

DATA ANALYSIS PLAN

**Integrated Next-generation Surveillance in Global Health: COVID-19 surveillance
secondary data source analysis**

I-TECH HQ

April 28, 2022

Version 1.0

COVID-19 Secondary data sources analysis Goal and Objectives:

Purpose: to describe and identify the COVID-19 surveillance system by analyzing the available open access datasets using CDC's public health surveillance systems evaluation tools and indicators for the period 2020 – 2022.

Objective 1: To identify and catalog open datasets containing data from COVID-19 surveillance systems.

Objective 2: To assess surveillance system quality by means of data quality and completeness indicators

Main questions:

1. What is the completeness and quality of the COVID-19 epidemiological datasets?
2. Does the completeness and quality of information vary by location and time?
3. Are there any other characteristics from the open data that can provide information about the comprehensiveness of the COVID-19 surveillance system

Methods:

This will be an exploratory data analysis of COVID-19 open-access datasets from Peru. The datasets are ready to download without prior request from the Peruvian Government. Further additional datasets can be accessed and provided by request. Public health surveillance system assessment indicators will be constructed from the datasets. The datasets contain id numbers that enable merging, which will be required to build some of the indicators described below.

Timeline:

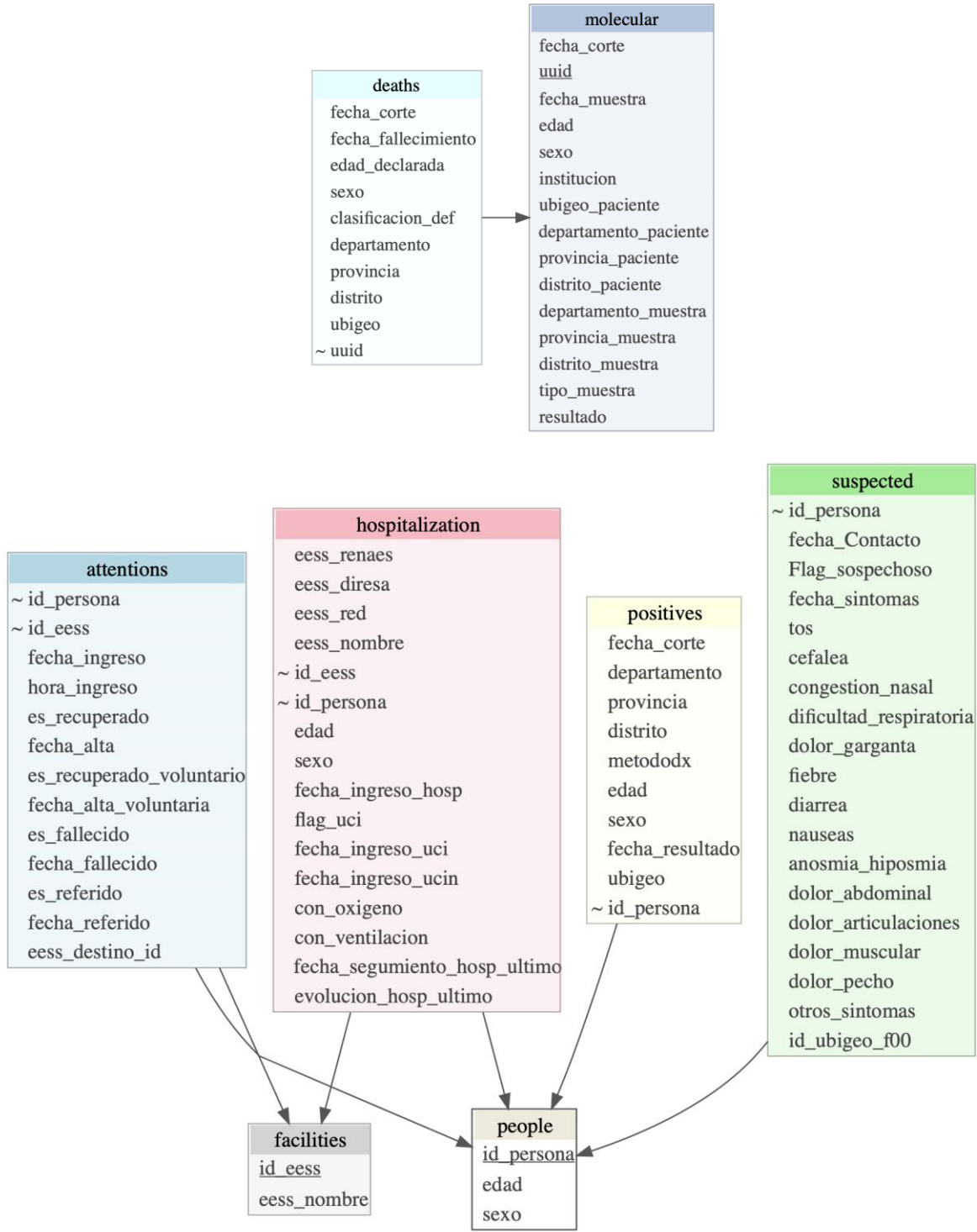
Data sources: Data sources for the COVID-19 surveillance secondary data source assessment are shown in Table 1 below.

Table 1. Secondary Data sources of COVID-19 Surveillance

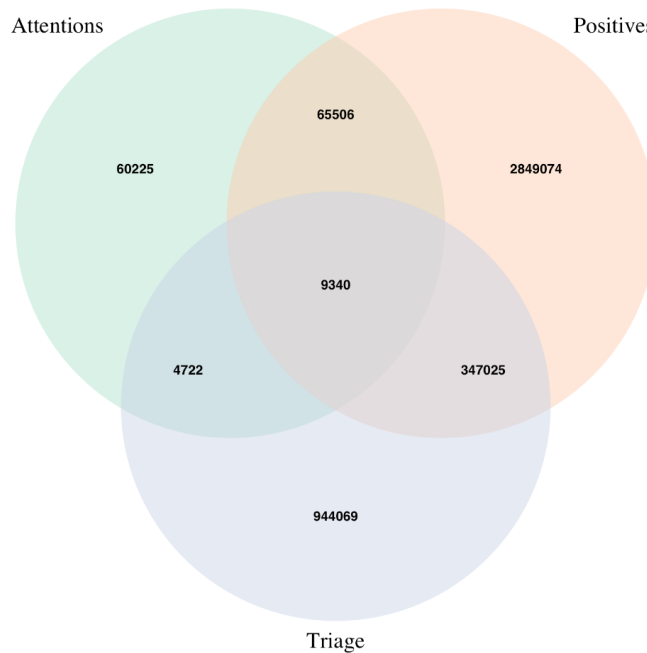
Data sources	Level	Source	Variables
COVID-19 deaths dataset	Individual	Vital registration system	Date, age, sex, location, id,...
PCR tests dataset	Individual	INS PCR NetLab system	Date, age, sex, institution, location, sample, result, ...
Positive results dataset	Individual	MoH dataset (SISCOVID?)	Date, diagnostic method, location, age, sex, id, ...
Triage	Individual	F00 form (SISCOVID?)	Date, symptoms, location
Clinical Attention	Individual	SISCOVID?	Date, attention time, diseased, id...
Hospitalizations	Individual	F500	Date, institution, location, age, sex, id ...

Data Model

Proposed entity-relationship model:



Distribution of patients in datasets that share the *id_persona* identifier to guide analysis and indicators:



Data management

All data management and analyses will be conducted by XXXX Data will be downloaded from the Peruvian Open Data portal.

Indicator definitions

#	Indicator definition and relevance	Numerator and denominator/ Calculation	Data source	Possible disaggregation
Attribute 1. Completeness				
1a	<p>Proportion of cases recorded in a database with no missing required information</p> <p>Measures whether the information recorded about the cases is correct.</p>	<p>Numerator: Total cases recorded with no missing required information</p> <p>Denominator: Total cases recorded, including unknown and missing items</p>	<ul style="list-style-type: none"> • PCR tests dataset • Positive results dataset • COVID-19 deaths dataset • Triage • Clinical Attention • Hospitalizations 	<ul style="list-style-type: none"> • Dataset • Institution • Department • Time
1b	<p>Proportion of missing information required by field</p>	<p>Numerator: # completed data fields</p> <p>Denominator: # total data fields including</p>	<ul style="list-style-type: none"> • PCR tests dataset • Positive results dataset • COVID-19 deaths dataset 	

#	Indicator definition and relevance	Numerator and denominator/ Calculation	Data source	Possible disaggregation
	Refers to the extent of errors within the system.	unknown and missing items	<ul style="list-style-type: none"> • Triage • Clinical Attention • Hospitalizations 	
Attribute 2. Validity				
2a	Proportion of cases/deaths complying with case definition	Numerator: # cases/deaths meeting case definition Denominator: # number cases/deaths (IHME???)	<ul style="list-style-type: none"> • COVID-19 deaths dataset 	<ul style="list-style-type: none"> • Dataset • Region • Department • Time
2b	Proportion of inconsistencies and errors within a dataset (e.g., inconsistent dates)	Numerator: # cases with coding errors within the dataset Denominator: # cases in the dataset	<ul style="list-style-type: none"> • PCR tests dataset • Positive results dataset • COVID-19 deaths dataset • Triage • Clinical Attention • Hospitalizations 	
Attribute 4. Concordance				
3a	PCR Testing results concordance Refers to the proportion of concordance in positive PCR tests between two different data sources	Numerator: # Positives PCR results in PCR Tests dataset Denominator: # Positives PCR results in Positive results dataset	<ul style="list-style-type: none"> • PCR tests dataset • Positive results dataset 	<ul style="list-style-type: none"> • Dataset • Region • Department • Time
3b	Virologic death criteria concordance with PCR positive tests Refers to the proportion of concordance between COVID-19 deaths classified by virologic SINADEF criteria and PCR positive tests	Numerator: # PCR positive tests Denominator: # COVID-19 related Deaths classified as “Virologic” SINADEF criteria	<ul style="list-style-type: none"> • PCR tests dataset • COVID-19 deaths dataset 	
3c	Hospitalization and clinical attention dataset concordance	Numerator: # unique individuals in the Hospitalization dataset	<ul style="list-style-type: none"> • Hospitalizations • Clinical Attention 	

#	Indicator definition and relevance	Numerator and denominator/ Calculation	Data source	Possible disaggregation
		Denominator: # unique individuals in the attention dataset		
Attribute 4. Timeliness and reactivity				
4a	<p>Main time delays</p> <p>Refers to calculating time delays that are important for notifiable communicable diseases.</p>	<p>Delay to report = date of report - date of onset (or date of first symptoms)</p> <p>Delay to lab result = date of lab result – date of start of symptoms</p> <p>Delay to attention = date of attention – day of onset of symptoms</p> <p>Delay to hospitalization = date of hospitalization – day of onset of symptoms</p> <p>Delay to ICU = date of ICU admission – date of onset of symptoms</p>	<ul style="list-style-type: none"> • Positive results dataset • Triage • Hospitalizations • Clinical Attention 	<ul style="list-style-type: none"> • Dataset • Region • Department • Time
4b	Proportion of Triage that was verified within 48 hours of detection	<p>Numerator: # cases verified</p> <p>Denominator: # outbreaks that should have been verified</p>	<ul style="list-style-type: none"> • Positive results dataset • Triage 	
4c	Average time interval between the date of onset and date of the lab result	<p>Numerator: Time to lab report</p> <p>Denominator: Total number of times to lab reported</p>	<ul style="list-style-type: none"> • Positive results dataset • Triage 	

Analysis

Descriptive analysis by indicator will provide counts and percentages disaggregated by the proposed dimensions. Additionally, for disaggregated categories and indicators, a bivariate analysis will be constructed to determine the association between indicators and dimensions.

Choropleth maps of indicators will be constructed to represent the spatial distribution

Example Dummy Tables and Graphs

Table 1: Dataset summary – PCR Tests

Characteristic	n (%)
Number of PCR tests	
Tests results Positive Negative Other	

Table 2: Dataset summary – Positive tests

Characteristic	n (%)
Number of positive tests	
Diagnostic method PCR Antigen Rapid test	

Table 3: Dataset summary – COVID-19 Deaths

Characteristic	n (%)
Number Covid-19 Deaths	
Death criteria definition Clinical Epidemiologic Radiologic Serologic SINADEF Virologic	

Table 4: 1a. Proportion of cases recorded with no missing information

Indicator	n (%)
By Dataset Positive Tests PCR Tests COVID-19 Deaths	
By Region / Department	
By Institution (MINSAs, EsSalud, etc.)	
By Result	

Table 5: 1b. Proportion of missing information required by field

Indicator	n (%)
PCR Tests Reporting Date UUID Sample date Age	
Positive Tests Reporting Date Department Province District ...	
COVID-19 Deaths Reporting Date Date of death Age Sex ...	