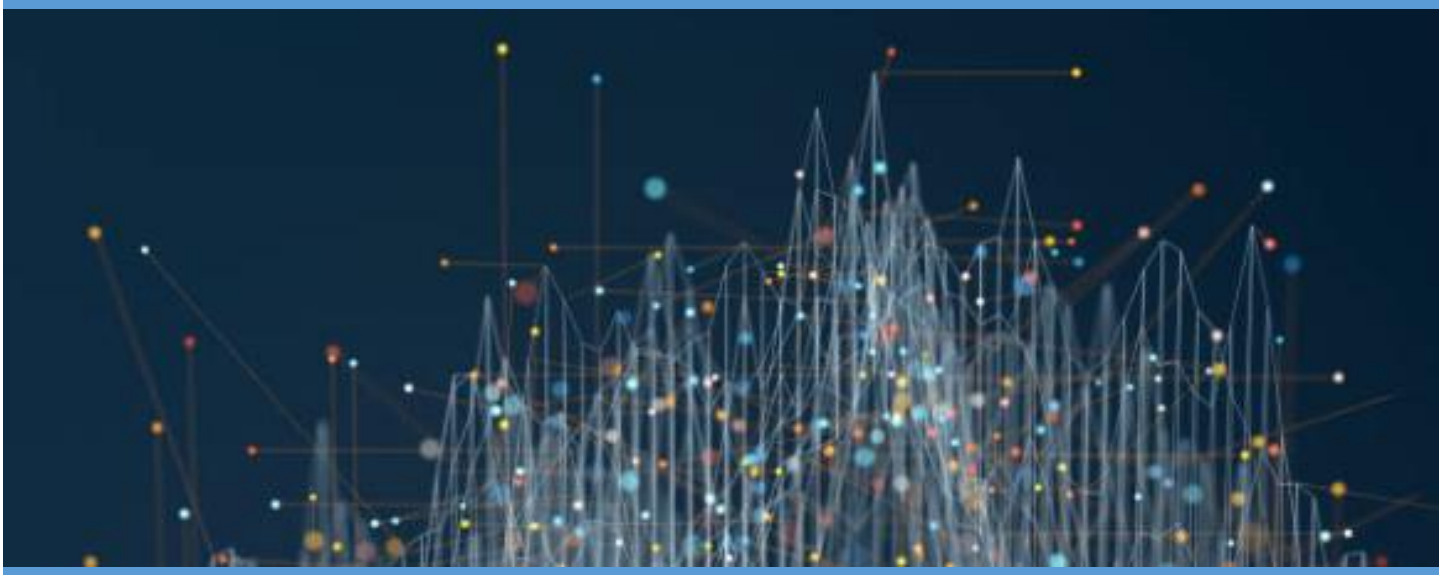
DEVELOPING, EVALUATING, AND DISSEMINATING DEFINITIONS FOR SYNDROMIC SURVEILLANCE IN PUBLIC HEALTH PRACTICE: A GUIDANCE DOCUMENT



JANUARY 2019
VERSION 1

Contents

Acknowledgements

This document was a result of the work completed by the International Society of Disease Surveillance Syndrome Definition Committee.

Authors

Rasneet S. Kumar
Office of Epidemiology
Maricopa County Department of Public Health

Nelson Adekoya
Centers for Disease Control and Prevention

Natasha Close
Office of Communicable Disease Epidemiology
Washington State Department of Health

Zachary Stein
Bureau of Epidemiology
Kansas Department of Health and Environment

Jill Baber
Division of Disease Control
North Dakota Department of Health

Syndrome Definition Committee
International Society of Disease Surveillance

The authors would like to thank the reviewers for their expertise

Jessica R. White
Office of Epidemiology
Maricopa County Department of Public Health

Daniel Bedford
Office of Health Informatics
Wisconsin Department of Health Services

Background

Document objective

Syndromic Surveillance for United States federal, state, and local health authorities is a relatively new field involving the passive collection of healthcare visits in a way that is timely. This passive collection is followed by rapid analysis to respond to public health events.

The ability to develop syndrome definitions is key to supporting public health practitioners in monitoring emergency department (ED) visits in near-real time and aid in situational awareness (e.g., heat-related illnesses, food-related illnesses, influenza-like illnesses, and ice-related injuries) or public health threat detection. Well-tested syndrome definitions are essential to syndromic surveillance practice.

This 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.

Starting in November of 2016, the ISDS Syndrome Definition Committee began a group approach to developing a syndrome definition that could be documented and used as an example in best practices. The topic selected by the committee was a suicide-related syndrome and is referenced throughout this guidance document. The methods described in this document should transfer to most syndromic surveillance platforms, but this guidance document was based on the National Syndromic Surveillance Program (NSSP) BioSense Platform, ESSENCE (Electronic Surveillance System for the Early Notification of Community-Based Epidemics).

What is a syndrome definition?

A syndrome definition is a set of criteria that identify visits to a healthcare facility. In syndromic surveillance, the terms queries, definitions, syndromes, subsyndromes, groupings, and classifications are often used interchangeably. A syndrome definition is comprised of one or all of the following components:

1. *Discharge diagnosis codes*
   These standardized codes provide a method to search the clinically assigned diagnoses to capture records of interest. There are two coding systems commonly found in syndromic data: International Classification of Diseases, Clinical Modification (ICD CM) and Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT).
2. *Keywords*
   Terms related to the topic of interest can be used to search free text fields, such as patient chief complaint, admission reason, and triage notes to capture records.
3. *Negations*
   A query may define discharge diagnosis codes or keywords that should *not* be identified. These are used to exclude records falsely identified with the query.

Building a syndrome definition

Before you begin building your syndrome definition, you should determine if related syndromes already exist. Building a new syndrome may not be necessary if previously validated work can be utilized. Check the ISDS Knowledge Repository, NC Detect Emergency Department Case Definitions, and pre-built syndromes within your syndromic surveillance application. Additionally, it's a good idea to ask colleagues, which can be done efficiently by posting to the ISDS forums. As data sharing across jurisdictions becomes more prevalent, there is value in consistent use of syndromes across jurisdictions. The benefits to using an existing syndrome include consistency in surveillance practice across jurisdictions, pre-validated methods, and speed and ease of use.

An existing syndrome may not fit a jurisdiction's specific needs or purposes. When using borrowed syndrome definitions, you may find they lack your region-specific terminology, do not accommodate differences in your variables, need modifications to suit your platform, or do not fully cover the topic of interest. If you decide to build a syndrome, the following guidelines can assist in the process.

Scope and purpose

Building a syndrome begins by defining what you want to surveil and why. Considerations and questions you should ask include:

- **Public health concern.** What is it and what possible ways can it present in syndromic data?
- **Sensitivity versus specificity**. Is the aim to capture as many related visits as possible with some "false positives" or include as few false positives as possible, even if it means missing some visits?
- **Time frame**. What time frame would you like information to cover and what time frame do your data cover?
- **Data sources**. Which data sources will you use? Are there specific geographies or facility types that you want to include or exclude? Do data from these locations include the data elements necessary for your syndrome?
- **Purpose**. Will the definition be used for trend analysis or active case detection? Will syndromic surveillance data capture actionable information for the specific public health concern?

The SDC Experience

The SDC surveyed the syndromic surveillance community to identify syndrome definition needs. The feedback received highlighted the need for a robust suicide-related syndrome.

When designing the suicide-related syndrome, SDC members shared feedback from key stakeholders in their jurisdiction regarding the need and purpose to surveil both self-harm and suicide ideation. Based on this feedback, the decision was made to create a broad definition which returns self-harm, suicide ideation and suicide attempt visits. This definition can be later modified for subsets of this population.

The SDC aimed to create a definition which would capture as many visits as possible, while limiting false positives. This was achieved by using keywords, ICD-10-CM, ICD-9-CM and SNOMED-CT to capture visits and negations to limit false positives. We queried all historical chief complaints in their original form, admission reason, and discharge diagnosis fields.

Lastly, we evaluated the definition on both emergency department and inpatient visits from several state and local jurisdictions as well as national level data available in NSSP ESSENCE. When using other data sources with this definition, jurisdictions should re-evaluate its effectiveness.

Compiling the components

Once you have determined the scope of the records you are attempting to identify and the purpose of identifying these records, you are ready to start gathering the components of a syndrome definition.

Discharge Diagnosis

In some cases, there are diagnosis codes for the specific injury, illness, or event of interest. These can be very useful for identifying visits of interest, but might be rarely used, incorrectly used, delayed, or missing. When considering use of diagnosis codes, you will want to balance timeliness and sensitivity with specificity. For example, there may be a code for *Salmonella*, but if it is rarely used, it may be of little value. Conversely, if you are looking for healthcare associated infections, there may not be a code that specifically identifies the records of interest. Diagnosis codes are sometimes desirable over keyword searches as they require clinical assessment in order to be assigned and a presumably a health care provider has determined the presence of a specific condition.

Consider whether you want to search for only the top-level diagnosis code to capture a broader range of visits that may be of interest rather than narrowing down to a specific code (e.g., B16 = Acute Hepatitis B vs. B16.0 = Acute Hepatitis B with hepatic coma). Facilities may transmit discharge diagnosis codes with or without a decimal point, so consider adding both to your syntax (e.g.,B16.0 and B160) Additionally, when gathering a set of relevant codes, keep in mind the coding systems that are used by your data providers. For example, if you are looking for visits that occurred prior to 10/1/15, you will want to include ICD-9-CM codes. Please note, ICD-9-CM codes may continue to be used by some data providers after 10/1/15. If focusing on visits occurring after 10/1/15, include ICD-10-CM codes. Be aware that there is also some overlap of ICD-9-CM and ICD-10 CM codes that may result in capture of the wrong records if high-level codes are used. Some data providers are also sending SNOMED codes in addition to, or in place of, ICD codes, so these may need to be considered as well. Diagnosis codes most frequently appear in either Discharge Diagnosis or Admit Reason fields.

Keywords

Keywords can include signs, symptoms or other terms that indicate the presence of the illness, injury, exposure, population, or involvement in an event of interest. Many conditions have very non-specific signs and symptoms. As a result, searches for signs and symptoms may not provide enough specificity, and thus, depending on the purpose of the search, may not be useful. If your goal is to find everyone that *might* have your condition of interest, this may not be an issue.

If looking for records tied to an event (e.g., motor vehicle collision, food poisoning), it can be very useful to include some keywords in your search. When using fields that contain rich text (e.g., triage notes), you are more likely to need negations and accommodate keywords of interest used when gathering patient medical history but that do not always indicate relevance to the acute condition of interest.

Fields that are particularly useful for these types of searches are Triage Notes, Clinical Impression, Admit Reason, and Chief Complaint.

When compiling keywords, also consider:

- Misspellings/typos – Is the word frequently misspelled? Does it frequently have typos? Do you want singular and plural or past and present forms? If so, try to shorten the word to search for just a portion of the word to avoid some of these common issues. It is important to include any possible spellings. Incorporating regular expressions into syndrome definitions can also help limit the number of terms that need to be included to account for these variations.
- Acronyms and abbreviations– Often, short-hand will be used for keywords of interest (e.g., FL = flu – like or Florida). Consider what abbreviations or acronyms may be used for your keywords and include them in your search. Keep in mind that the Chief Complaint parsed field in ESSENCE expands many common acronyms and short-hand, so if using that field, you will want to search for the expanded terms rather than short-hand. You can investigate the handling of terms in the "more" tab in ESSENCE.
- Clinical vs. lay terms –Some fields will contain clinical vernacular while other fields contain lay descriptions of the condition of interest. Match your search terms to the fields that you are searching. If you are searching chief complaint, you will likely want to think about how a member of the general population might describe the condition, whereas Admit Reason or Clinical Impression can have more clinical terminology.

Keywords and discharge diagnosis can be identified in several ways, including:

- Literature on relevant signs/symptoms, terms, and diagnosis codes for the outcome of interest
- Syndromic case definitions from other sites
- Diagnosis code reference guides and tools to determine which code(s) to include/exclude and desired level of specificity
- Diagnosis code-positive visits in the syndromic surveillance application, followed by reviewing the resulting records to identify potential keywords present in free text fields (e.g., Chief Complaint, Triage notes, Clinical impression)
- Systematic text mining (e.g., word clustering, chief complaint terms or diagnosis code frequency) of records to identify additional terms to include/exclude once you have an initial list of keywords or codes.
- Consulting with subject matter experts on symptomology, discharge diagnosis use and other considerations

To compile a starting definition, the SDC put out a request to the syndromic community for existing suicide-related syndromes. We collected four definitions from New York City, Seattle and King County, Kansas and an existing ESSENCE definition. We also received feedback from committee members and consulted with subject matter experts from the CDC, Division of Violence Prevention, Surveillance Branch on additional keywords and diagnosis to include in the definition. Additionally, we reviewed online resources for missing diagnosis codes

A list of valuable resources to assist in compiling keywords or diagnosis codes can be found here.

Negations (or exclusion terms)

Once you have created your list of terms, codes, and various spellings or abbreviations, you should also think about negation terms. Some of the keywords or partial keywords may pick up records you are not interested in (e.g., "flu" is contained in "reflux"). It may be necessary to include exclusion terms to avoid capturing these records. Often, the need for exclusion terms will become apparent later in the validation process, but it's good to think about any that may obscure your results early in the process.

Once you have compiled all your potential syndrome components, it is time to program your syndrome definition into your syndromic surveillance application. A text editor can be useful to see the whole definition, ensure matching parentheses, and modify as necessary. Try to think about how a visit for your condition of interest may be unique. Consider which fields may be useful and the pros and cons of each. Being familiar with your data and knowing what is in each field (Are chief complaints usually a single word, short phrase, or longer sentences?) and how often each field is populated (Are Triage Notes often or seldom available?) are very useful. It's also important to consider other criteria that may help to identify the visits of interest. Is it limited to a certain facility location and facility type? Are there any other important distinguishing patient characteristics (e.g., age groups, sex, geography, race, ethnicity, patient class)?

Refining the syndrome

You have developed an initial syndrome definition, it is now time to refine it. There are several methodologies to refine a syndrome including a manual record level review, systematic text mining, and more advanced machine learning systems. In this document, we outline the methods employed by the SDC to refine the suicide-related syndrome.

1. Visualize a time-series graph of your query. It is helpful to view a time-series first to assess how many records are captured. This is a quick way to determine whether the definition is too narrow or too broad. In some cases, you may have no records returned and may want to rethink your query. In other cases, there are large numbers of visits indicating that some inclusion or exclusion criteria may need refining.

2. Manually review the record-level data and categorize each record as probable, ruled out or undetermined for the scope of your definition. A record is categorized as probable if a specified discharge diagnosis is listed or if the free text field (i.e. Chief Complaint) clearly relates to the topic of interest. A record is categorized as ruled out if it is clearly not related to the condition of interest (i.e., a false positive). All other records are categorized as undetermined.

Manual review of records helps discover additional keywords and negations. It will help identify needed syntax modifications, especially when using partial words. Query text strings may be part of other, unrelated words that you hadn't thought about.

3. Calculate the percentage of probable records captured by each keyword. This can be done using a statistical software such as SAS or R (Sample SAS code). The percent probable is a good indicator of how well the keyword is performing to capture true positive records.

   Percent Probable = (# Probable records identified / # Total records identified) * 100

4. Modify your syndrome based on the insight from manual review. Low performing terms can either be removed or modified, and negations can be added to improve performance. Always keep your initial sensitivity and specificity goals in mind. Low performing terms can be included in the query if the goal is to create a sensitive definition. For a more specific definition, it is ideal to remove low performing terms in order to limit false positives.

   In some cases, certain keywords or codes may capture zero or very few records. If they are also not capturing false positives, then it may be best to keep those terms in your definition for the sake of completeness. If, however, they are adding in false positives or your list of terms is already very long, you may opt to leave them out of your syndrome definition and lose those few true records that were being captured. In other scenarios, your list of terms may be fine, but the fields you are searching are causing false positives. For example, if your term is appearing in triage notes, but not in the context of the acute injury/illness, you may want to consider removing triage notes from your list of fields searched. It may be overcomplicating your query and not actually capturing many, if any, additional records of interest.

5. Repeat steps 1-5 until you are satisfied with both sensitivity and specificity.

After the initial compilation of keywords and diagnosis codes, the Suicide-related definition was run on the historical chief complaint, admission reason and discharge diagnosis fields in NSSP ESSENCE. Manual review of records was completed. If a record had a self-harm or suicide related discharge diagnosis or if the free text fields described self-harm, suicide ideation or attempt, the record was categorized as probable. If the record was clearly unrelated and a false positive, it was categorized as ruled out. All other records were categorized as undetermined.

Table 1.Example of manual review categorization

Admission Reason	Chief Complaint History	Discharge Diagnosis	Determination from manual review
Depression	Depression	R45851	Probable
SUICIDE ATTEMPT LIP LAC JUMPED OFF ROOF	{1};SUICIDE ATTEMPT LIP LAC JUMPED OFF ROOF;	F41.9	Probable
LEFT SHOULDER PAIN	{1};Was repositioning pt this am by herself. Left shoulder did not hurt initially but throughout the day became more painful;		Ruled Out
Mental Health Evaluation	MENTAL HEALTH EVALUATION	F329	Undetermined

The percent probable was calculated for each keyword. Low performing terms were either removed or negations were added to increase performance.

Table 2. Example of percent probable calculation

Keyword	N	Probable	Undetermined	Ruled Out	% Probable
SUCI	28	28	0	0	100%
IDEATION	109	107	1	1	98%
SELF and LACERA	15	12	0	3	80%
SELF and CUT	29	21	1	7	72%
LIFE and END	7	3	0	4	43%
SELF and INJUR	20	7	0	13	35%
SUID	5	1	0	0	20%
GO ON	13	0	1	12	0%

Based on this review, we removed terms such as "Go on" that had 0% probable records. We added negations (end of life care) for terms like "Life and End" to improve performance.

This process was repeated several times using national, state and local data until the ideal sensitivity and specificity were achieved. The final version of the suicide- related syndrome definition captured 94% probable

A syndrome definition will never have a perfect sensitivity and specificity. The purpose of your surveillance should guide when a syndrome definition is ready to be used in practice. For example, if the purpose is to monitor trends over time, a more specific definition may work, however if the goal is to detect cases, it may be useful to create a sensitive definition. Lastly, a syndrome definition must be updated periodically to adjust for changes over time in systems, reporting requirements, and data.

Dissemination

Syndromic surveillance is a field that involves a lot of collaboration and sharing among the community of public health practitioners. Even a precursory syndrome definition might be the best syndrome definition that exists for that type of surveillance. You should never assume that someone else has or does it better; share your queries with those around you!

The SDC Suicide-related syndrome can be found in the ISDS Knowledge Repository, Syndrome Definition Library and as a CCDD Category in NSSP ESSENCE.

Sharing should be the next step. Sharing not only allows others to try your query, but also allows the community to comment on and improve your work. Some methods for sharing include:

ISDS Syndrome Definition Library (SDL)

, within the Knowledge Repository, is a place to post your developed syndrome definitions, receive feedback from the community, and search for syndrome definitions developed by others. It is curated by the ISDS Syndrome Definition Committee (SDC). The SDL has minimal requirements; the SDC wants to encourage users to post drafts, post their own versions of definitions already in the SDL, as well as post updated versions of definitions already posted. All SDL posts get credited to the author and the organization they represent. So far, the SDL has received submissions from Local Health Departments, State Health Departments, CDC NSSP, ISDS, and CSTE Workgroups.

NSSP ESSENCE CCDD Category

 is a query disseminated by NSSP and incorporated into NSSP ESSENCE. If you have a query that you'd like to turn into a CCDD Category, the process can be started by contacting NSSP, who will then work with you to validate your query. Once a query has been validated, it is written as a structured query language (SQL) statement that is added with the next ESSENCE update cycle. This SQL statement will look at all historical NSSP ESSENCE data and apply your new CCDD Category tag to any visit meeting the criteria. Working forward, your CCDD Category will also tag any new visits entering NSSP ESSENCE processing. The result is your CCDD Category being listed under the CCDD Categories field in the NSSP ESSENCE Query Portal for all NSSP ESSENCE users to use.

There are multiple reasons to pursue turning your query into a CCDD Category:

- Query Sharing – NSSP ESSENCE users regularly check for new CCDD Categories and try them out on their data. This helps advance syndromic use within jurisdiction. If others are using your definition, you may also get feedback on how it can be further refined.

- Increased Confidence by Users – Due to the validation steps required for a new CCDD Category, other users may be more likely to use a query once it's published as a CCDD Category.
- Repeatability and Ease of Use – Rather than copying free-text query language and setting other query criteria, a CCDD Category allows a 1-button click to use very advanced queries. This ensures your query is accessible to all users and allows a higher level of repeatability among NSSP ESSENCE users.
- Faster Runtimes – Turning a very complex query into a tag that automatically gets applied to NSSP ESSENCE visits means that you're querying the tag itself and all the processing is done in advance. This results in a much faster query. Faster queries on the front-end also means less processing load on the NSSP servers.

In some states, sharing the raw code for the SDC Suicide-Related Definition with users saw limited use, but use of the definition increased after it was incorporated as a CCDD Category.

CCDD Category submission process

This process can be started through the NSSP Service Desk at http://support.syndromicsurveillance.org

Submit a ticket under ESSENCE→ General Questions and include "CCDD Category" in the Summary. Support staff will work with you to check the query's code, applicable fields, and finalize a name for CCDD Category creation.

Final Notes

After putting your syndrome definition through the multiple checks outlined in this guidance document, you should have a syndrome that performs reliably and predictably on your target data. Syndromic surveillance practice is a fluid process and a good, solid syndrome definition leads to deeper knowledge on a subject and often many new questions to answer. Always be looking forward to next steps like validating against other data sources, incorporating user feedback, generalizing your query to the whole USA, narrowing it to a specific hospital or region, breaking it into its subsequent pieces, or even using it as a small piece of a much larger query. Most of all, use your definitions and the definitions of others to enhance and build public health practice.

Resources

ICD-10 codes: http://www.icd10data.com
ICD -9 codes: http://www.icd9data.com
SNOMED: http://browser.ihtsdotools.org/
ISDS Knowledge Repository: https://www.surveillancerepository.org/
ISDS Syndrome Library: https://www.surveillancerepository.org/search/syndrome
ISDS Forums: https://www.healthsurveillance.org/forums/Default.aspx
NC Detect Emergency Department Case Definitions: http://ncdetect.org/case-definitions/

Sample SAS code for percent probable calculation

After manually reviewing the line list, the following code can assist in calculating the percentage of probable records captured by each keyword.

```
***CREATES A VARIABLE FOR EACH KEYWORD AND ASSIGNED "1" IF THE RECORD
CONTAINS THAT KEYWORD;
Data SUICIDE;
SET WORK.SUICIDE;    ***CHANGE THE NAME OF THE DATSET TO MATCH THE
IMPORTED DATASET;
IF    find(upcase(ChiefComplaintOrig), "SELF") > 0 AND
find(upcase(ChiefComplaintOrig), "HARM") > 0 THEN SELFandHARM = 1;
ELSE SELFandHARM = 0;
RUN;


***CALCULATES THE %PROBABLE FOR EACH KEYWORD;
TITLE "SELF and HARM";
PROC FREQ DATA= SUICIDE;
TABLES CLASSIFICATION; **VARIABLE NAME CONTAINING PROBABLE, RULED OUT
& UNDETERMINED CLASSIFICATION;
WHERE SELFandHARM =1;
RUN;
```

References

1. Wagner MM, Hogan WR, Chapman WW, Gesteland PH. Chief complaints and ICD codes. Handbook of biosurveillance. Edited by: Wagner MM, Moore AW, Aryel RM. New York: Academic Press, 2006, pp333-60.
2. Warns-Petit E, Morignat E, Artois M, Calavas. Unsupervised clustering of wildlife necropsy data for syndromic surveillance. MBC Veterinary Research 2010;6:56.