Syndrome Definition Development Process

NSSP CoP Syndrome Definition Subcommittee

Table of Contents

Purpose

The purpose of this document is to standardize and formalize the process for adding new and updating existing syndrome definitions (CCDD Categories) in the NSSP-ESSENCE instance of the BioSense platform. While jurisdictions that maintain their own local versions of ESSENCE do not need to follow this process, this can still serve as a guidance document for best practices when developing new or updating existing definitions.

Process Overview

Below are the four broad steps that are involved in developing and adding new or updated definitions to the NSSP-ESSENCE platform. Steps 1-2 are helpful steps for evaluating and developing definitions even if you do not intend to add them to ESSENCE.

1. Research and Requirements Gathering
2. Iterative Testing and Validation
3. External Validation
4. CDC Review and Addition to ESSENCE

Existing Resources

An overview of the existing resources for developing, testing, and validating syndrome definitions:

- Syndrome Definition Guidance Document
    - Focuses on several key areas related to syndrome definition creation, including the basics behind a syndrome definition, steps to build a syndrome, evaluation of a new (or old) definition, and dissemination.
- Syndrome Definition Evaluation Toolkit
    - R code to evaluate 1-3 definitions for use case applicability, determining differences, manually reviewing line level results, etc.
    - Companion tool to the “Refining the syndrome” section of the Syndrome Definition Guidance Document, found on page 8-9.
- NSSP-ESSENCE Resources
    - Free-text Coding in NSSP-ESSENCE
    - How to Use RStudio with NSSP-ESSENCE APIs
    - CDC-Developed Rnssp R Package
- NSSP CoP Knowledge Repository
    - Syndrome Definition Subcommittee call recordings
    - Syndrome Definition Fact Sheets
- NSSP CoP Meetings
    - Monthly meetings for a broad range of topics related to syndromic surveillance in the US, including the Syndrome Definition Subcommittee.
    - Link to join or update:

https://app.smartsheet.com/b/form/f69e3a17f4a044f68b3bd350e52f9e57

- NSSP CoP Slack Channel
    - Space for federal, state, local, and tribal public health users to collaborate on syndromic surveillance topics.
    - #syndrome-definitions channel
    - Link to join: https://cste.co1.qualtrics.com/jfe/form/SV_721Ioa5BmlANKS1
- NSSP CoP Definition Development Tracker
    - Add new ideas and view status of definitions already in the development/validation process.
    - Access using the link found in the #syndrome-definitions Slack channel or by emailing syndromic@cste.org for access link.

Worksheets and Templates for Development Stages

With each stage of the Definition Development Process, there is an accompanying template to help guide you through that stage. The fields in each template are meant to be carried through to the next stage, so think of each template as the draft for the next stage's template. The Jurisdiction Validation worksheet must be completed to move forward with recruiting additional jurisdictions to formally validate the definition. The Fact Sheet and Technical Brief is required to submit to CDC for review and must be approved by all authors prior to the definition being added to ESSENCE. Although the other templates are not required, they are strongly recommended as a tool for development, even if you do not plan to have your definition added to ESSENCE.

Style Guide

Dictionary

- Definition: The inclusion and exclusion criteria (primarily ESSENCE regex-based) used to classify visits as [condition]-like or suspect [condition]
- Query: The set of parameters used to limit results based on the driving questions, such as age, data source, etc.
- CCDD Category: Definitions that have been added to NSSP-ESSENCE that automatically categorize visits during the BioSense platform processing stage

Historical Context for Field and Definition Names

CCDD Category

ESSENCE started with categories or "Syndromes" which looked at the Chief Complaint to tag visits automatically as data streams into the platform. After some growth of syndromic surveillance, there was a desire to categorize visits not only based on the Chief Complaint, but also the Discharge Diagnosis codes the visit received. At the time, the best way to accomplish this was to combine the "Chief Complaint Parsed" field with the "Discharge Diagnosis" field into a new field called, "Chief Complaint and Discharge Diagnosis" which is often shortened to "CCDD". After this change, there was a need to implement some automation to the task, similar to the syndromes being labeled as data flows to the platform. Since these "new" categories were all running in the CCDD field, they were called "CCDD Categories".

Since then, ESSENCE has widely expanded in query functionality with additional fields as well as unique ways to leverage combinations of fields beyond just the "CCDD". Even though the fields queried are often much more than this single "CCDD" field, it is still the most commonly leveraged field for new definitions, and we have retained the legacy name of "CCDD Categories" when talking about CDC- and community-developed definitions that is used to classify visits automatically.

Field Names in Definition Names

One of the key expansions in query functionality in ESSENCE was the ability to search for free text terms or codes using additional fields beyond just CCDD. As these additional fields have been incorporated into new versions of definitions, these field names have been incorporated to help distinguish significant changes from previous versions. One example of this is the CDC Fentanyl Overdose v2 Parsed definition which is applied to the new CCDD Parsed field.

Use of "CDC" in Definition Names

The addition of "CDC" at the beginning of some definition names began as CDC subject matter experts for the condition of interest were involved in the development of the definition. This practice will be discontinued moving forward in favor of listing specific grants and/or programs that are sponsoring the work to develop a new definition. This is intended to help clarify what definitions are used for specific projects.

Naming Conventions

These naming conventions are in effect as of July 2024 to standardize definitions names while keeping them succinct and informative.

Definition Name Structure:

Definition Name Component	Requirement	Description	Examples
Grant or Project Affiliation	Optional	Abbreviated version of the grant or project that is affiliated with the development and use of the definition	OD2A AVERT
Condition of Interest	Required	A concise description of the condition	Unintentional Firearm Injury
Scope	Optional	To be applied in instances where there are multiple versions where one is more broad (emphasis on capturing as many True Positive visits as possible) and one is more narrow (emphasis on capturing as few False Positive visits as possible)	Broad Narrow
Diagnosis Code-Based	Optional	Applied to indicate whether the definition includes only diagnosis codes	DD
Field Specification	Optional	To be applied if there are multiple versions of the same definition and use case where the only difference is which free text fields are leveraged to return results. This segment should indicate what fields are used (assuming the default is CCDDparsed). If a previous version of a definition was applied to CCDD (or DD) but the new version is applied to CCDD Parsed (or DD Parsed), the newest version should include "Parsed".	Triage Notes Clinical Impression Admit Reason
Age Restriction	Optional	Only to be applied when an age restriction is built into the definition. In cases where an age restriction is recommended but not built-in, please include the recommended age rage in the definition description and use language such as "Recommended age range: [ages]".	Limited to <18 Limited to 65+

Version	Required	Version number of the definition. If it is a new definition, must be v1. Whole numbers only.	v1 v2

Examples:

- AVERT Unintentional Firearm Injury (Triage Notes) v1
- Long COVID Broad v1
- Long COVID Broad (Clinical Impression) v1
- Long COVID Broad (CCDDParsed Only) v1
- OD2A Methamphetamine DD v2

Definition Process

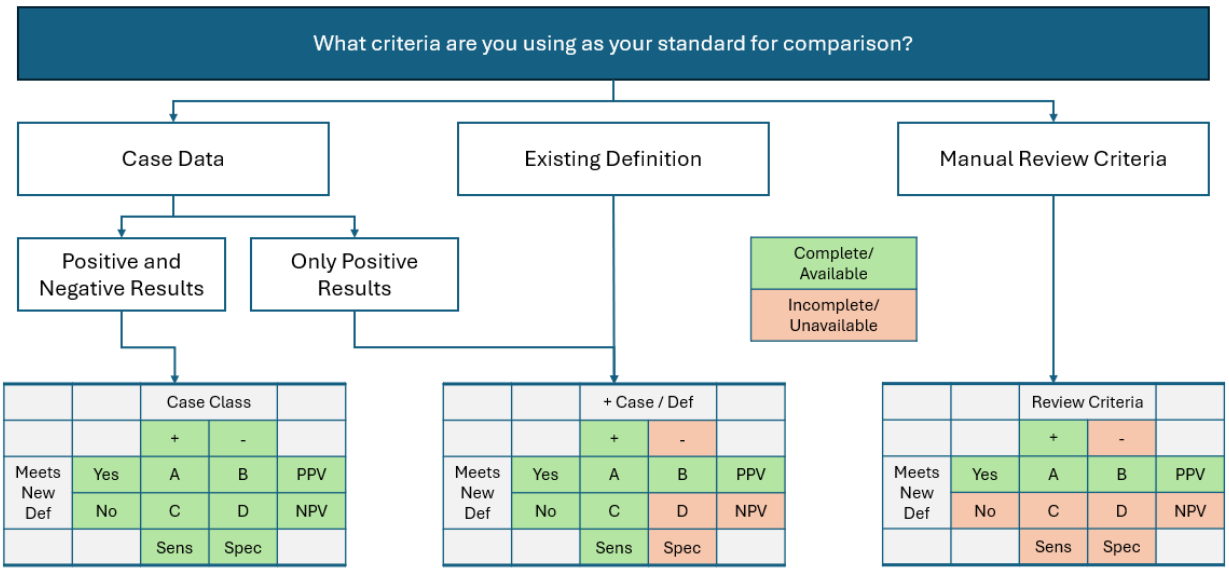
Research and Requirements Gathering

Optional Document: Research and Requirements Gathering Worksheet

This worksheet is intended to help you narrow the scope of your definition as much as possible as you are doing background research and gathering necessary information and requirements for the definition. Although **this worksheet is optional**, it is a useful set of questions that are a recommended place to start.

- Identify the Condition of Interest.
- Create a 2-3 sentence description of what the proposed definition will capture in ESSENCE data.
    - This description may change as you go through the process of development and validation.
- Questions that will help you describe the justification for the proposed definition:
    - What definitions exist that may include concepts related to the condition of interest?
    - What codes, terms, or concepts are **missing** from existing definitions that you feel would justify either a new definition or a new version of an existing definition?
    - What codes, terms or concepts are **included** in existing definitions that result in too many false positives or return a set of results that does not align with your specific needs?
    - Is this definition meant to identify visits where the condition of interest is the reason for visit, or where the condition of interest is identified in visit information but may not be the primary reason for visit?
    - Is there a grant, program, or project that is motivating the development of this definition? If so, please describe.
    - What other information about circumstances, questions, requests, etc. informs the development of this definition?
- Questions about anticipated results and resources to determine use case priorities:
    - Based on what you know about how frequently cases or instances of this condition occur from other data sources, how would you categorize the expected volume of results in a given period of time (e.g., week, month, etc.)? Low, medium, or high?
    - Are there seasonal patterns you want to be able to observe? If so, describe them.
    - Do you have a specific application for the results in mind?
    - What resources do you have available to follow up with or conduct further analysis of the results?
    - How much historical data do you anticipate wanting or needing to look through?
    - Set priority levels for each potential use case activity using the descriptions below to help guide your decision:
        - Case Finding: The volume of visits is low enough and the definition parameters are specific enough to be used for identifying visits that warrant additional follow-up to determine if the patient meets the criteria for a confirmed case. Please note: Visit results alone are not sufficient to determine that the patient meets the case definition.

        - Trend Monitoring: The volume of visits is high enough and enough visits have occurred over time for trend classification models to be able to determine either directionality and/or baselines and confidence intervals.
        - Early Outbreak Detection: The volume of visits may be low, but detection algorithms could identify statistically significant short-term increases that may suggest an outbreak. Further spatiotemporal analysis and review will often be required to confirm an outbreak.
        - Emergent Condition: Visit volumes are low because this is a relatively new, and potentially novel, condition of interest. At a minimum, results can inform situational awareness to determine if further resources, review, and/or analysis need to be conducted for a timely and appropriate public health response.
- Questions about what resources you will have available to validate the new definition, including providing validation result metrics:
    - What is your top priority when thinking about the scope of results you want to capture using this definition? Do you want to emphasize casting a broad net to capture as many true positives as possible knowing you will capture more false positives? Do you want to emphasize a narrower scope where you are minimizing the number of false positives at the risk of missing some true positives? Are you aiming for a reasonable middle ground between the two?
    - What resources will you be using to validate your definition?
        - Existing definition(s)
        - Manual review criteria
        - Reported cases (positives only) that will be linked to visit data (rarely utilized)
        - Reported cases (positives and negatives) that will be linked to visit data (rarely utilized)
    - Use this validation metric differential diagram to determine what validation metrics will be available to you based on what you are using to compare your results to:



- $PPV = \frac{A}{A+B}$ → Always available and should always be considered/utilized

    - $Sensitivity = \frac{A}{A+C}$ → Only available when comparing to existing definition(s) or positive cases
    - $Specificity = \frac{D}{B+D}$ → Only available when comparing to positive and negative cases
    - $NPV = \frac{D}{C+D}$ → Only available when comparing to positive and negative cases
- Gather code and term requirements to match justification and intended use case.
    - It is recommended to consult with clinical subject matter experts (SMEs) on specific codes and terms that are relevant to the condition of interest.
    - Consider if codes and terms need specific inclusion (meaning combinations of codes and terms are required) or exclusions (meaning certain codes or terms, including misspellings or things like "history of [condition]" cannot be present) criteria.
    - During this stage, this can be a general list of codes and terms you think will meaningfully contribute to this definition. As you work through this process, you may choose to add inclusion or exclusion criteria, only include more specific codes, or add spelling variations. The level of detail should be aligned with your intended use case (refer to the previous sections for guiding questions and considerations).

Iterative Testing and Validation

Optional Document: Iterative Testing and Validation Template

This worksheet is intended to help you summarize your research and requirements gathering information into what will eventually become the Jurisdiction Validation Worksheet that will be used by other jurisdictions to help with validation efforts. Although **this worksheet is optional**, it is recommended to use this to keep track of changes or decisions made about what to include or exclude in a definition.

- Summarize information collected for the definition justification
- Set your criteria for classifying visits as True Positive/Undetermined/False Positive for the PPV calculation.
- Start with an initial version based on the code and term requirements you gathered.
- Make iterative improvements to inclusions and exclusion criteria as you evaluate line-level results.
- Calculate preliminary validation metrics to determine if your definition is performing as expected.
- Review definitions internally, consulting with SMEs who may have suggestions on what to prioritize.

Jurisdiction Validation

Required Document: Jurisdiction Validation Template

Validators should provide feedback on the readability and usability of this document as well as the performance of the definition as it will eventually provide important information for end users who will apply this definition for public health practice.

- Revise/update your criteria for classifying visits as True Positive/Undetermined/False Positive for the PPV calculation based on your own iterative review.

- The Definition Proposal Form is intended to be a guide for other jurisdictions who are participating in the validation process. Sections of this document will become sections of the final Syndrome Definition Fact Sheet. Validators should provide feedback on the readability and usability of this document as it will eventually provide important information for end users who will apply this definition for public health practice.

Validation by 2 or more STLT participants is required, except in emergency scenarios designated by NSSP leadership. Jurisdictions may be recruited through any of the following methods:

- Request time on a monthly Syndrome Definition Subcommittee call to give a 2–3-minute elevator pitch overview and request for validators.
    - Slides optional
    - Email syndromic@cste.org for contact information for current subcommittee co-chairs
- Direct communication with jurisdictions who may have already expressed interest or be involved in grant or programs with required definition development participation.
- Request recommendations from NSSP partners who may have awareness of jurisdictional availability and interest.

CDC Review and Addition to ESSENCE

Required Document: Definition Fact Sheet and Technical Brief

The Definition Technical Brief is intended to provide critical background information for end users who will apply this definition for public health practice about the purpose, intended use, performance, and considerations of the proposed definition. Sections of this document will come from sections of the Definition Proposal Form. While CDC partners will review this document during the CDC Review and Addition to ESSENCE stage and may provide feedback or suggestions, they are not necessarily considered "approvers" for definitions that were led by STLT partners.

Once all participating validators have reached a consensus, the lead developer will connect with NSSP partners to move forward with final reviews before being added to the NSSP-ESSENCE platform. Email nssp@cdc.gov and include the final draft of the Definition Fact Sheet and Technical Brief.

Outlined below are the steps taken by NSSP and their approximate timelines:

- Technical Brief review (3-6 weeks)
    - Technical briefs for definitions developed by non-CDC jurisdictions will not need to go through CDC clearance, however CDC will still act as a reviewer to ensure the content aligns with existing technical briefs and that all requirements have been met.
- Syntax review (3-6 weeks)
    - Consult with CDC SMEs
    - Check for syntax mistakes (i.e., missing carets or commas)
    - Run final version of query in CCQV
    - Consult with definition POC on any recommended changes or clarifying questions
    - Generate and share text mining report

- Prepare materials for JHU APL to incorporate into NSSP-ESSENCE

Note: The Definition Technical Brief must be ready for publication (including CDC clearance and/or individual jurisdiction clearance where applicable) before the definition can be added to NSSP-ESSENCE.