[S106] Representation of Drug Use in Biomedical Standards, Clinical Text, and Research Measures

Elizabeth W. Carter, MS[1], Indra Neil Sarkar, PhD, MLIS[2], Genevieve B. Melton, MD, PhD[3], Elizabeth S. Chen, PhD[2]

[1]University of Vermont, [2]Brown University, [3]University of Minnesota,

AMIA

www.amia.org

Disclosure

No relationships with
commercial interests to disclose.

Learning Objectives

- After participating in this activity, the learner should be better able to:
  - Identify data sources for studying the representation of social and behavioral factors
  - Describe an approach for analyzing data sources to identify elements and values for a particular social or behavioral factor
  - Understand the current collection of drug use information in disparate data sources

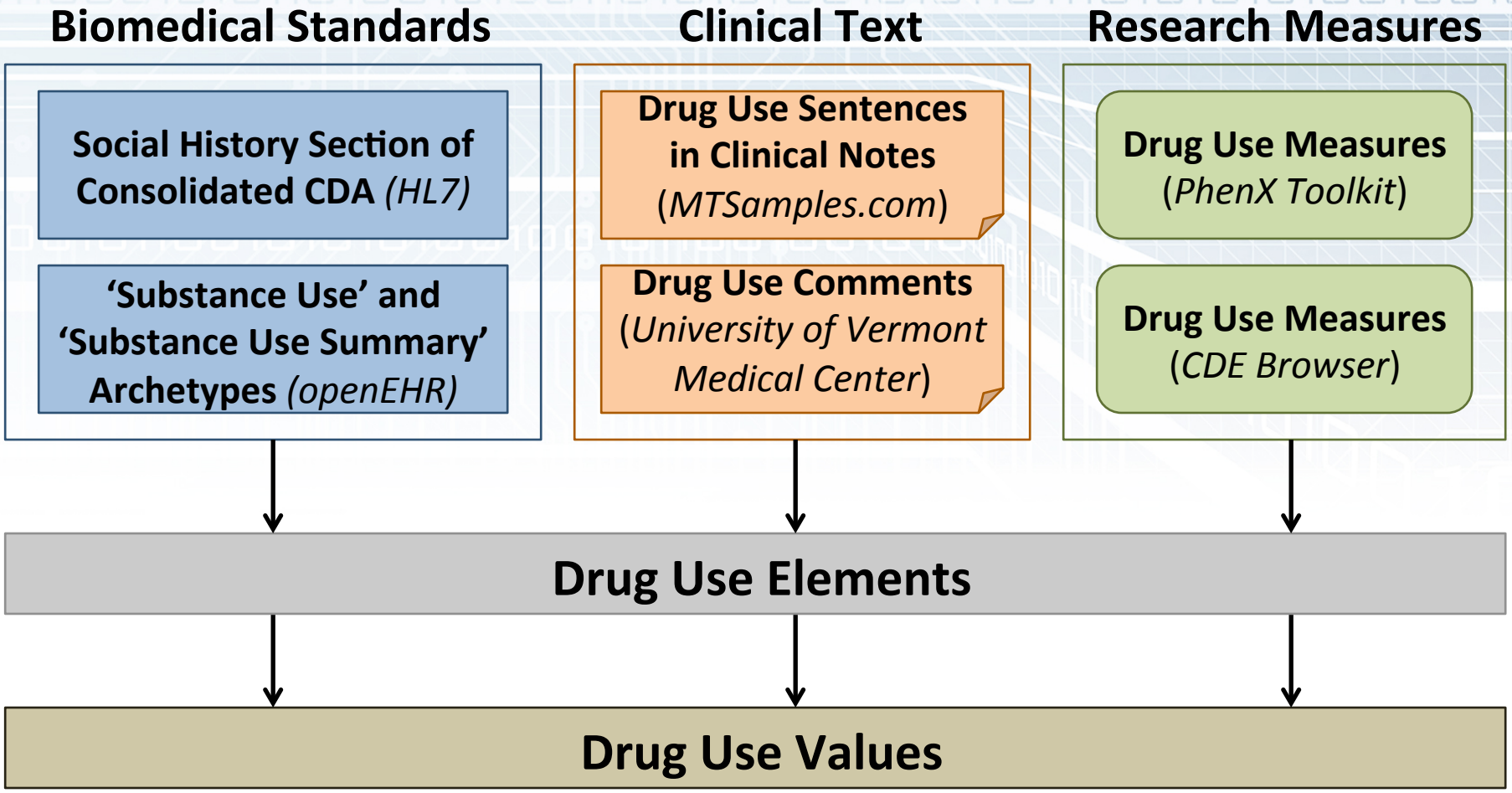
Introduction

- Drug use is a prominent cause of morbidity and mortality in the United States
- Potential of electronic health record (EHR) systems to support drug use screening, diagnosis, intervention, and treatment
- Need for standardized collection and subsequent use of drug use information
- Study objective: examine representation of drug use in multiple sources for informing EHR and standards development

Prior Studies

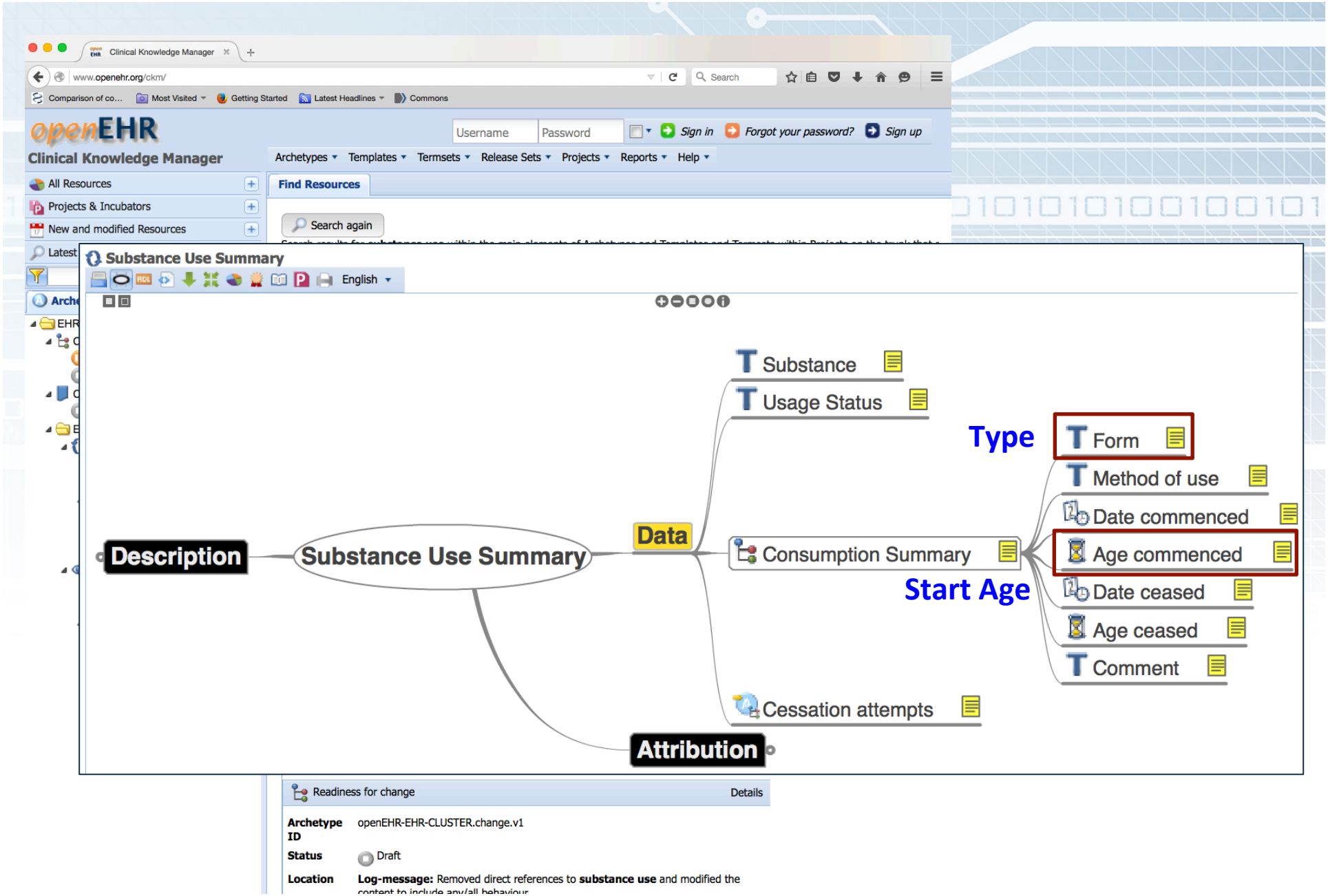
- Categorize social history in clinical notes – **drug use** among top 8 statement types (Chen *et al.*, 2011)
- Model social and behavioral factors, including **drug use**, in public health surveys (Melton *et al.*, 2012)
- Analyze contents of free-text **alcohol use** documentation in the EHR (Chen and Garcia-Webb, 2014)
- Analyze contents and quality of free-text **tobacco use** documentation in the EHR (Chen *et al.*, 2014)
- Extract **alcohol, drug, and tobacco use** information from clinical notes using natural language processing techniques (Wang *et al.*, 2015)

Overview of Approach



Analysis of Biomedical Standards

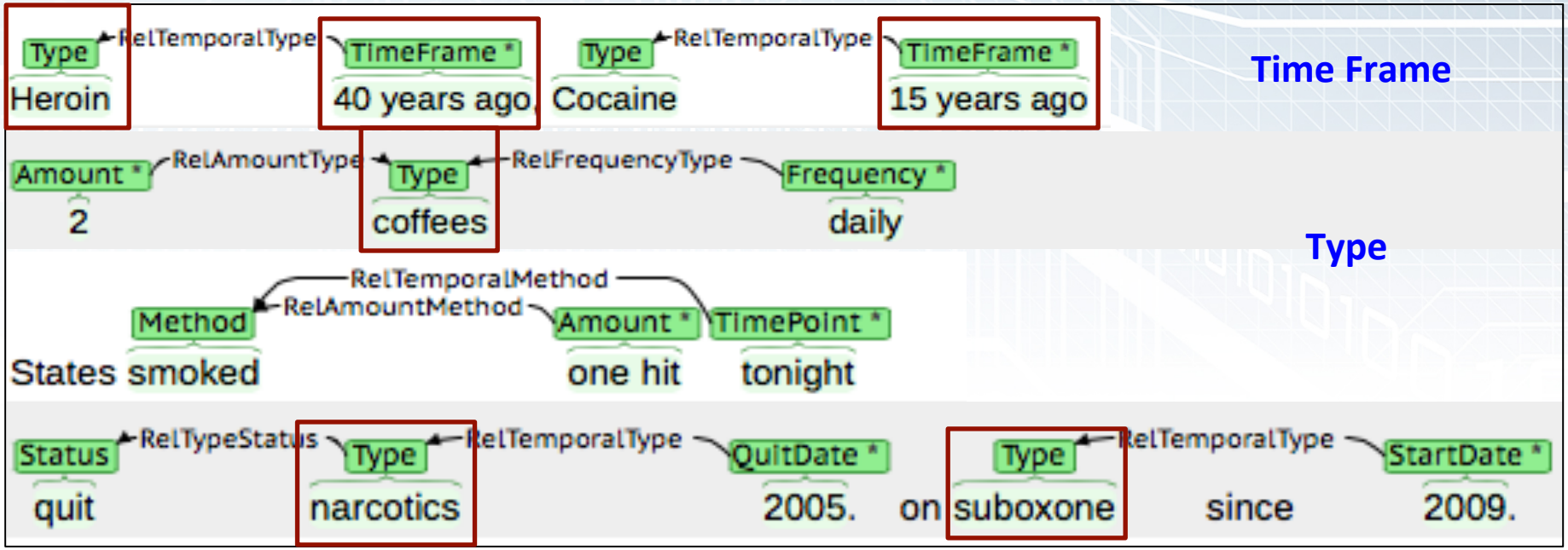
- Review existing standards for elements and values related to drug and substance use
  - Social History Observation template in HL7 Consolidated CDA implementation guide
  - Substance Use Summary and Substance Use archetypes in openEHR Clinical Knowledge Manager
- Merge elements with those identified in prior studies
- Use combined set of elements to create initial drug use annotation schema/guidelines for analyzing clinical text and research measures


Clinical Knowledge Manager
www.openehr.org/ckm/
openEHR
Clinical Knowledge Manager
Archetypes
Templates
Termsets
Release Sets
Projects
Reports
Help
All Resources
Projects & Incubators
New and modified Resources
Find Resources
Search again
Substance Use Summary
English
Substance
Usage Status
Type
Form
Method of use
Date commenced
Age commenced
Start Age
Date ceased
Age ceased
Comment
Description
Substance Use Summary
Data
Consumption Summary
Cessation attempts
Attribution
Readiness for change
Details
Archetype ID
openEHR-EHR-CLUSTER.change.v1
Status
Draft
Location
Log-message: Removed direct references to substance use and modified the

Analysis of Clinical Text

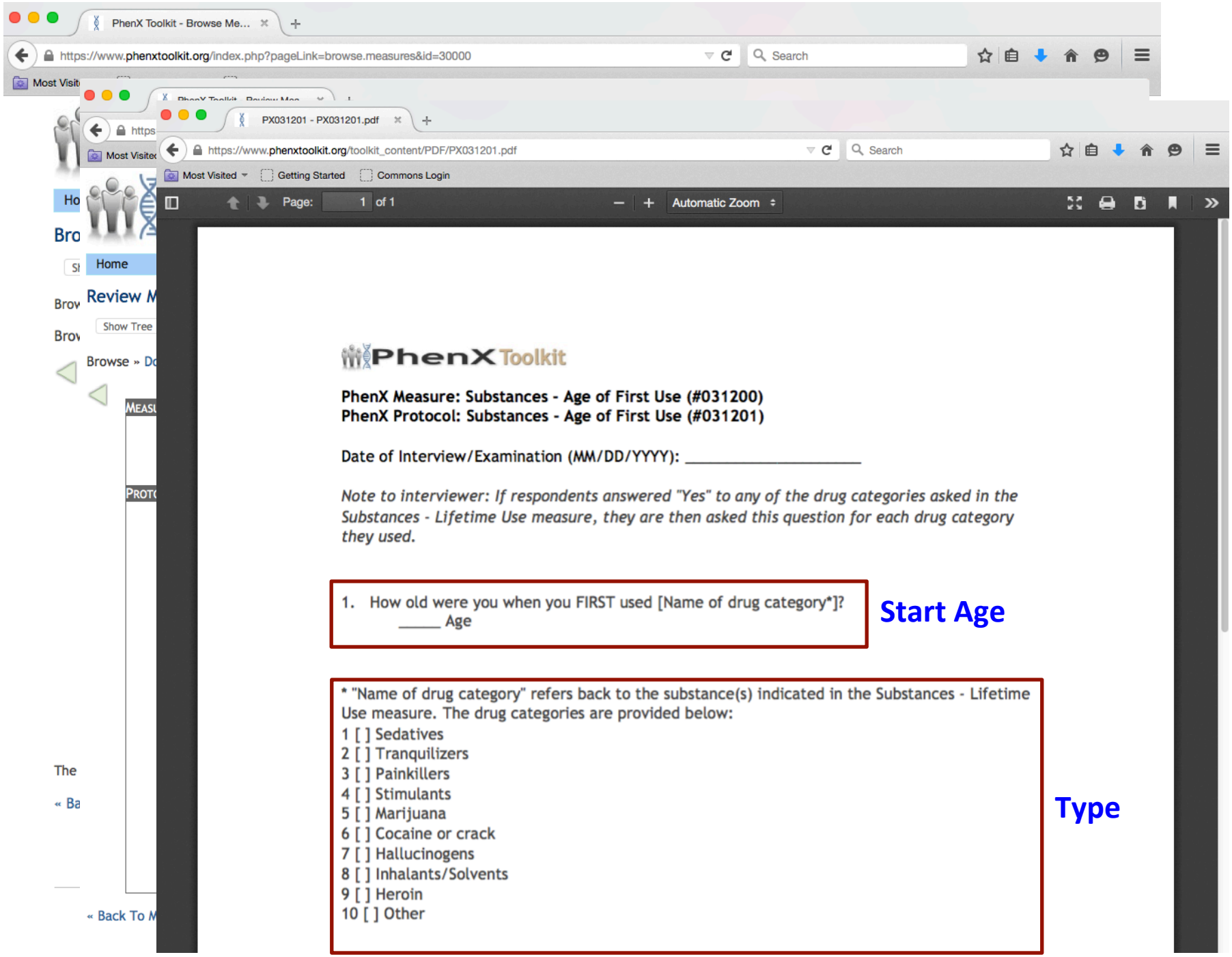
- Identify drug use sentences in consult notes from MTSamples.com (MTS) using GATE
- Extract drug use comments from Epic EHR social history module at University of Vermont Medical Center (UVMMC)
- Enhance drug use annotation schema/guidelines based on review of a random sample of drug use comments
- Annotate entities and relationships in drug use sentences and comments using BRAT
- Combine and group MTS and UVMMC annotations based on common words, phrases, or patterns

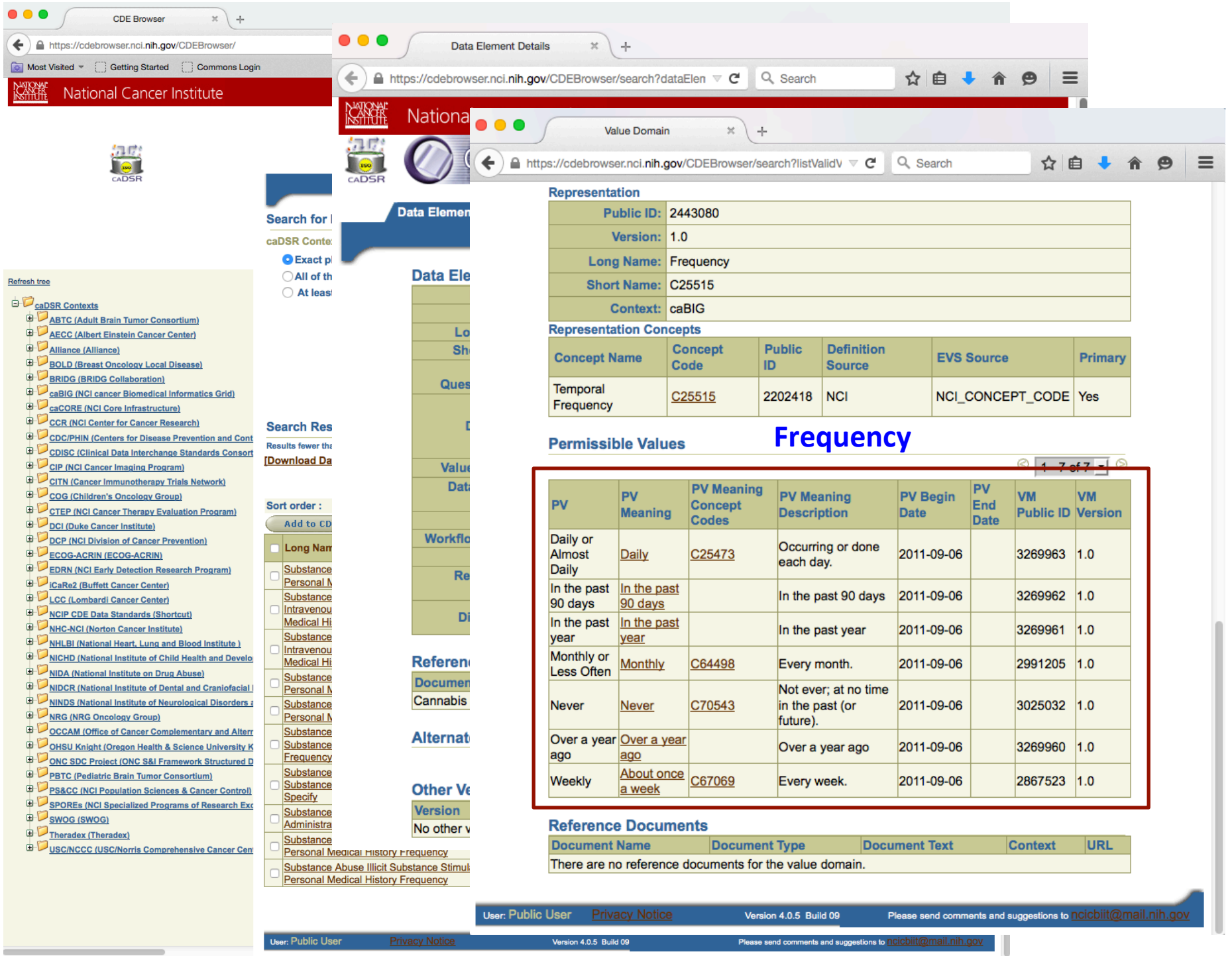
Annotation of Clinical Text



Analysis of Research Measures

- Search for measures in Consensus Measures for Phenotypes and eXposures (PhenX) Toolkit, particularly in the "Alcohol, Tobacco and Other Substances" domain
- Search for measures in caDSR using the Common Data Element (CDE) Browser
- Exclude measures not focused on drug use (e.g., those explicitly for substance abuse and cessation or opinions about drug use and abuse)
- Extract questions and responses associated with each remaining drug use-specific measure
- Annotate questions and responses using enhanced drug use annotation schema/guidelines
- Combine and group PhenX Toolkit and CDE Browser annotations based on common words, phrases, or patterns


PhenX Toolkit - Browse Me...
https://www.phenxtoolkit.org/index.php?pageLink=browse.measures&id=30000
Search
PX031201 - PX031201.pdf
https://www.phenxtoolkit.org/toolkit_content/PDF/PX031201.pdf
Most Visited
Getting Started
Commons Login
Page: 1 of 1
Automatic Zoom
Home
Review M
Show Tree
Browse » Do
« Back To M
PhenX Toolkit
PhenX Measure: Substances - Age of First Use (#031200)
PhenX Protocol: Substances - Age of First Use (#031201)
Date of Interview/Examination (MM/DD/YYYY): _____
Note to interviewer: If respondents answered "Yes" to any of the drug categories asked in the Substances - Lifetime Use measure, they are then asked this question for each drug category they used.
1. How old were you when you FIRST used [Name of drug category*]?
_____ Age
Start Age
* "Name of drug category" refers back to the substance(s) indicated in the Substances - Lifetime Use measure. The drug categories are provided below:
1 [ ] Sedatives
2 [ ] Tranquilizers
3 [ ] Painkillers
4 [ ] Stimulants
5 [ ] Marijuana
6 [ ] Cocaine or crack
7 [ ] Hallucinogens
8 [ ] Inhalants/Solvents
9 [ ] Heroin
10 [ ] Other
Type


CDE Browser
https://cdebrowser.nci.nih.gov/CDEBrowser/
Most Visited
Getting Started
Commons Login
National Cancer Institute
caDSR
Refresh tree
caDSR Contexts
ABTC (Adult Brain Tumor Consortium)
AECC (Albert Einstein Cancer Center)
Alliance (Alliance)
BOLD (Breast Oncology Local Disease)
BRIDG (BRIDG Collaboration)
caBIG (NCI cancer Biomedical Informatics Grid)
caCORE (NCI Core Infrastructure)
CCR (NCI Center for Cancer Research)
CDC/PHIN (Centers for Disease Prevention and Cont
CDISC (Clinical Data Interchange Standards Consort
CIP (NCI Cancer Imaging Program)
CITN (Cancer Immunotherapy Trials Network)
COG (Children's Oncology Group)
CTEP (NCI Cancer Therapy Evaluation Program)
DCI (Duke Cancer Institute)
DCP (NCI Division of Cancer Prevention)
ECOG-ACRIN (ECOG-ACRIN)
EDRN (NCI Early Detection Research Program)
iCaRe2 (Buffett Cancer Center)
LCC (Lombardi Cancer Center)
NCIP CDE Data Standards (Shortcut)
NHC-NCI (Norton Cancer Institute)
NHLBI (National Heart, Lung and Blood Institute)
NICHD (National Institute of Child Health and Develo
NIDA (National Institute on Drug Abuse)
NIDCR (National Institute of Dental and Craniofacial
NINDS (National Institute of Neurological Disorders
NRG (NRG Oncology Group)
OCCAM (Office of Cancer Complementary and Alterr
OHSU Knight (Oregon Health & Science University K
ONC SDC Project (ONC S&I Framework Structured D
PBTC (Pediatric Brain Tumor Consortium)
PS&CC (NCI Population Sciences & Cancer Control)
SPOREs (NCI Specialized Programs of Research Exc
SWOG (SWOG)
Theradex (Theradex)
USC/NCCC (USC/Norris Comprehensive Cancer Cen
Data Element Details
https://cdebrowser.nci.nih.gov/CDEBrowser/search?dataElen
Search
Search for
caDSR Conte
Exact p
All of th
At leas
Search Res
Results fewer tha
[Download Da
Sort order :
Add to CD
Long Nam
Substance Personal M
Substance Intravenou Medical Hi
Substance Intravenou Medical Hi
Substance Personal M
Substance Personal M
Substance Substance Frequency
Substance Substance Specify
Substance Administra
Substance Personal Medical History Frequency
Substance Abuse Illicit Substance Stimul Personal Medical History Frequency
Data Elemen
Data Ele
Lo
Sh
Ques
Valu
Dat
Workflo
Re
Di
Referen
Documen
Cannabis
Alternat
Other Ve
Version
No other
User: Public User
Privacy Notice
Version 4.0.5 Build 09
Please send comments and suggestions to ncicbiit@mail.nih.gov
Value Domain
https://cdebrowser.nci.nih.gov/CDEBrowser/search?listValidV
Search
Representation
Public ID: 2443080
Version: 1.0
Long Name: Frequency
Short Name: C25515
Context: caBIG
Representation Concepts
Concept Name | Concept Code | Public ID | Definition Source | EVS Source | Primary
Temporal Frequency | C25515 | 2202418 | NCI | NCI_CONCEPT_CODE | Yes
Permissible Values
Frequency
1 - 7 of 7
PV | PV Meaning | PV Meaning Concept Codes | PV Meaning Description | PV Begin Date | PV End Date | VM Public ID | VM Version
Daily or Almost Daily | Daily | C25473 | Occurring or done each day. | 2011-09-06 | | 3269963 | 1.0
In the past 90 days | In the past 90 days | | In the past 90 days | 2011-09-06 | | 3269962 | 1.0
In the past year | In the past year | | In the past year | 2011-09-06 | | 3269961 | 1.0
Monthly or Less Often | Monthly | C64498 | Every month. | 2011-09-06 | | 2991205 | 1.0
Never | Never | C70543 | Not ever; at no time in the past (or future). | 2011-09-06 | | 3025032 | 1.0
Over a year ago | Over a year ago | | Over a year ago | 2011-09-06 | | 3269960 | 1.0
Weekly | About once a week | C67069 | Every week. | 2011-09-06 | | 2867523 | 1.0
Reference Documents
Document Name | Document Type | Document Text | Context | URL
There are no reference documents for the value domain.
User: Public User
Privacy Notice
Version 4.0.5 Build 09
Please send comments and suggestions to ncicbiit@mail.nih.gov

Overview of Results

- Clinical Text

Source	# Notes	# Sentences	# Annotations*
MTS	491	130	**465**
UVMMC	-	450	**1,205**

* κ of 0.92 and 0.94 for annotation of 10% of MTS sentences and UVMMC comments

- Research Measures

Source	# Measures	# Questions	# Annotations*
PhenX Toolkit	10	191	**2,019**
CDE Browser	40	40	**343**

* κ of 0.95 and 0.94 for annotation of 10% of PhenX Toolkit and CDE Browser questions

#	Element	HL7	openEHR	MTS	UVMMC	PhenX Toolkit	CDE Browser
1	Status	◆	◆	◆	◆		◆
2	Certainty			◆	◆	◆	◆
3	**Negation**		◆	◆	◆	◆	◆
4	Temporal				◆	◆	◆
5	Start Date	◆	◆		◆		
6	Start Age		◆		◆	◆	◆
7	Quit Date	◆	◆		◆		◆
8	Quit Age		◆				
9	*Duration of Use*				◆	◆	◆
10	Duration Since Quit				◆	◆	
11	*Duration Since Time Point*			◆	◆	◆	
12	**Time Frame**			◆	◆	◆	◆
13	*Time Point*			◆	◆	◆	◆
14	**Method**	◆	◆	◆	◆	◆	◆

#	Element	HL7	openEHR	MTS	UVMMC	PhenX Toolkit	CDE Browser
15	**Type**	◆	◆	◆	◆	◆	◆
16	Subtype			◆	◆		
17	**Amount**	◆	◆	◆	◆	◆	◆
18	**Frequency**	◆	◆	◆	◆	◆	◆
19	Context			◆	◆	◆	◆
20	Situation			◆	◆		
21	Location				◆		
22	Subject					◆	
23	Change		◆				
24	Triggers		◆				
25	Evidence of Dependence		◆	◆	◆	◆	◆
26	Cessation		◆	◆	◆	◆	
27	Other		◆	◆	◆	◆	
	Total # of Elements	***7***	***15***	***16***	***23***	***18***	***15***

Distribution of Values for *Type*

Clinical Text *472 Total Values; 109 Unique Values; 34 Groups*	Research Measures *824 Total Values; 408 Unique Values; 203 Groups*
Marijuana: 83 (17.6%) [11] Cannabis Marijana, marijuana, marijuanna, MJ Medical marijuana, prescribed marijuana	**Drugs:** 69 (8.4%) [16] Any other drugs, drug products Drug non-medical use only Drugs not prescribed by a doctor
Caffeine: 77 (16.3%) [10] Caffein, caffeine Coffee, coffees, ice coffee, green tea, tea Cola, pepsi, soda, energy drinks	**Marijuana:** 50 (6.1%) [12] Cannabinoids, cannabis Hash, hash oil, hashish Marijuana, mary jane
Illegal drugs: 62 (13.1%) [10] Illegal drug, illicit substance All illicit, illicit drug, illicits Multiple illicit drugs	**Cocaine:** 47 (5.7%) [10] Angel dust Cocaine, coke, crack Cocaine in chunk form

Distribution of Values for *Frequency*

Clinical Text ***223 Total Values; 81 Unique Values; 7 Groups***	**Research Measures** ***672 Total Values; 52 Unique Values; 7 Groups***
Occasionally: 114 (51.1%) [13] Occasionally, off and on, on occasion Every few months Every once in a while Intermittent	**[#]:** 576 (85.7%) [18] Exact number: e.g., 0, 1, 2 Range: e.g., 1-2, 6-9, 20-39 >/<: e.g., 100 times or more Calculation: e.g., average # times
[#] [time unit] per [time unit]: 102 (45.7%) [54] /day, daily /week, weekly 1-2 times per month, 3x a day Every other day, per day everyday	**[#] [time unit] per [time unit]:** 51 (7.6%) [22] Daily, weekly, monthly 1 or 2 days a week, 2x weekly About once a day, / day < once a month, monthly or less often
Other: 5 (2.1%) [4] Chronic Minimal, not usually Part-time	**Other:** 12 (1.6%) [5] More often than prescribed Most frequently, usually Never

Distribution of Values for *Time Frame*

Clinical Text *90 Total Values; 48 Unique Values; 9 Groups*	Research Measures *182 Total Values; 31 Unique Values; 9 Groups*
Past: 30 (33.3%) [6] Former Past, prior Prev, previous	**Past [#] [time unit]:** 65 (35.7%) [7] During the last 30 days, past 30 days During the last 12 months, past year Past 30 days up to and including today
Current: 18 (20.0%) [5] Current, currently Now, present This time	**[#] [time unit]:** 60 (33.0%) [6] _day(s), _week(s), _month(s), _year(s) A day Number of days_[range:0-30]
[#] [time unit] ago: 9 (10.0%) [9] Ten years ago 2 weeks ago 1 month ago	**Lifetime:** 43 (23.6%) [4] During your life, in your lifetime Any time in her life Any time in their lives

Discussion

- Total of 27 elements represented across biomedical standards, clinical text, and research measures
- Disparate lists of values for elements highlight clinical vs. research purposes of sources
- Range of variant names (e.g., slang or street name) and levels of granularity (e.g., drug groupings) observed for *Type* element
- Next steps include analysis of additional sources and alignment with standardized vocabularies

Conclusion

- Limited efforts focused on standardizing drug use information in comparison with tobacco and alcohol use
- Analysis of biomedical standards, clinical text, and research measures
  - Provide broad perspective on current representation of drug use information
  - Contribute to relevant standards development efforts for supporting integration and interoperability
  - Inform enhanced and standardized collection of drug use information in EHR and other systems

Questions

Thank you!

https://sites.google.com/site/sfhereproject/

This work was supported in part by NIH/NLM grant R01LM011364.

NIH
U.S. National Library of Medicine