Initiative has subsequently **formalized a process** for stakeholder review, facilitated discussion, and the approval process for adoption

Discussion

- Key lesson learned: **Importance of identifying and articulating foundational strategic decisions**, including interoperability within the analytic framework
- Limitation: Dimensions were **assessed by the CDM Working Group**, rather than by a formal survey of all participants; however, the working group represented individuals with the deep expertise necessary to make informed recommendations

Acknowledgments The project described was supported by the PCORnet National Patient-Centered Clinical Research Network. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the Patient-Centered Outcomes Research Institute.



<https://github.com/CDMFORUM/CDM-GUIDANCE/blob/master/2015-03-20.AMIA-CRI.2015.data.domains.poster.pdf>



pcorner[®]

The National Patient-Centered Clinical Research Network