

**FEi Baseline Sensitive Data Value Sets Review**



**April 20, 2016**

SAMHSA Health IT

Health Information Technology

Prepared by Serafina Versaggi, Versaggi Consulting

**Table of Contents**

[1 Introduction 2](#_Toc448936722)

[2 Scope 2](#_Toc448936723)

[2.1 Rationale 2](#_Toc448936724)

[3 Findings 3](#_Toc448936725)

[3.1 Approach/artifact 3](#_Toc448936726)

[3.2 Invalid Codes in Baseline Data 4](#_Toc448936727)

[3.2.1 Valid ICD-9-CM Format and Codes 4](#_Toc448936728)

[3.2.2 Valid ICD-10-CM Format and Codes 6](#_Toc448936729)

[3.2.3 LOINC 7](#_Toc448936730)

[3.2.4 SNOMED CT 7](#_Toc448936731)

[3.2.5 RxNorm 8](#_Toc448936732)

[3.2.6 CPT 8](#_Toc448936733)

[3.2.7 HCPCS (Level II codes) 10](#_Toc448936734)

[3.3 Incomplete Coverage 11](#_Toc448936735)

[4 Recommendations 11](#_Toc448936736)

[4.1 Document Approach for Defining and Curating Sensitive Health Data Value Sets 11](#_Toc448936737)

[4.2 Value set naming conventions 11](#_Toc448936738)

[4.3 Ensure Sensitive Data Value Sets are Comprehensive 12](#_Toc448936739)

[4.4 Perform GAP Analysis then Ballot Draft VSAC Value Sets 12](#_Toc448936740)

[4.4.1 Test Consent2Share using selected VSAC harmonization candidate value sets 12](#_Toc448936741)

[4.5 Use Valid Code System Files to Create Value Sets Using Code System Source Files 13](#_Toc448936742)

[5 Appendix 13](#_Toc448936743)

[5.1 ICD-9-CM 13](#_Toc448936744)

[5.2 ICD-10-CM 14](#_Toc448936745)

[5.3 LOINC 14](#_Toc448936746)

[5.4 SNOMED CT 14](#_Toc448936747)

[5.5 CPT / HCPCS 14](#_Toc448936748)

Table of Figures

[Figure 1: ICD-9-CM Code Format 6](#_Toc448936700)

[Figure 2: ICD-10-CM Code Format 7](#_Toc448936701)

[Figure 3: Category II 10](#_Toc448936702)

# Introduction

This report summarizes a visual inspection of FEi’s sensitive data value sets and includes recommendations for enhancing the quality and coverage for these data. The report serves as Deliverable #4 in the statement of work (SOW) prepared under FEi Subcontract # S16V-01.

This task was undertaken following questions I had regarding the original source for the FEi sensitive data (a spreadsheet provided by Ioana Singureanu, enclosed).[[1]](#footnote-1) I was asked to perform a high level review of these data, identify anomalies and make suggestions for improving the completeness and quality of the data.

# Scope

The original scope for this project was limited to sensitive mental health and substance use disorders as represented in the FEi baseline data and to map FEi ICD-10-CM as well as ICD-10-CM concepts in the DSM-5 manual to SNOMED CT codes.

As work progressed, the FEi baseline data (valueset\_conceptCodes\_Final.xlsx downloaded from the FEi FTP site) was used to guide the development of each of Substance Use map and VSAC value set template that comprise deliverable #2, so that all of the relevant codes[[2]](#footnote-2) are included in the resulting maps and value sets.

As the result, while this review is not exhaustive, I became very familiar with the baseline data and can provide the following observations for value sets comprised of LOINC, RxNorm, and CPT/HCPCS code systems in addition to the ICD9 & 10 code findings. A detailed issues list is documented in the [Appendix](#_Appendix).

## Rationale

The scope for this effort was limited to diagnoses in this phase because of time and resource constraints. At the same time, it is important to identify the scope and breadth for these value sets, and it is also not necessary to map every type of clinical information (procedures/tests, medications, etc.) to SNOMED CT.

There are Common Standards and Implementation Specifications for Electronic Exchange of Information as specified by Meaningful Use Stage 2 final rules that define a common dataset for all summary of care records, including structured, coded data to be formatted uniformly and sent securely during transitions of care, upon discharge, and to be shared with the patient themselves. These include:

* Patient name and demographic information including preferred language (ISO 639-2 alpha-3)
* Sex, race/ethnicity (OMB Ethnicity) and date of birth
* Vital signs including height, weight, blood pressure, and smoking status (SNOMED CT)
* Encounter diagnosis (SNOMED CT or ICD-10-CM)
* Procedures (SNOMED CT)
* Medications (RxNorm) and medication allergies (RxNorm)
* Laboratory test results (LOINC)
* Immunizations (CVX)
* Functional status including activities of daily living, cognitive and disability status
* Care plan field including goals and instructions
* Care team including primary care provider of record
* Reason for referral and referring provider’s name and office contact information (for providers)
* Discharge instructions (for hospitals)

The Meaningful Use legislation specifies the relevant code systems associated with data elements specifically included EHR-System certification criteria in each new version of the catalog[[3]](#footnote-3) (e.g., diagnoses as ICD or SCT; diagnostic tests as LOINC; medications (ordered, dispensed and administered) as RxNorm; Procedures and Encounters as either ICD-10-PCS (ICD-9-CM for services rendered prior to Oct. 1, 2015), CPT-4 or HCPCS, etc.). Each new version of the Data Element Catalog defines the relevant code systems.

The Common Data element Catalog could help guide the development (and maintenance) of SAMHSA sensitive data value sets by identifying the vocabulary certified systems are expected to be use when capturing / exchanging data and for automating the quality measures reporting process.

The frequency by which various code systems are also contributes to the strategy for creating, mapping and maintaining these value sets. (E.g., RxNorm is updated weekly; ICD-10-CM yearly for implementation on Oct 1 each year; SNOMED CT twice yearly Jan 1 & July 1.)

# Findings

This section describes problems revealed by visual inspection of data downloaded into an MS Excel spreadsheet from the FEi FTP site.

There are two concerns with the ICD-9-CM and ICD-10-CM value sets. The first is accuracy (of the codes); the second, the breadth of coverage/inclusiveness for the value set, discussed in [section 3.2](#_Incomplete_sensitive_value)below.

## Approach/artifact

When an invalid code was found in the baseline data, I updated format in that cell of the Excel spreadsheet: valueset\_conceptCodes\_AmendedSV.xlsx so that the correct code is displayed. These invalid codes are suspected in part to be due to import issues when the data are downloaded into Excel.

I highlighted cells and entered comments in those cells throughout the spreadsheet describing the discrepancy/issue in case there is interest in the details (Spreadsheet *valueset\_conceptCodes\_AmendedSV.xlsx* enclosed with this report).

## Invalid Codes in Baseline Data

Accuracy of the data is the first concern.

A number of invalid ICD and LOINC codes appear to have been introduced by typos (having been manually entered), or created by Microsoft Excel cell format issues (e.g., missing significant digits, date or scientific notation data formats) when the data are downloaded from the FEi production database and presented in a spreadsheet.

The following sections provide information about the ICD-9-CM and ICD-10-CM code systems to help understand issues related to the data.

### Valid ICD-9-CM Format and Codes

The ICD-9-CM code system consists of three volumes covering diagnosis and procedure codes.

*ICD-9-CM Diagnosis codes* are contained in volumes 1 & 2. Codes are 3, 4, or 5 digits in length.

Three digit codes represent heading categories that may be further subdivided by the use of fourth and/or fifth digits which provide greater detail. On rare occasions, some three character codes are not further subdivided (e.g., 319 Unspecified intellectual disabilities).

Three-digit codes are only found in patient/client data when the category is not further subdivided (otherwise the code is not specific enough for billing or other reporting purposes).

ICD-9-CM codes are defined/stored as left justified strings without a decimal point in most files[[4]](#footnote-4). Valid ICD-9-CM codes have a minimum length of a three characters. A decimal point is inserted following the third character from the left if and only if the code is greater than three characters. For example, 291 is 291; whereas 2910 and 29100 are 291.0 and 291.00 respectively.

Codes in the 001-099 range include significant leading zero(s), meaning zero(s) are always stored and displayed, e.g., 042 Human immunodeficiency virus [HIV] disease.

*ICD-9-CM Procedure codes* (Volume 3) are either 3 or 4 digits in length. Two digit codes are included in Volume 3 of ICD-9-CM as category headings only. These categories contain codes that are further subdivided by the use of a third and/or fourth digit following a decimal point between the second and third digit.

There are no services, encounters or procedures that could be defined using ICD-9-CM (or ICD-10-PCS) procedure codes currently in the FEi baseline sensitive data value sets.

[Figure 1](#Fig1) depicts the format and meaning for each character of an ICD-9-CM code.

#### Finding: Incorrect ICD-9-CM codes

Many invalid ICD-9-CM codes may have introduced when data were downloaded into Excel. (Microsoft Excel may automatically apply a built-in number format to the cell based on characteristics of the number.)[[5]](#footnote-5)

Because some ICD-9-CM code descriptions did not match the code and because some LOINC codes appear as dates, I made an assumption that invalid codes were introduced by Excel. It is critical that only valid codes are included before these value sets are submitted even for draft publication in VSAC.

Each unique ICD9 & ICD10 code has only one description and each ICD9 and ICD10 description has only one code associated with each other.

In FEi baseline data, there are instances of a single code with more than one description[[6]](#footnote-6).

See [Appendix](#_Appendix) for detailed findings related to the FEi baseline ICD-9-CM codes and value sets.

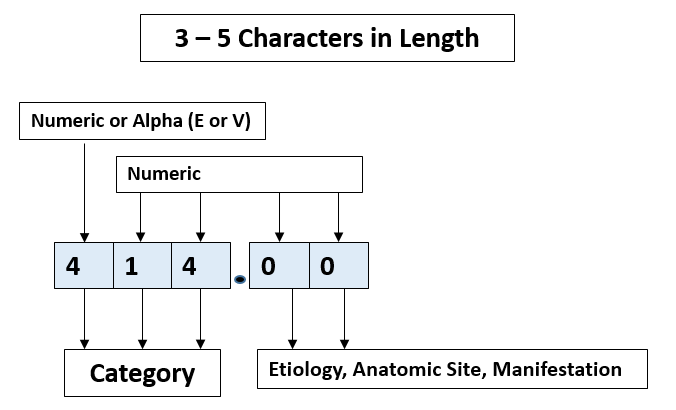


Figure 1: ICD-9-CM Code Format

### Valid ICD-10-CM Format and Codes

While the single ICD-9-CM code system encompassed diagnoses (Volumes 1, 2) and procedures (Vol 3), the U.S. edition of ICD-10 is defined by two code systems – ICD-10-CM correlating to diagnoses and ICD-10-PCS covering procedures.[[7]](#footnote-7)

The ICD-10-CM Tabular List contains categories, subcategories and codes. Codes may be 3, 4, 5, 6 or 7 characters long[[8]](#footnote-8). Each level of subdivision after a category is a subcategory.

Characters for categories, subcategories and codes may be either alpha or numeric. All categories consist of 3 characters. A three-character category that is not further subdivided is equivalent to a code.

Subcategories are comprised of 4 or 5 characters.

The final level of subdivision is a code.

Figure 2 describes the format and meaning behind each character in ICD-10-CM.

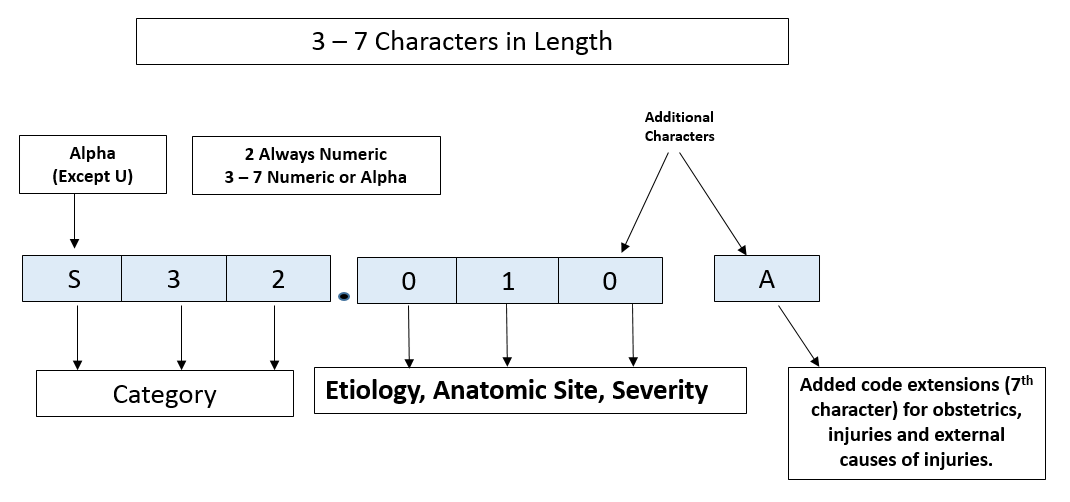


Figure 2: ICD-10-CM Code Format

#### Finding: Invalid ICD-10-CM codes

Some invalid ICD-10-CM codes found in the FEi baseline data may be due to typographical errors if codes were manually entered into the database (the source).

These errors did not appear to be due to Excel cell formatting, and the invalid codes have been highlighted and commented in the amended spreadsheet included with this report.

See [Appendix](#_Appendix) for detailed findings related to the FEi baseline ICD-10-CM codes and value sets.

### LOINC

A new approach for identifying sensitive LOINC codes should be developed.

LOINC codes are used to identify the clinical laboratory tests (orders & some results) and a number of clinical observations captured by certified EHR-Systems as mandated by Meaningful Use. It isn’t sufficient to merely identify a subset of clinical data (e.g., diagnoses and encounter types) associated with sensitive health information categories (e.g., HIV, mental health, substance use, trauma, etc.). More comprehensive coverage could be attained by selecting additional LOINC codes for inclusion in the existing value sets (enabling data segmentation-capable systems to better isolate and protected sensitive health information from unintended (re-)disclosure.

In addition to gaps in coverage, a number of invalid codes were found in the database probably introduced by Excel formatting issues (code represented as date).

### SNOMED CT

Given that the project was to develop SNOMED CT versions of sensitive value sets, it wasn’t surprising to find only a handful of SNOMED CT codes currently in the FEi baseline data.

#### Misspelled SNOMED CT code system name

The SNOMED CT code system name was misspelled in a number of *Value set names* in the FEi database at the time this analysis was undertaken. Since then, SNOWMED was updated to SNOMED[[9]](#footnote-9) in the database.

#### Inconsistent placement of SNOMED CT in value sets

The few SNOMED CT codes that were found were included in value sets containing almost exclusively ICD-9-CM codes and whose value set name included ICD-9-CM and SNOMED ([misspelled](#_Misspelled_SNOMED_CT) initially). Although value sets are not limited to codes derived from a single code system and version (for example in VSAC Grouping[[10]](#footnote-10) value sets), instead SNOMED CT codes should be moved to their SNOMED CT-equivalent value sets (the ones based on the process to create maps between ICD-10-CM and SCT drafted in Deliverables 1 & 2).

#### Codes Selected from the Inappropriate SNOMED CT Hierarchy

One SNOMED CT code was taken from the wrong SNOMED CT hierarchy inaccurately representing the desired concept in that value set.

Presence of codes rooted at a different concept normally indicates an incorrect choice of codes.

### RxNorm

RxNorm codes are used to Coverage in each sensitive category is lacking as there are a number of medications that are used for Alcohol abuse treatment that are missing from the value set named *Alcohol-Abuse-Treatment-RxNorm.*

<https://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html>

### CPT

The Current Procedural Terminology (CPT) code set is maintained by the American Medical Association. It is copyright protected by the AMA.[[11]](#footnote-11) These proprietary codes are used to describe medical, surgical, and diagnostic services for administrative, financial, and analytical purposes.

The FEi baseline file contains a number of CPT and HCPCS codes. It’s my understanding that a license must be purchased for applications to “use” CPT codes in their system, but there are no restrictions for creating value sets containing valid CPT-4 codes.

CPT is comprised of three categories:

* Category I: Procedures that are consistent with contemporary medical practice and are widely performed.
* Category II: Supplementary tracking codes that can be used for performance measures.[[12]](#footnote-12)
* Category III: Temporary codes for emerging technology, services and procedures.

All three categories comprise CPT Level I codes and are different from the HCPCS (Level II) code system.

#### CPT Category I Codes

Category I codes are the proprietary five-digit numeric codes included in the main body of CPT and the section of CPT that most people think of when referring to CPT.

These codes represent procedures that are consistent with contemporary medical practice and are widely performed.

* Procedure or service approved by the Food and Drug Administration (FDA)
* Procedure or service commonly performed by health care professionals nationwide
* Procedure or service's clinical efficacy is proven and documented

Category I codes are updated annually (effective date January 1) and are broken into six sections.

1. Evaluation and Management
2. Anesthesiology
3. Surgery
4. Radiology
5. Pathology and Laboratory
6. Medicine

Category I codes consist 5 numeric digits (no alpha characters).

#### CPT Category II Codes

Category II codes are supplemental tracking codes intended to collect information about the quality of care delivered by encoding a number of clinical components (services and/or test results) that may be typically included in evaluation and management (E&M) services to support performance measures.

The use of category II codes is optional, not required for correct coding and may not be used as a substitute for Category I codes. As the result, it is important that the sensitive information value sets drawn from the CPT code system include most importantly appropriate CPT Category I (AKA CPT Level I) codes in addition to Category II codes that may be associated with those E&M services because these codes are NOT accepted for outpatient billing purposes (by CMS and likely, by any private insurer either) and therefore the likelihood that these codes exist in the local HER-Systems databases is low unless they are coding for internal reporting or performance measurement purposes.

Category II Codes are alphanumeric and consist of four digits followed by the alpha character 'F’. These codes are released annually as part of the general CPT code set and published twice a year, Jan.1 and July 1, on the AMA Web site. The most current listing of CPT II codes can be found on the AMA Web site at <http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt/about-cpt/category-ii-codes.page?>

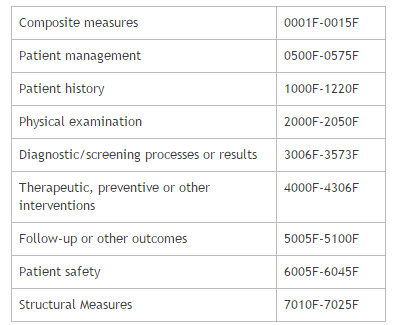


Figure 3: Category II

#### CPT Category III Codes

CPT Category III consist of a set of temporary codes for used emerging technology, services, and procedures.

Category III codes allow data collection for these services/procedures instead of the use of a Category I code such as 99499 -Unlisted evaluation and management service. Use of unlisted codes does not provide the opportunity to collect information that is important in the evaluation of health care delivery and the formation of public and private policy. The use of the Category III codes in EHR-Systems will allow physicians and other qualified health care professionals, insurers, health services researchers, and health policy experts to identify emerging technology, services, and procedures for clinical efficacy, utilization and outcomes.

Category III codes may not conform to the usual requirements for CPT Category I codes established by the Editorial Panel. For Category I codes, the Panel requires that the service/procedure be performed by many health care professionals in clinical practice in multiple locations and that FDA approval, as appropriate, has already been received. The nature of emerging technology, services, and procedures is such that these requirements may not be met. For these reasons, temporary codes for emerging technology, services, and procedures have been placed in a separate section of the CPT code set and the codes are differentiated from Category I CPT codes by the use of alphanumeric characters.

CPT Category III codes have an alpha character as the 5th character in the string (i.e., four digits followed by the letter T).

Codes in this section may or may not eventually receive a Category I CPT code. In either case, in general, a given Category III code will be archived five years from the date of initial publication or extension unless a modification of the archival date is specifically noted at the time of a revision or change to a code (e.g., addition of parenthetical instructions, reinstatement).

Services and procedures described by Category III codes which have been archived after five years, without conversion, must be reported using the Category I unlisted code unless another specific cross-reference is established at the time of archiving.

New codes or revised codes in this section are released semi-annually via the AMA CPT Internet site, to expedite dissemination for reporting. The full set of temporary codes for emerging technology, services, and procedures are published annually in the CPT code set. Go to www.ama-assn.org/go/cpt for the most current listing.

### HCPCS (Level II codes)

The [Healthcare Common Procedure Coding System (HCPCS)](https://www.cms.gov/Medicare/Coding/HCPCSReleaseCodeSets/Alpha-Numeric-HCPCS.html) is an “open-source” subset of health care procedure codes based on CPT. The Healthcare Common Procedure Coding System (HCPCS) is a collection of codes that represent procedures, supplies, products and services which may be provided to Medicare beneficiaries and to individuals enrolled in private health insurance programs. The codes are divided into two levels[[13]](#footnote-13) described below.

* **Level I** consists of the CPT-4 code system. Codes and descriptors copyrighted by the American Medical Association's current procedural terminology, fourth edition (CPT-4). These are 5 digit (numeric) codes representing physician and non-physician services.
* **Level II** codes are subset into the HCPCS code system. These codes are used primarily to identify products, supplies, and services not included in the CPT, such as ambulance services and durable medical equipment, prosthetics, orthotics, and supplies (DMEPOS) when used outside a physician's office. HCPCS codes are also referred to as alpha-numeric codes because they are five position codes consisting of a single alphabetical letter followed by 4 numeric digits, while CPT codes are identified using 5 numeric digits or 4 digits followed by an alpha character.[[14]](#footnote-14)

## Incomplete Coverage

The FEi baseline value sets contain only a subset of the relevant codes available in each code system necessary to provide protection from the unintended disclosure of protected health data during an exchange by systems capable of applying security tags to comply with data segmentation for privacy.

To provide comprehensive coverage, all relevant codes for sensitive category descending from a particular hierarchy within a given code system should be included in each value set. For example, Value Set Name Category *Alcohol use and Alcoholism Information* has incomplete and/or missing value sets that could be drawn from code systems such as LOINC, RxNorm, CPT, HCPCS and ICD-9-CM procedures (or ICD-10-PCS) and represented in FEi baseline data.

Each value set definition should include its purpose of use (fit), along with the inclusion and exclusion criteria that describe the content / codes included in the value set.

See the [Recommendation](#_Ensure_Sensitive_Data) section for details.

# Recommendations

## Document Approach for Defining and Curating Sensitive Health Data Value Sets

Document the requirements and use case for each value set and sensitive data category identified by SAMHSA. This is an important step to refine the process for curating and maintaining sensitive data value sets in the VSAC. The requirements will be input to the value set definition: i.e., value set purpose, code system source; inclusion and exclusion rules, and will guide the content development and subsequent use for these value sets once in the public domain.

This recommendation is based on an assumption that FEi’s terminology services for Consent2Share will be able to use the VSAC Value Sets, and that once published in VSAC, EHR-Systems will be able to use these value sets to infer information that should be redacted by a rules engine based on privacy policies and patients’ consent directives whenever protected health information is exchanged between EHR-Systems and/or sent to a Health Information Exchange (HIE). See [section 4.3.1](#_Test_Consent2Share_using).

## Value set naming conventions

Value set name and value set category naming conventions should reflect the content and purpose for the value set as best as possible. Some of the FEi value set names (and categories) can be misleading and/or are inconsistent.

Value set names should reflect their purpose and convey the distinguishing characteristics of member concepts (disorders versus procedures; medications versus diagnostic tests, etc.) and it’s OK to append the code system name from which concept codes are drawn.

While value sets are not necessarily limited to codes selected from a single code system (for example, in the case of Grouping value sets), most value sets in the VSAC other than Grouping are drawn from a single code system.

## Ensure Sensitive Data Value Sets are Comprehensive

Currently, FEi baseline data value sets do not provide the coverage intended for that sensitive category or concept, since only a subset of are relevant (sensitive) codes from a given code system are currently included in the baseline data value sets.

For example, only a subset of codes are currently present in FEi value set name Alcohol-Abuse-ICD-10 when the value set should include all concepts[[15]](#footnote-15) from the ICD-10-CM chapter F10 (Alcohol related disorders) and possibly others, to trap any instance of alcohol-related sensitive condition that could appear in client/patient data (should they request those data not be shared).

## Perform GAP Analysis then Ballot Draft VSAC Value Sets

Another way to ensure completeness is to search the VSAC for similarly-named value sets and identifying candidates for harmonization.

Value sets currently published in VSAC by other stewards have similar ‘sensitive’ value set names and contain similar concepts which we consider sensitive (e.g., substance use disorders, BH Mood Stabilizer Medication RxNorm, etc.). These value sets have been defined for different purposes, such as quality measures, but they are also a potential source for additional concepts that may have been missed when initially composing the SAMHSA sensitive data value sets.

Going on the assumption that other stewards’ value sets include codes likely to be in use by certified EHR-Systems, performing an inventory and gap analysis against the draft SAMHSA/FEi sensitive value sets can help ensure codes are not omitted by oversight.

Once the draft value sets have been defined (i.e., Deliverable 1 & 2 harmonized with the above), the ballot process within a standards organization such as HL7 provides a mechanism for socializing these value sets, as well as gaining expertise and buy-in from subject matter experts who will improve the quality of these drafts through the review process.

The ballot process is an important means for educating a wider stake holder audience about the meaning and worth of these value sets, thereby hopefully gaining greater acceptance and uptake by organizations who wish to enable electronic data segmentation for privacy (DS4P). These value sets contain the critical (and missing) vocabulary required for security tagging based on client/patient consent and applicable jurisdictional policies when exchanging data electronically between disparate EHR-Systems.

### Test Consent2Share using selected VSAC harmonization candidate value sets

I recognize that FEi needs access to the newly defined value sets for use in C2S as soon as possible.

Given my recommendation to ballot the draft value sets prior to publication in VSAC, I suggest using some of the similarly defined value sets currently published in VSAC for testing purposes. Once the SAMHSA sensitive value sets have been approved for publication, they will function exactly as the value sets used for testing because the structure of the value sets are identical (except the outcome will be different based on the content obviously).

## Use Valid Code System Files to Create Value Sets Using Code System Source Files

Once the content development approach and strategy for curating and maintaining these value sets in VSAC has been defined and FEi has tested and can use the VSAC value sets internally in the application(s), codes should always be selected directly from source files (e.g., [the CMS website for ICD codes](https://www.cms.gov/Medicare/Coding/ICD10/2016-ICD-10-CM-and-GEMs.html), [Regenstrief website for LOINC codes](https://loinc.org/downloads/files/loinc-snomed-ct-expression-associations-technology-preview/gotoCopyrightedFile), [RxNorm from NLM](https://www.nlm.nih.gov/research/umls/rxnorm/docs/rxnormfiles.html), etc.) instead of using indirect sources to cut & paste from various website tools to browse code system content.

By using the code systems’ file definitions (e.g., 1:1 code/description; code format) invalid codes can be eliminated from the FEi baseline data.

The Excel mapping worksheets (Deliverables 1 & 2) will be used to prepare the draft value sets suggested for publication in VSAC. A copy of an example covering *Alcohol disorders* has been included with this deliverable as an example reference.

The VSAC Authoring Tool provides a descriptor match check as a built in function. The load process will be streamlined if the original mapping spreadsheets contain only valid code/description combinations because the VSAC Authoring Tool will perform this validation during batch import of codes from the spreadsheet.

# Appendix

The following list of issues were found in the sensitive FEi baseline value sets. Findings have been classified by code system.

## ICD-9-CM

1. Due to the invalid format of many ICD-9-CM codes in the spreadsheet, many codes are associated with the wrong description (e.g., value set Psychiatric-Disorders-ICD-9-SNOWMED contains code 309 Adjustment Disorder with Depressed Mood. The correct code for that description is 309.0 Code 309 description is the category Adjustment Reaction). Other times, the same unique code appears multiple times associated with different descriptions, which is invalid for ICD.
2. ICD-9-CM 305.3 with two different descriptions[[16]](#footnote-16). The first, 305.3 Nondependent hallucinogen abuse, the second, 305.3 Hallucinogen abuse, unspecified. The code matching description Hallucinogen abuse, unspecified is 305.30 which led me to believe Excel created the anomaly (and hopefully both unique codes exist in the FEi Vaseline data (305.3 as well as the more appropriate 305.30). 305.31, 305.32 & 305.33 should be included in this value set as well.
3. 042, 078.1, 094.2, 097.9, 098.2, 099.41 are missing the leading zero in the FEi data as displayed in Excel.

## ICD-10-CM

F06.01 Catatonic disorder due to known physiological condition should be F06.1 (there is no F06.01). This appears to be due to a typo as Excel formatting could not have introduce this invalid code.

## LOINC

1. Numerous LOINC codes and/or LOINC code parts[[17]](#footnote-17) are represented as Date or another format resulting in invalid LOINC codes.
2. What to do with LOINC code parts included in the file. Since the file does not organize the LOINC parts with the parent code, it wasn’t clear whether how LOINC part codes are intended to be used by applications applying data segmentation operations.

## SNOMED CT

* 1. Although only six SNOMED CT codes were found in the baseline data (based on being associated with the SNOMED CT code systems), the SCT codes were included in value sets containing other code system concepts. In one case, a SNOMED CT disorder (diagnosis code) was included in a Medications value Set (RxNorm codes for HIV medications).
  2. SNOMED CT code SCTID: 417877006 Cocaine – pharmaceutical (substance) taken from the Substance hierarchy and included in the Cocaine disorders value set (Cocaine-Abuse-ICD-9-SNOWMED) instead of from the appropriate SNOMED CT hierarchy (Clinical Findings) for Cocaine related disorders (e.g., SCTID: 78267003 Cocaine abuse (disorder)).

## CPT / HCPCS[[18]](#footnote-18)

The CPT and HCPCS codes need review. Issues with codes in the baseline data fall into the same general categories.

Some codes have been incorrectly categorized as CPT-4 (by the code system name) in the baseline data when they are HCPCS codes[[19]](#footnote-19).

Many sensitive CPT and HCPCS codes are missing from the current value sets as well, a common issue across all value sets that is discussed in more detail in the following section.

#### Incomplete coverage

* + - 1. A number of relevant CPT and HCPCS codes are missing from the sensitive value sets.
      2. There is no value set assigned to the ALC category containing relevant CPT or HCPCS codes.

#### Inconsistent assignment and/or value set naming convention

Some codes have been assigned to the wrong code system and wrong value set category:

1. HCPCS code 4320F Patient counseled regarding psychosocial AND pharmacologic treatment options for alcohol dependence (SUD) is assigned to category ETH (Substance use information sensitivity) instead of ALC (Alcohol use and Alcoholism Information).

1. The format of the data within Ioana’s spreadsheet was not suitable for analysis. After spending time reworking the data to create one row per mapped code, Ken provided access to the FEi FTP site. [↑](#footnote-ref-1)
2. If an invalid code was found in FEi value set(s), that code would not be included in the new file, but the correct version of the concept would be added in its place. For instance, in the *Cocaine-Abuse-ICD-9-SNOWMED* value set, the SNOMED CT code for *substance* Cocaine was included in lieu of a code from the correct SNOMED CT concept hierarchy (Clinical Findings under which disorders are classified). The code for the substance Cocaine would not be found in a diagnosis field in a real client/patient database. [↑](#footnote-ref-2)
3. See The Data Element Catalog which identifies data element names required for capture in Electronic Health Record (EHR) technology certified under the 2014 Edition of the Office of the National Coordinator for Health Information Technology (ONC) Standards and Certification Criteria. <https://www.nlm.nih.gov/healthit/dec/> [↑](#footnote-ref-3)
4. All ICD-9-CM diagnosis codes (other than V codes) have a maximum length of five characters with the decimal point implied between the third and fourth digit when the length is greater than three characters. For V codes, the decimal is implied between the second and third digit. V codes are Diagnosis Codes whose first position is a ‘V’ and are used as to encode diagnoses falling into the Supplementary Classification of Factors Influencing Health Status and Contact with Health Services. [↑](#footnote-ref-4)
5. Microsoft Excel automatically applies a built-in number format to a cell, based on the following criteria:

   * *If a number contains a slash mark (/) or hyphen (-), it may be converted to a date format. (LOINC)*
   * If a number contains a colon (:), or is followed by a space and the letter A or P, it may be converted to a time format.
   * If a number contains the letter E (in uppercase or lowercase letters; for example, 10e5), or the number contains more characters than can be displayed based on the column width and font, the number may be converted to scientific notation, or exponential, format.

   *If a number contains leading zeros, the leading zeros are dropped (IC-9-CM)* [↑](#footnote-ref-5)
6. Even though the DSM-5 may have multiple descriptions for a single code, and the same code may fall into multiple categories within the DSM-5, an ICD code has a 1:1 code : long description relationship. DSM-5 is not a coding system, but a methodology for guiding clinicians to a particular diagnosis. The DSM-5 manual is comprised of 20 chapters classifying disorders based on their similarities to one another. [↑](#footnote-ref-6)
7. ICD-10-PCS procedure codes are used only to encode acute care (inpatient) hospital procedures (they are input into the DRG derived for that stay and used for reimbursement and statistical reporting purposes). FEi baseline sensitive data value sets do not currently include any sensitive procedures represented as ICD-10-PCS. [↑](#footnote-ref-7)
8. ICD-10-CM and ICD-10-PCS codes are also formatted as left justified strings in their respective source files. But compared to ICD-9-CM, ICD-10 diagnoses codes (CM) and procedure codes (PCS) have different minimum and maximum lengths (3 – 7 characters for diagnoses; 7 characters for procedures). [↑](#footnote-ref-8)
9. Technically if the intent is to represent the code system name in the value set name, it should be SNOMED CT, not SNOMED.

   *SNOMED* is an acronym for the Systematized Nomenclature of Medicine (a clinical terminology) originally developed by the College of American Pathologists and now owned and maintained by the International Health Terminology Standards Development Organisation (IHTSDO).

   SNOMED Clinical Terms (SNOMED CT) is the most recent version and was preceded by SNOMED RT and SNOMED International. SNOMED CT is considered to be the most comprehensive, multilingual healthcare terminology in the world, created as a result of the merger of SNOMED RT and NHS Clinical Terms Version 3.

   There are also Extensions as well as Editions of SNOMED CT. The extensions comprise a set of terminology components and derivatives that add to and are dependent on the SNOMED CT International Edition (e.g., U.S. edition). [↑](#footnote-ref-9)
10. A Grouping value set is a specialized type of value set containing a list of value sets sharing a common purpose and containing similar clinical concepts. In this case, the VSAC Grouping value sets contain a list of the OIDS associated with the value sets whose concepts codes appear in the Grouping value set (the union of ICD-9-CM, ICD-10-CM, and SNOMED CT). [↑](#footnote-ref-10)
11. CPT-4 codes including both long and short descriptions shall be used in accordance with the CMS/AMA agreement. Any other use violates the AMA copyright. [↑](#footnote-ref-11)
12. See Appendix for more details related to the CPT and HCPCS code systems. [↑](#footnote-ref-12)
13. Originally, the HCPCS was comprised of three levels. Level III codes, also called local codes, were developed by state Medicaid agencies, Medicare contractors, and private insurers for use in specific programs and jurisdictions. The use of Level III codes was discontinued on December 31, 2003. [↑](#footnote-ref-13)
14. Codes and descriptors copyrighted by the American Dental Association's current dental terminology, seventh edition (CDT-2011/12). These are 5 position alpha-numeric codes comprising the d series. All other level II codes and descriptors are approved and maintained jointly by the alpha-numeric editorial panel (consisting of CMS, the Health Insurance Association of America, and the Blue Cross and Blue Shield Association). [↑](#footnote-ref-14)
15. Deemed germane upon review [↑](#footnote-ref-15)
16. It appears the FEi baseline database has changed since I first downloaded and analyzed the data. Since then, a newly downloaded version of the file only has one record for 305.3. Nonetheless, this is still a problem because 305.3 is the category code and unless there is a rule’s engine that redacts all codes descending from 305.3 through 305.4, sensitive data will be missed. 305.3 Nondependent hallucinogen abuse is a subcategory that is not specific for use in patient charts. But there are a number of valid codes descending from 305.3 that are valid and which should be included in the sensitive data value set: 305.30 Hallucinogen abuse, unspecified (and others, through 305.33). [↑](#footnote-ref-16)
17. **A fully specified name in LOINC (associated with the LOINC code) includes the following attributes:** <component/analyte>:<kind of property (substance concentration, mass, volume)>:<time aspect of the measurement>:<type of sample (blood, urine, serum, etc.)>:<scale of measurement, qualitative versus quantitative, etc.>:<method if relevant, e.g., RIA, Immune Blot, etc.>. [↑](#footnote-ref-17)
18. Information included in this section has been gleaned from various sites including [*Advance Healthcare Network for Health Information Professionals*](http://health-information.advanceweb.com/Article/Understand-the-Three-CPT-Code-Categories-2.aspx) [↑](#footnote-ref-18)
19. These unique code systems are assigned unique OIDS: HCPCS Level II Alphanumeric Code System OID: 2.16.840.1.113883.6.285 & CPT Code System OID: 2.16.840.1.113883.6.12. [↑](#footnote-ref-19)