



4. Understanding Clinical Data and Workflow

Defining Data Elements Guide

One of the fundamental decisions you must make when developing an EHR-based surveillance program revolves around determining and defining the required data elements for the program and how they will be encoded. This step helps you to answer these important planning questions:

1. What data do we need from an EHR system to meet our surveillance goals?
2. Are those data currently captured in the EHR?
3. What standards are used to encode those data? How consistently are they used within a health care organization and across organizations?
4. Who is actually collecting and entering the data, for what purpose(s), and with what attention to quality (this last question is addressed in the *Understanding Clinical Data and Workflow* tool in this section)?

Defining data elements must be clearly linked to your surveillance goals, the population being surveilled, the indicators you are monitoring and the reports you anticipate needing to generate. The following section provides more detail on each step.

Surveillance system goals

First, articulate your surveillance goals in a way that makes clear what data you need to collect. As important, those goals should help you determine what data you don't need to collect: adding any new data element to a clinical workflow and possibly the EHR has non-trivial costs associated with it in terms of staff training, software enhancements and testing, and ongoing staff monitoring for data quality. A health care organization is unlikely to agree to such a request without a clear return on that investment via a benefit (see *Making the Value Case*). If the health care organization does agree to your request, work with them to understand where the data element or elements would be collected and what the organization would have to do to collect it—for example, make changes to workflow, add training, or conduct ongoing monitoring.

Based on the definition that your team came up with earlier in making the value case, enter a clearly articulated goal or set of goals for the surveillance program.

The following statements provide examples of clearly articulated surveillance system goals:

- The surveillance program will monitor risk factors for CVD in <your jurisdiction or region's name> and provide timely information to partners to inform decisions and interventions.
- The program seeks to provide actionable information and value to all partners to improve health outcomes and will be sustainable in part by only collecting data already captured by clinical partners for Meaningful Use, quality reporting or other purposes.



4. Understanding Clinical Data and Workflow

Defining Data Elements Guide

Population of interest

Define the population being surveilled, in terms of characteristics such as age, gender, race and ethnicity, geography and condition to help further clarify required data elements.

The following statements provide examples of defining a population of interest:

- All adolescents 12-18 years of age in the jurisdiction with at least one health care visit in the past year.
- The population defined by the following numerator and denominator definitions:
 - Numerator: All persons who are seen in <define the types or names of participating healthcare organizations> and for whom data on any of the indicators is recorded in the EHR.
 - Denominator: All persons residing within the jurisdiction.

Surveillance indicators and reports

Describe any indicators to surveil and any reports that you want to generate to specifically define the data elements required by those reports.

The following examples, provided courtesy of the Colorado Health Observations Regional Data Service (CHORDS), show draft indicators:

Tobacco use and exposure (unduplicated counts to the extent feasible)

- Adults (over 25) with any tobacco use
- Adults (over 25) smoking cigarettes
- Young adults (18-24) with any tobacco use
- Young adults (18-24) smoking cigarettes
- Adults exposed to second hand smoke
- Children (under 18) exposed to second hand smoke

Cardiovascular disease

- Adults with elevated blood pressure reading at most recent visit
- Adults with diagnosed hypertension



4. Understanding Clinical Data and Workflow

Defining Data Elements Guide

- Adults diagnosed hypertensive with a normal blood pressure at most recent visit
- Adults with elevated lipids
- Adults diagnosed with hypercholesterolemia
 - Adults diagnosed with hypercholesterolemia with elevated at most recent visit
 - Adults diagnosed with hypercholesterolemia with a current statin prescription

Healthy eating and active living

- Adults obese or overweight at most recent visit (BMI ≥ 25.0)
 - Adults obese at most recent visit (BMI ≥ 30.0)
 - Adults overweight at most recent visit (BMI ≥ 25.0)
- Children at an excessive weight (obese or overweight) at most recent visit (BMI percentile ≥ 85)
 - Percent of children (under 18) overweight at most recent visit (BMI percentile = 85 – 94)
 - Percent of children (under 18) obese at most recent visit (BMI percentile ≥ 95)

Data elements definition

Finally, refer to the tables below for examples when defining the data elements that you need for your surveillance program. The tables, which are only partially completed, provide examples that should be used and are for illustration purposes only.



4. Understanding Clinical Data and Workflow

Defining Data Elements Guide

Table 1: Surveillance System Data Requirements

Determine exactly what data elements your system needs to achieve the surveillance system goals. Include data needed to measure any indicators of interest, support specific analyses, or generate any desired reports. Document them in the worksheet. Then document the standard code system and value set for each of those data elements, when such a standardized vocabulary exists. Consider developing this data model as a collaborative activity with your data trading partners (see *Understanding Clinical Data and Workflows* for information on structured vs. unstructured data).

Concept Name	Concept Definition	Code System	Value set	Format	Required
Patient birth date	The date the person was born	None	None	YYYYMMDD	Yes
Patient address	The address of the patient's residence	None	None	Street, apartment, city, state, zip as separate fields	Yes
Ethnicity	Indicates if the patient is Hispanic or not Hispanic	CDCREC (CDC race and ethnicity codes)	2135-2 (Hispanic), 2186-5 (not Hispanic)	String, length 6	If known
Tobacco use	Use of any tobacco product	SNOMED CT	449868002 <i>or</i> 428041000124106 <i>or</i> 428071000124103 <i>or</i> 428061000124105	String, length 15	Yes
Second hand smoke exposure	Exposed to smoke of any kind	Local	None, will vary	String, length 20	If known
Hypertension	Diagnosis based on repeated blood pressure reading	ICD-9	401.9 (Hypertension NOS)	String, length 6	Yes
Hypercholesterolemia	Diagnosis based on blood tests for lipid levels	ICD-9	272.0 - 272.4 (Hyperlipidemia NEC/NOS)	String, length 5	Yes
Obesity	Adults obese at most recent visit	None	BMI as calculated value	Real number	Yes
			Height	Real number	Yes, to validate BMI calculation
			Weight	Real number	Yes, to validate BMI calculation



4. Understanding Clinical Data and Workflow

Defining Data Elements Guide

Table 2: Sending System Data (Discrepancies from Surveillance System (in bold))

For each of your data trading partners, document the format and content of the data from their EHR systems. List any missing or incompatible concepts and value sets in Table 3 for further exploration.

Concept Name	Concept Definition	Code System	Value set	Format	Compatible
Patient birth date	The date the person was born	None	None	MM/DD/YYYY	Can map
Patient address	The address of the patient's residence	None	None	Street, apartment, city, state, zip as separate fields	Yes
Screening date	The date the A1c test was drawn	None	None	MM/DD/YYYY	Can map
Hispanic?	Indicates if the patient is Hispanic or not Hispanic	Local	H (Hispanic), NH (not Hispanic), U (unknown)	String, length 2	Can map
Tobacco use	Use of any tobacco product	Local	None	String, length 15	Natural language processing to extract and classify
Second hand smoke exposure	Exposed to smoke of any kind	None	None, free text only	String, length variable	No
Hypertension	Diagnosis based on repeated blood pressure reading	ICD-9	401.9	String, length 6	Yes
Hypercholesterolemia	Diagnosis based on blood tests for lipid levels	ICD-9	272.0 - 272.4	String, length 5	Yes
Obesity	Adults obese at most recent visit	None	Height and Weight only	Real numbers	Calculate BMI

