S29: PCORnet Implementation of PopMedNet Data Assessment Tools

Jessica Sturtevant, ScM

Harvard Pilgrim Health Care Institute & Harvard Medical School

PCORnet Distributed Research Network Operations Center (DRN OC)

AMIA

www.amia.org

Disclosure

- No disclosure or relationships with commercial interests.

Learning Objectives

- Evaluate extensibility of the PopMedNet Rapid Data Assessment Tool (RDAT) across different distributed research networks, diverse data sources and different common data models.

PCORnet Implementation of PopMedNet Data Assessment Tools

PCORNET

www.amia.org

PCORnet Data Networks

11 (13) Clinical Data Research Networks (CDRNs)

+

18 (20) Patient-Powered Research Networks (PPRNs)

=

PCORnet Phase 1 (Phase 2):

A national infrastructure for patient-centered clinical research

PCORnet Goal

PCORnet seeks to improve the nation's capacity to conduct clinical research by creating a large, highly representative, national patient-centered network that supports more efficient clinical trials and observational studies.

www.amia.org

DataMart Geographic Distribution

This map depicts the number of PCORI funded Patient-Powered or Clinical Data Research Networks that have coverage in each state.

PCORnet Distributed Research Network

- Must address confidentiality, privacy, regulatory and proprietary concerns of numerous data partners

- Distributed queries collect and report on aggregate information

Data Characterization

- Examine Common Data Model variables, formats, time trends

- Categorical variables examined using frequencies

- Continuous variables examined using distributions
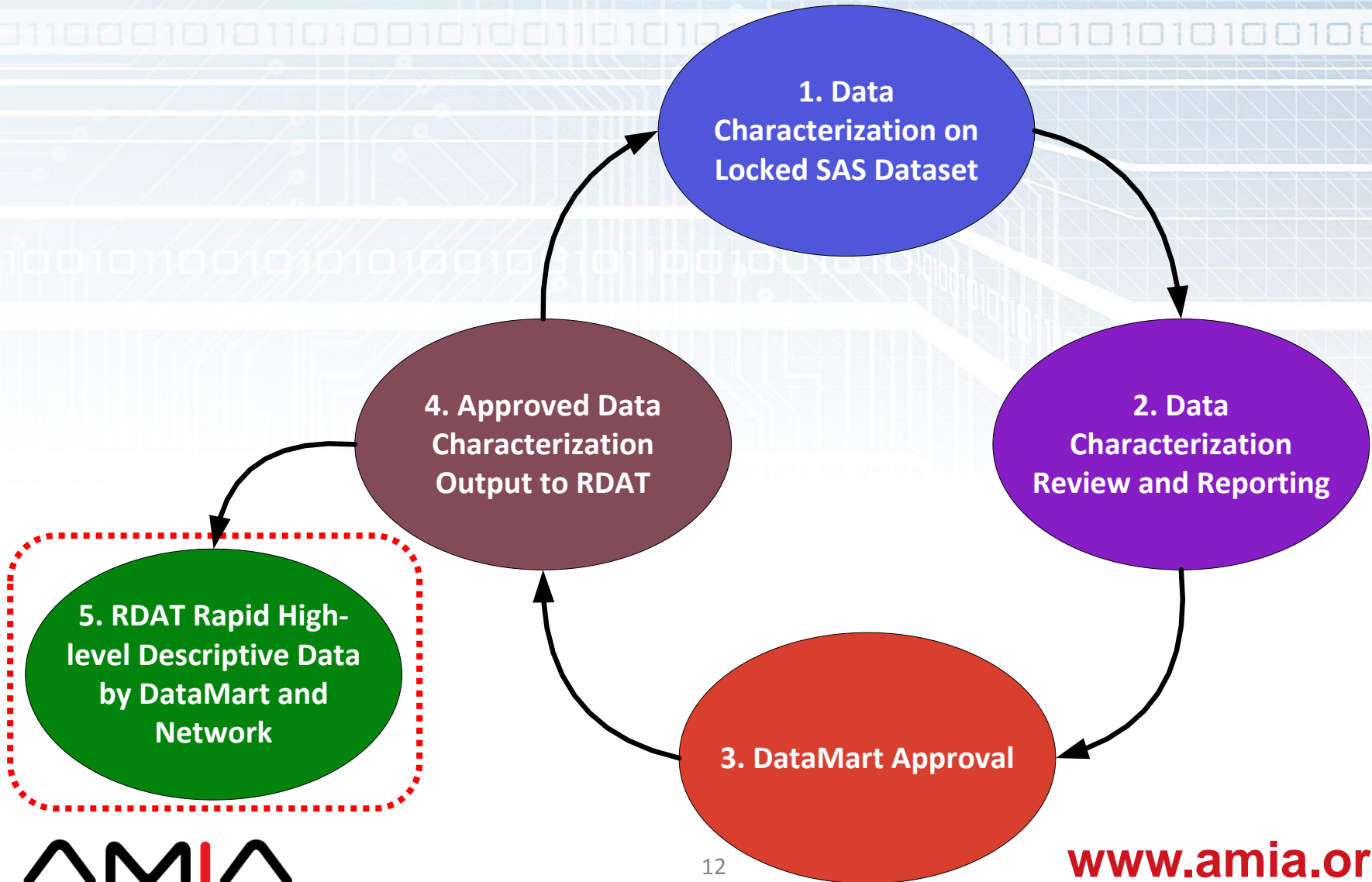
- Output is summary level/aggregate

PCORnet Implementation of PopMedNet Data Assessment Tools

RAPID DATA ASSESSMENT TOOL

Distributed Research Network Operations Center (DRN OC)

- DRN OC performs data characterization (DC) for each DataMart (DM)

- DRN OC stores aggregate DC output locally

- Approved, aggregate DC output queried directly via the PCORnet Rapid Data Assessment Tool (RDAT)

- RDAT facilitates rapid, easy access to descriptive data by DataMart and Network

www.amia.org

DRN OC Data Characterization Cycle

1. Data Characterization on Locked SAS Dataset

2. Data Characterization Review and Reporting

3. DataMart Approval

4. Approved Data Characterization Output to RDAT

5. RDAT Rapid High-level Descriptive Data by DataMart and Network
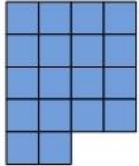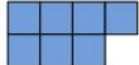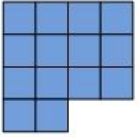
PCORnet DRN Query Tool: Rapid Data Assessment Tool (RDAT)

- PopMedNet (PMN)
  - Implemented for PCORnet Distributed Research Network (DRN) in 2014
- RDAT – PMN Query
  - Locally queries aggregate DC output
  - Supports preparatory-to-research (PTR)
  - Facilitates rapid data assessment to support clinical trials and observational research

www.amia.org

Standardized Data Characterization Output Model
RDAT Cross Network Extensibility

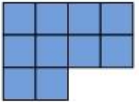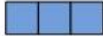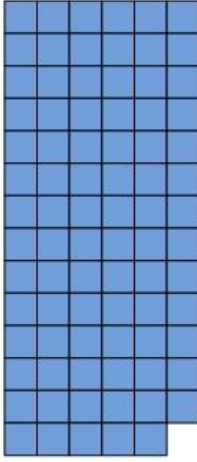| Mini-Sentinel	NIH Collaboratory DRN	Health Data Collaboration	MDPHnet	PCORnet
Mini-Sentinel	NIH Collaboratory Health Care Systems Research Collaboratory	health care systems research network ☀CRN	MDPH NET	pcornet The National Patient-Centered Clinical Research Network
Medical product safety surveillance	Research	Preparatory-to-research Feasibility	Public health surveillance	Research
18 Health Insurers/ Integrated Delivery Systems	**Feasibility Project** 7 Health Insurers/ Integrated Delivery Systems	**HCSRNnet** 14 Health Insurers/ Integrated Delivery Systems	3 Multi-Site Medical Group Practices	**11 Clinical Data Research Networks** contributing 64 DataMarts **18 Patient-Powered Research Networks** contributing 19 DataMarts
	LIRE Project 4 Health Insurers/ Integrated Delivery Systems	**CRNnet** 10 Health Insurers/ Integrated Delivery Systems		
		IMEDS-Pfizer 10 Health Insurers/ Integrated Delivery Systems		Current as of 10/27/2015

Rapid Data Assessment Tool (RDAT)

- Leveraged capabilities developed by the FDA Sentinel Project
  - Race
  - Ethnicity
  - Data Completeness
- PCORnet Enhancements
  - Age Distribution
  - Sex
  - Weight Distribution
  - Height Distribution
  - Diagnosis Codes
  - Procedure Codes

www.amia.org

RDAT Development

PopMedNet Query Type:
Rapid Data Assessment Tool (RDAT)

Leverage query terms developed by other PMN networks:

- Race
- Data
- Ethnicity

Develop query terms for PCORnet:

- Age
- Sex
- Height
- Weight
- DX
- PX

Future query term development:

- Smoking
- Date Range
- DX, PX, etc.
- Dispensing

Typical RDAT Use

1. Question CDM Tables & Variables

2. RDAT Query of pertinent variables

3. High-level Descriptive Data by DataMart and Network

4. Descriptive Data and Charts used to Assess Data Usability

5. Inform DataMart Use in Studies and Trials
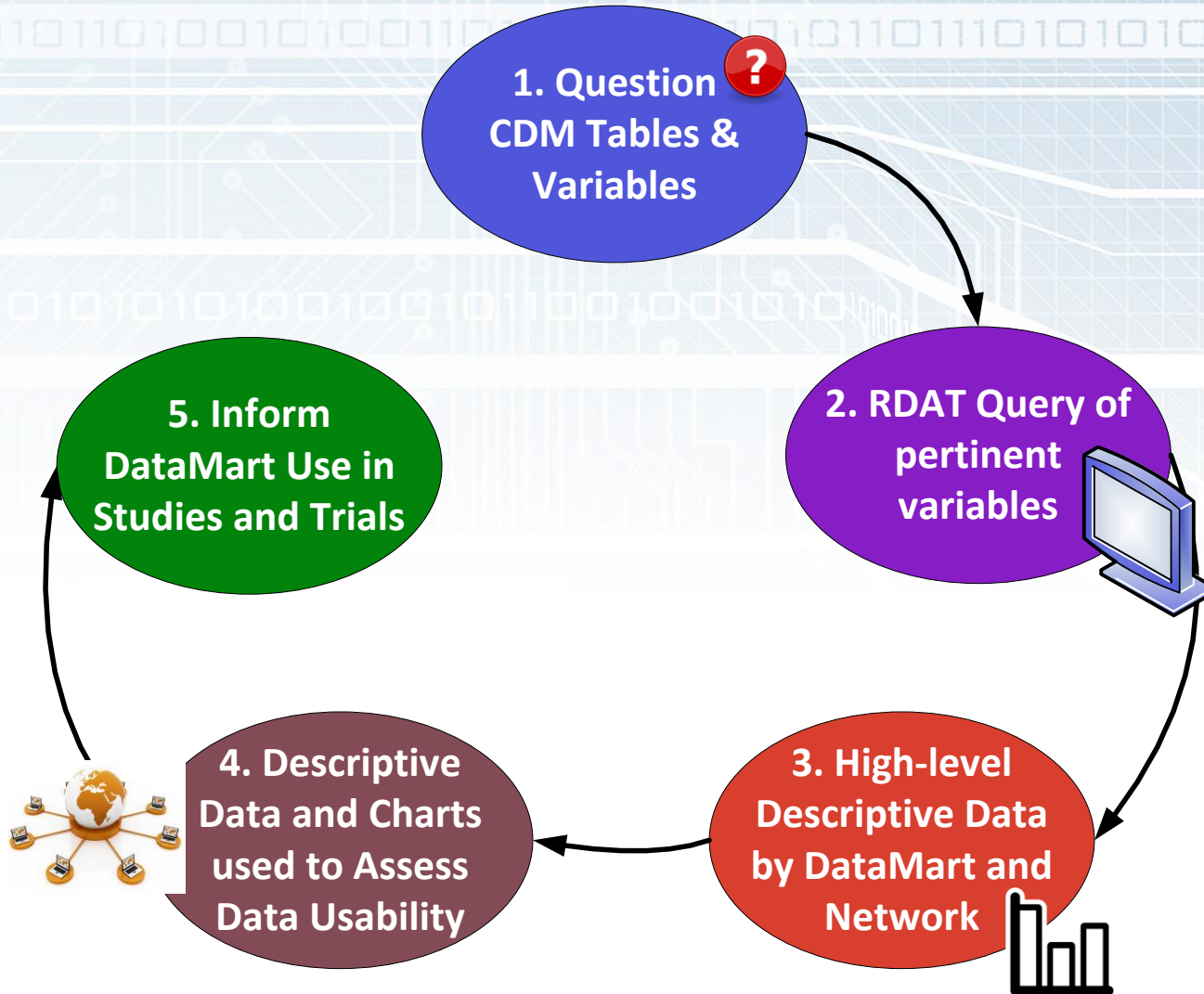
RDAT Query Response

www.amia.org
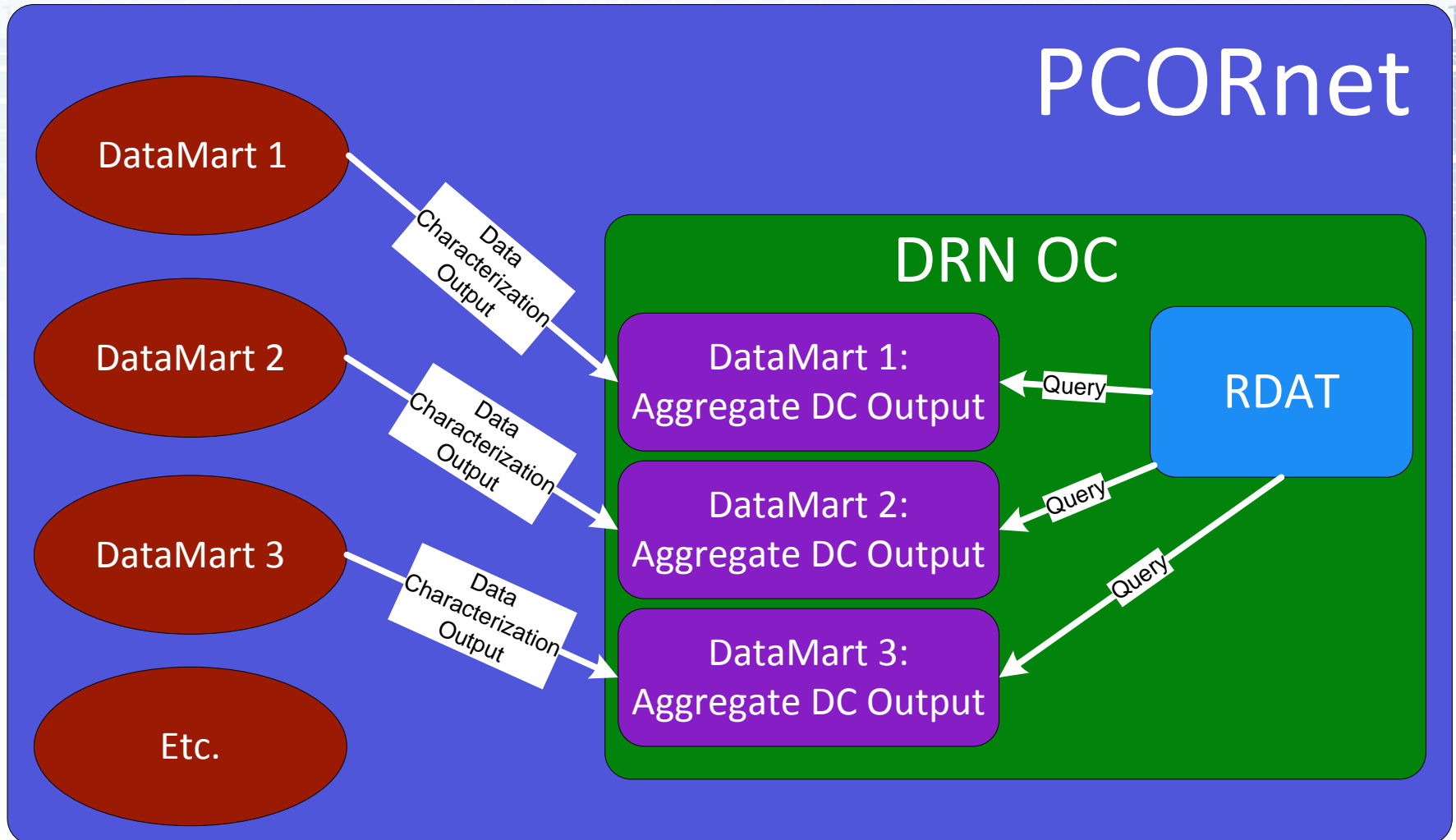
Rapid Data Assessment Tool

SINGLE DATAMART QUERY

The 15 PCORnet CDM Domains, v3.0

CONDITION v2.0

A condition represents a patient's diagnosed and self-reported health conditions and diseases. The patient's medical history and current state may both be represented.

DEATH v3.0

Reported mortality information for patients.

DEATH_CAUSE v3.0

The individual causes associated with a reported death.

DEMOGRAPHIC v1.0

Demographics record the direct attributes of individual patients.

DIAGNOSIS v1.0

Diagnosis codes indicate the results of diagnostic processes and medical coding within healthcare delivery.

DISPENSING v2.0

Outpatient pharmacy dispensing, such as prescriptions filled through a neighborhood pharmacy with a claim paid by an insurer. Outpatient dispensing is not commonly captured within healthcare systems.

ENROLLMENT v1.0

Enrollment is a concept that defines a period of time during which all medically-attended events are expected to be observed. This concept is often insurance-based, but other methods of defining enrollment are possible.

ENCOUNTER v1.0

Encounters are interactions between patients and providers within the context of healthcare delivery.

HARVEST v3.0

Attributes associated with the specific PCORnet datamart implementation

LAB_RESULT_CM v2.0

Laboratory result Common Measures (CM) use specific types of quantitative and qualitative measurements from blood and other body specimens. These standardized measures are defined in the same way across all PCORnet networks.

PCORNET_TRIAL v3.0

Patients who are enrolled in PCORnet clinical trials.

PRESCRIBING v3.0

Provider orders for medication dispensing and/or administration.

PRO_CM v2.0

Patient-Reported Outcome (PRO) Common Measures (CM) are standardized measures that are defined in the same way across all PCORnet networks. Each measure is recorded at the individual item level: an individual question/statement, paired with its standardized response options.

PROCEDURES v1.0

Procedure codes indicate the discreet medical interventions and diagnostic testing, such as surgical procedures, administered within healthcare delivery.
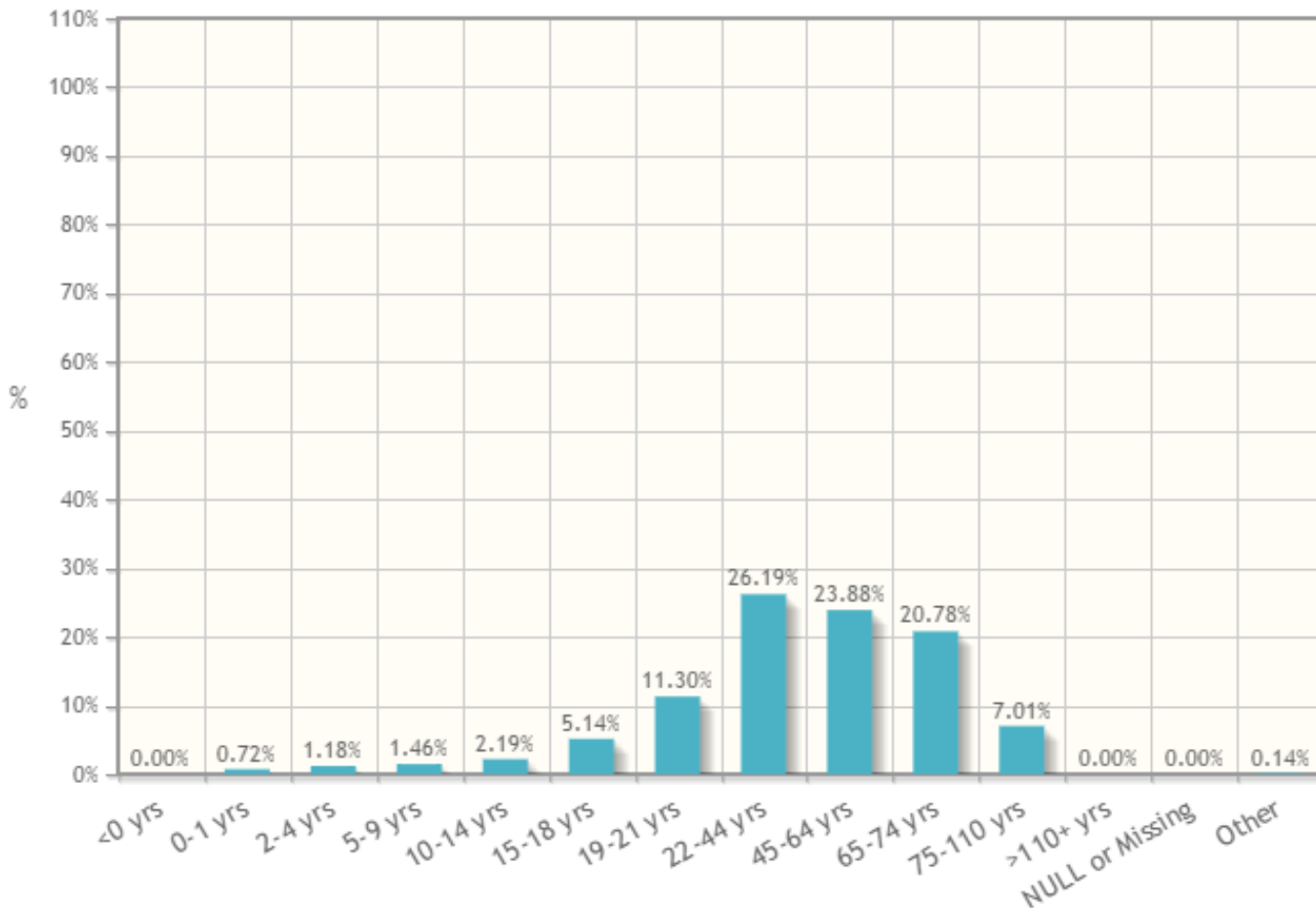
VITAL v1.0

Vital signs (such as height, weight, and blood pressure) directly measure an individual's current state of attributes.

www.amia.org

Age Group
Rapid Data Assessment Tool Query

www.amia.org

Race Frequency

Race	n	%
Unknown	88648115	62.78
American Indian or Alaska Native	247958	0.18
Asian	2937282	2.08
Black or African American	9249691	6.55
Native Hawaiian or Other Pacific Islander	592332	0.42
White	39538729	28
Missing	2	0

Race Distribution among Selected Data Partners*
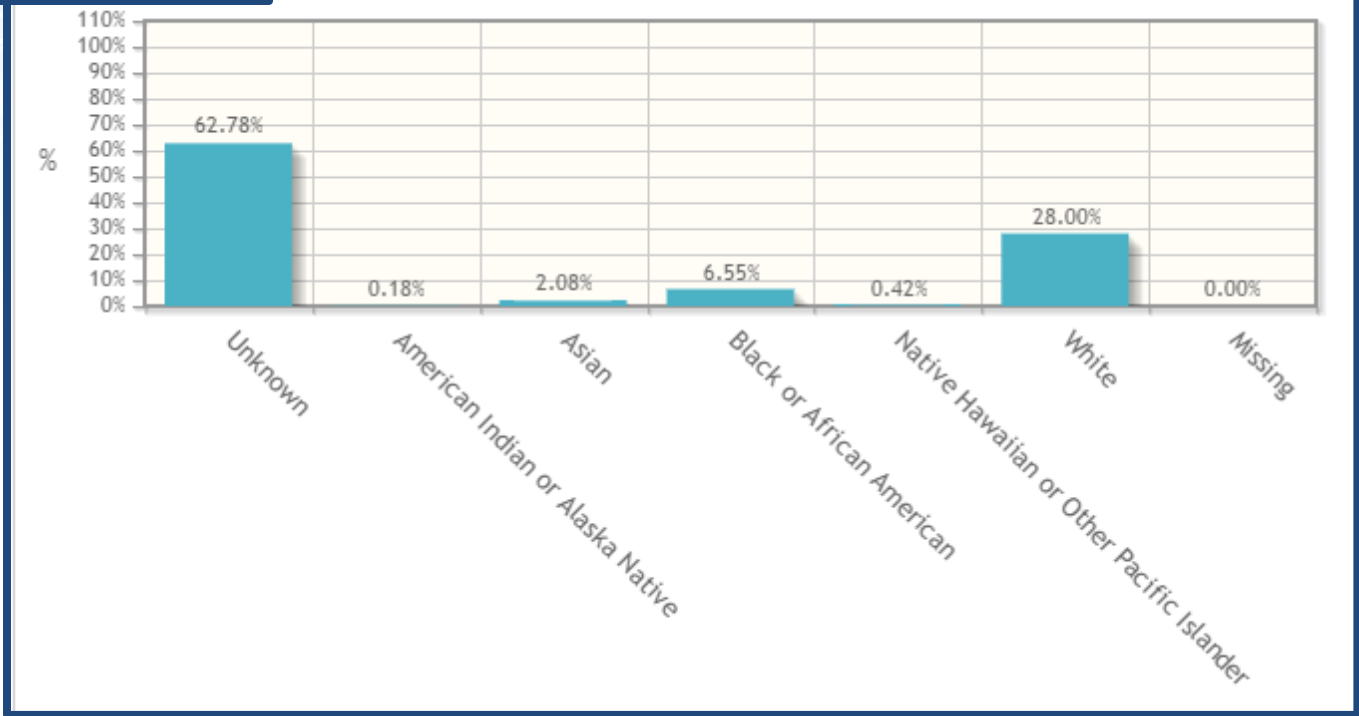
Sex Frequency

Sex	n	%
Ambiguous	312	0.04
Female	366549	44.52
Male	360638	43.80
No Information	35286	4.29
NULL or Missing	26379	3.20
Other	194	0.02
Unknown	6699	0.81
Values outside of CDM specifications	27304	3.32
Unselected Categories	0	0.00

Sex Distribution among Selected Data Partners*

www.amia.org

Weight
Rapid Data Assessment Tool Query

Weight Distribution among Selected Data Partners*

www.amia.org

Height Frequency

Height	n	%
0-10 in	0	0.00
11-20 in	0	0.00
21-45 in	11062	1.34
46-52 in	24449	2.97
53-58 in	78345	9.51
59-64 in	211150	25.64
65-70 in	245806	29.85
71-76 in	210697	25.58
77-82 in	13916	1.69
83-88 in	558	0.07
89-94 in	318	0.04
<0 in	0	0.00
NULL or Missing	20737	2.52
95+ in	4	0.00
Other	6519	0.79

Height Distribution among Selected Data Partners*
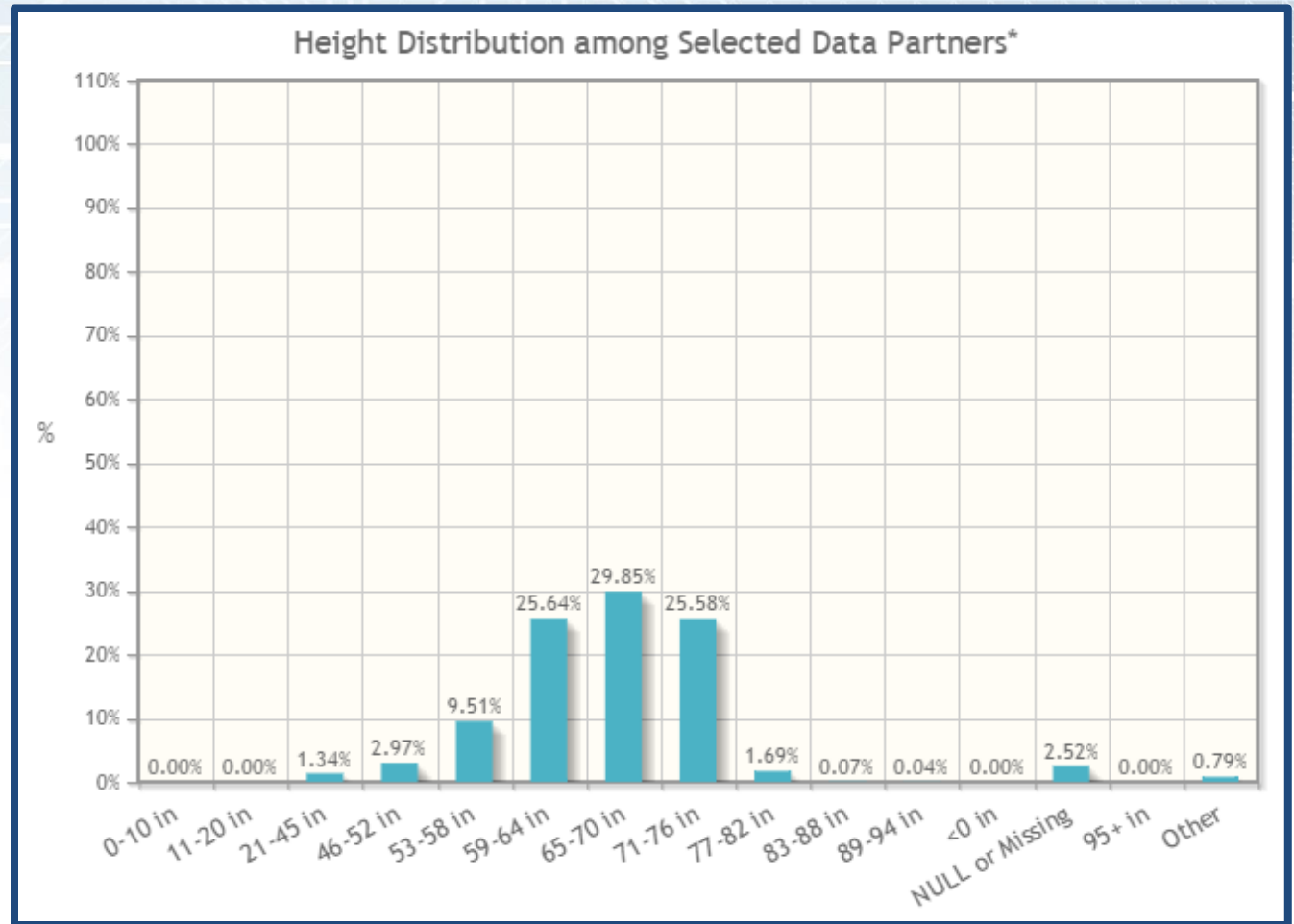
Rapid Data Assessment Tool

MULTIPLE DATAMART QUERY

www.amia.org

Sex Frequency Across DataMarts

Sex	AAAA		BBBB		CCCC	
	N	%	N	%	N	%
Ambiguous	40	0.03	78	0.06	11	0.01
Female	54561	43.11	60421	47.43	61224	46.29
Male	52861	41.77	58233	45.71	61412	46.43
No Information	260	0.21	1562	1.23	712	0.54
NULL or Missing	12145	9.6	1123	0.88	53	0.04
Other	25	0.02	12	0.01	124	0.09
Unknown	160	0.13	5609	4.4	841	0.64
Values outside of CDM specifications	6502	5.14	351	0.28	7891	5.97
Unselected Categories	0	0	0	0	0	0

Percent within Data Partner ▼

www.amia.org

Sex Frequency Across DataMarts

Sex Distribution within AAAA
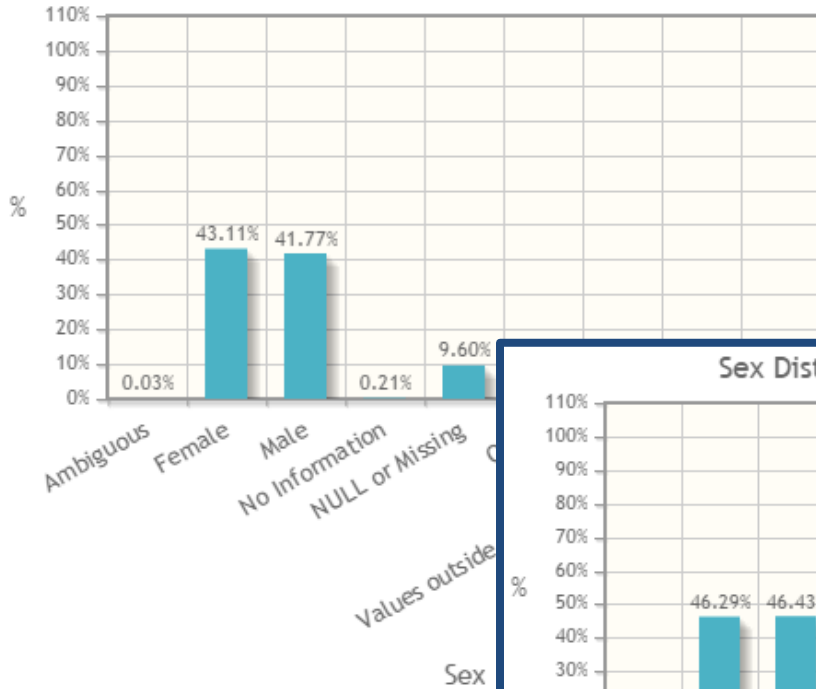
Sex Distribution within BBBB
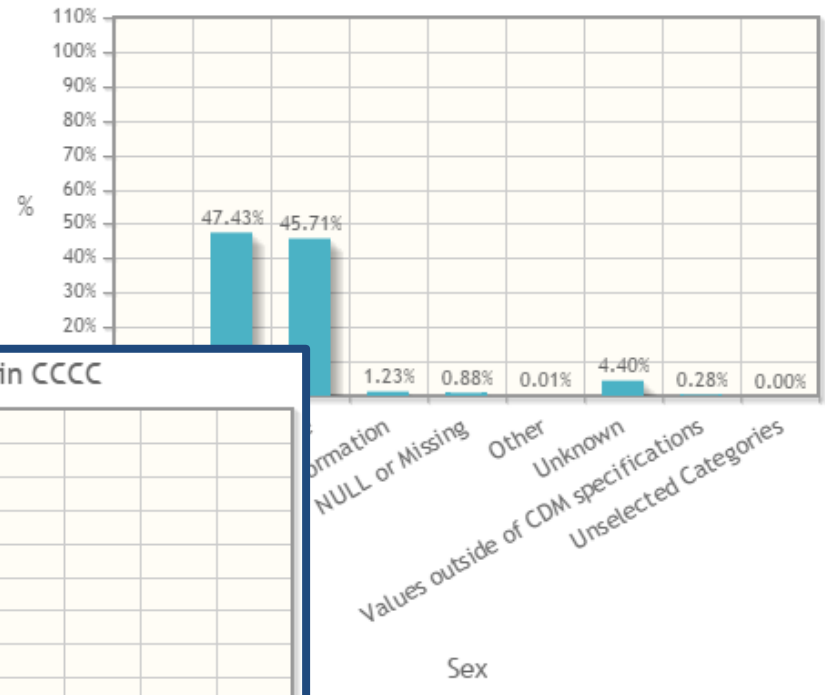
Sex Distribution within CCCC

Age Group Across DataMarts

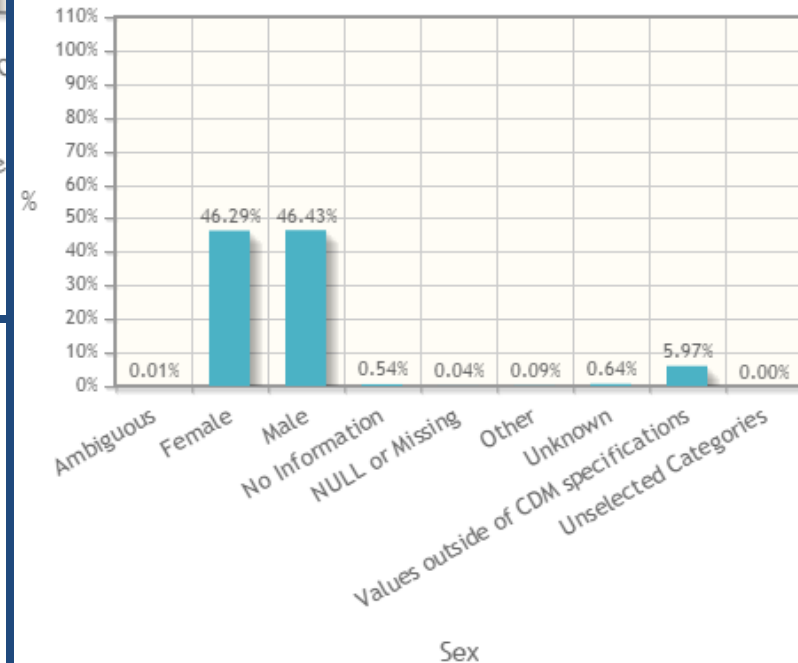| Age	AAAA		BBBB		CCCC	
	N	%	N	%	N	%
<0 yrs	0	0	0	0	0	0
0-1 yrs	1203	0.95	964	0.76	1123	0.86
2-4 yrs	2064	1.63	2531	1.99	1456	1.11
5-9 yrs	2237	1.77	2658	2.09	2389	1.82
10-14 yrs	3072	2.43	2987	2.35	2578	1.97
15-18 yrs	7748	6.13	7849	6.16	6998	5.35
19-21 yrs	12816	10.13	14523	11.4	13677	10.45
22-44 yrs	37892	29.96	36176	28.4	35236	26.91
45-64 yrs	26783	21.18	29535	23.19	31442	24.02
65-74 yrs	23564	18.63	22527	17.69	30941	23.63
75-110 yrs	8921	7.05	7411	5.82	5023	3.84
>110+ yrs	0	0	0	0	0	0
NULL or Missing	0	0	0	0	0	0
Other	157	0.12	206	0.16	56	0.04

Percent within Data Partner ▼

Age Group Across DataMarts

Age Distribution within AAAA
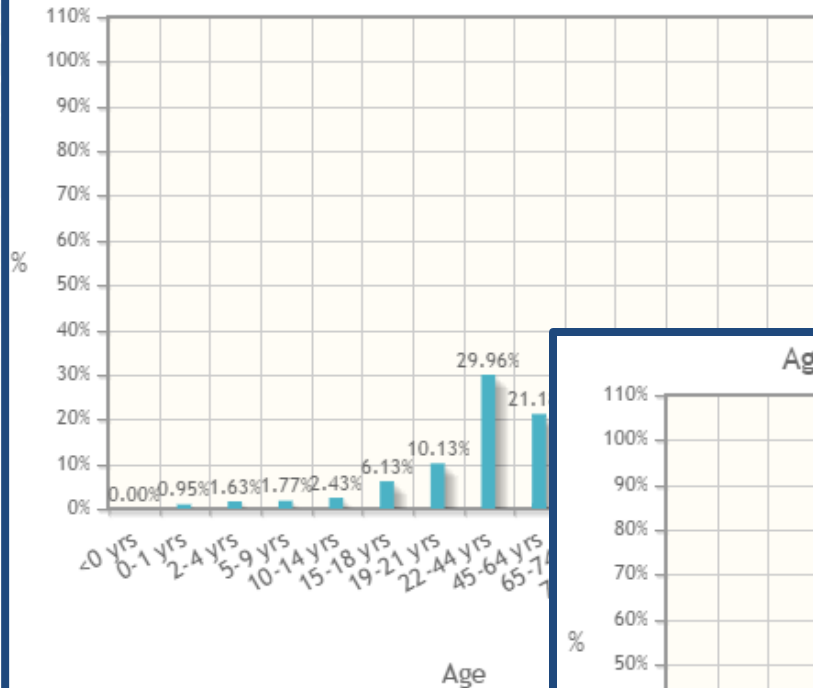
Age Distribution within BBBB

Age Distribution within CCCC

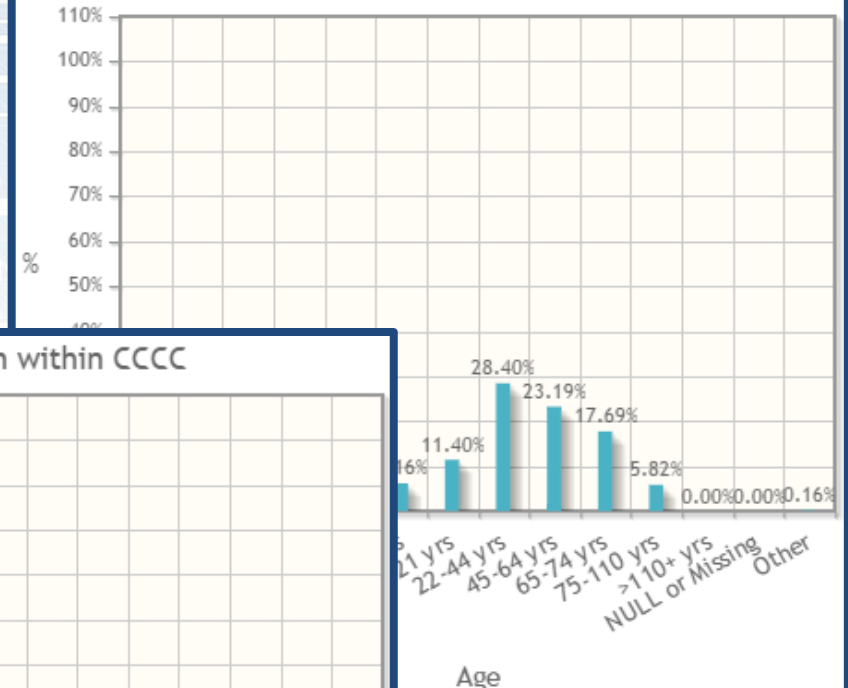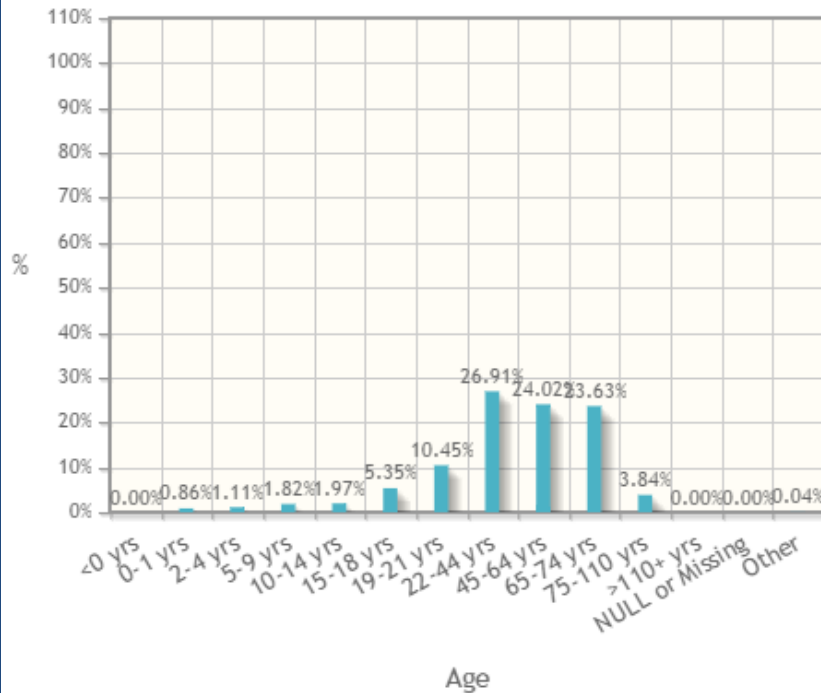# Weight Group Across DataMarts

Weight	AAAA		BBBB		CCCC	
Percent within Data Partner	N	%	N	%	N	%
<0 lbs	0	0	0	0	0	0
0-1 lbs	0	0	0	0	0	0
2-6 lbs	34	0.03	2	0	0	0
7-12 lbs	121	0.09	294	0.23	12	0.01
13-20 lbs	285	0.22	487	0.38	256	0.19
21-35 lbs	512	0.39	632	0.49	333	0.25
36-50 lbs	1254	0.95	768	0.6	1256	0.95
51-75 lbs	2771	2.11	1063	0.83	2893	2.19
75-100 lbs	3699	2.81	2741	2.14	11714	8.86
101-125 lbs	12186	9.27	14602	11.41	14502	10.96
126-150 lbs	13239	10.07	19369	15.14	14337	10.84
151-175 lbs	14997	11.41	27172	21.24	21836	16.51
176-200 lbs	21658	16.48	25132	19.64	21184	16.01
201-225 lbs	20782	15.81	25461	19.9	20069	15.17
226-250 lbs	12543	9.54	5623	4.39	14404	10.89
251-275 lbs	10356	7.88	2004	1.57	8506	6.43
276-300 lbs	9804	7.46	1134	0.89	30	0.02
301-350 lbs	0	0	0	0	0	0
NULL or Missing	3252	2.47	999	0.78	441	0.33
350+ lbs	386	0.29	22	0.02	471	0.36
Other	3577	2.72	451	0.35	39	0.03

Percent within Data Partner ▼

Weight Group Across DataMarts

Weight Distribution within AAAA
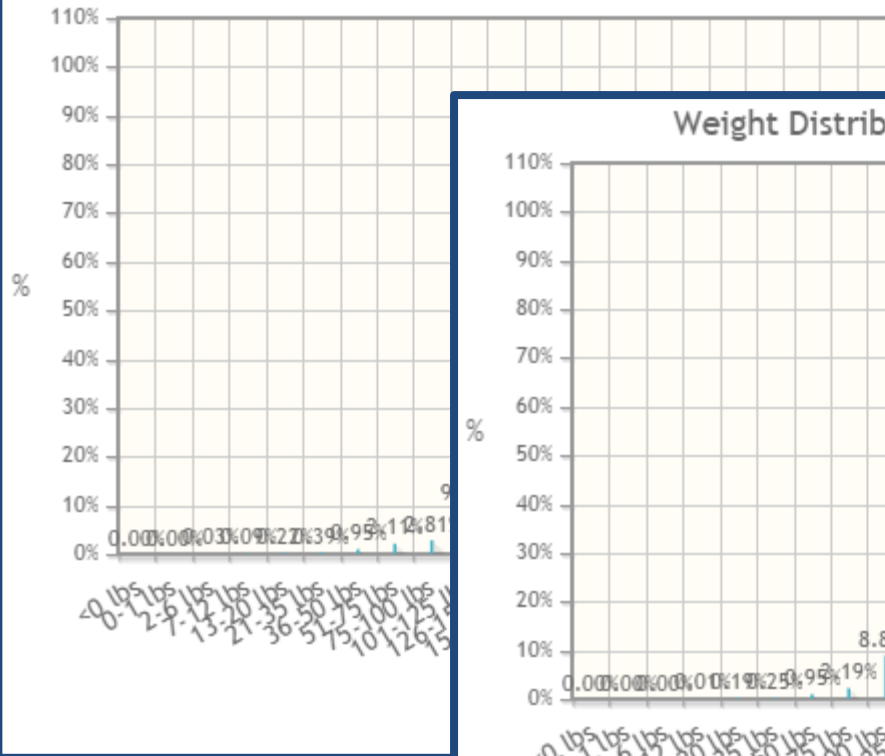
Weight Distribution within CCCC
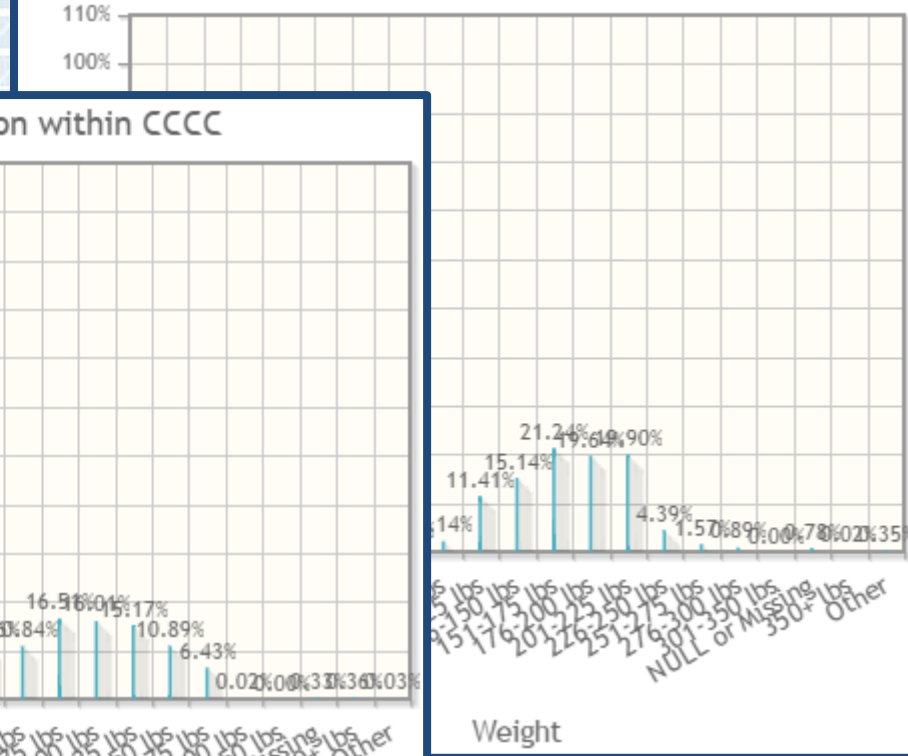
Weight Distribution within BBBB
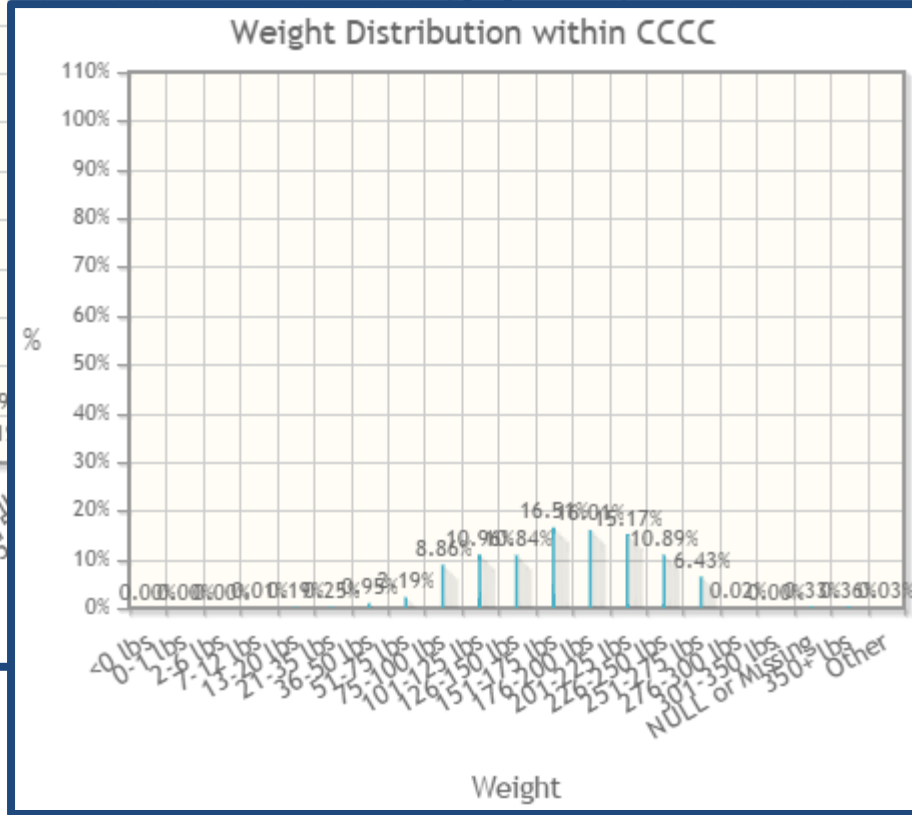
RDAT Development Next Steps

- Develop access controls for viewing DM, CDRN or aggregate network results

- Add terms such as smoking, dispensing, etc.

- Enhance visualizations

AMIA

PCORnet Next Steps

- Inform PCORnet DataMart participation in observational research studies and clinical trials

- Rapid access to data characterization results for PTR

- Inform study design (e.g. missingness, ranges, variable population)

www.amia.org

Questions

Thank you!

Jessica Sturtevant

jessica.lynn@post.harvard.edu

Acknowledgements

The PCORnet Distributed Research Network is a product of many data partners and collaborators. We gratefully acknowledge their work and individual contributions.

- Harvard Pilgrim Health Care Institute (HPHCI)
  - Jeffrey Brown
  - Kyle Erickson
  - Chayim Herzig-Marx
  - Casie Horgan
  - Lindsey Petro
  - Gabrielle Purcell
  - Jose Rodriguez
- Patient-Centered Outcomes Research Institute (PCORI)
  - Maryan Zirkle

- Duke Medicine
  - Lesley Curtis
  - Ed Hammond
  - Deb Harris
  - Jenny Ibarra
  - Michael Park
  - Laura Qualls
  - Shelley Rusincovitch
  - James Topping

www.amia.org

Further Reading

- Data Quality/DRN
  - Holmes J.H., et al., Developing a distributed research network to conduct population-based studies and safety surveillance. AMIA Annu Symp Proc, 2008. 973. http://www.ncbi.nlm.nih.gov/pubmed/18999251
  - Kahn M.G., et al., Transparent reporting of data quality in distributed data networks. eGEMs, 2015. 3:1(7). http://www.ncbi.nlm.nih.gov/pubmed/25992385
  - Kahn M.G., et al., A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. MedCare, 2012. 50(0). http://www.ncbi.nlm.nih.gov/pubmed/22692254
  - Maro J.C., et al., Design of a national distributed health data network. Ann Intern Med, 2009. 151(5): 341-4. http://www.ncbi.nlm.nih.gov/pubmed/19638403
  - Brown J.S., M. Kahn, and S. Toh, Data quality assessment for comparative effectiveness research in distributed data networks. Med Care, 2013. 51: S22-9. http://www.ncbi.nlm.nih.gov/pubmed/23793049
  - Weiskopf N.G., Weng C., Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc, 2013. 20(1): 144-51. http://www.ncbi.nlm.nih.gov/pubmed/22733976
  - Curtis L.H., J. Brown, R. Platt, Four health data networks illustrate the potential for a shared national multipurpose big-data network. Health Aff (Millwood), 2014. 33(7): 1178-86. http://www.ncbi.nlm.nih.gov/pubmed/25006144
  - Curtis L.H., et al., Design considerations, architecture, and use of the Mini-Sentinel distributed data system. Pharmacoepidemiol Drug Saf, 2012. 21 Suppl 1: 23-31. http://www.ncbi.nlm.nih.gov/pubmed/22262590

- PCORnet
  - PCORnet: The National Patient-Centered Clinical Research Network. Available from: http://pcornet.org/.
  - Collins, F.S., et al., PCORnet: turning a dream into reality. J Am Med Inform Assoc, 2014. 21(4): 576-7. http://www.ncbi.nlm.nih.gov/pubmed/24821744
  - Califf, R.M., The Patient-Centered Outcomes Research Network: a national infrastructure for comparative effectiveness research. N C Med J, 2014. 75(3): 204-10. http://www.ncbi.nlm.nih.gov/pubmed/24830497
  - Curtis, L.H., J. Brown, and R. Platt, Four health data networks illustrate the potential for a shared national multipurpose big-data network. Health Aff (Millwood), 2014. 33(7): 1178-86. http://www.ncbi.nlm.nih.gov/pubmed/25006144
  - Fleurence, R.L., et al., Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc, 2014. 21(4): 578-82. http://www.ncbi.nlm.nih.gov/pubmed/24821743

www.amia.org