Outbreak investigation tools: Epi core variables, Initial case investigation form and Data dictionary

Geneva, November 2019.

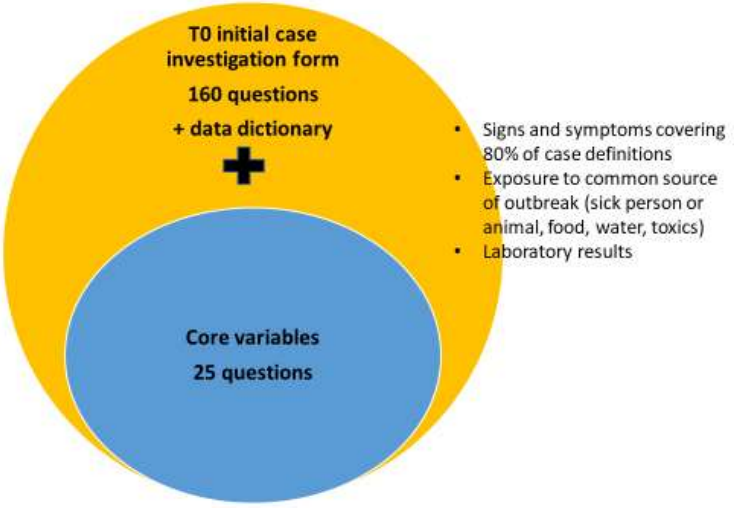
The *Data Collection Standard* is one pillar of the WHO outbreak toolkit project aimed to enhance the performance of patient-based data collection during disease outbreak investigations. Recurrent observations showed that the collection of too many variables can impair the quality and completeness of the data collected while adding little to the description of the outbreak. The WHO outbreak toolkit project adheres to the epidemiological principles of outbreak investigations, which is first to describe the outbreak by time, place and person to allow informed assessment of severity and risk of extension of the outbreak. The project developed a set of tools to correctly collect the necessary data.

The "Epi Core variables" represent the core information that needs to be captured during every case investigation, regardless of where the outbreak is occurring, or the causative agent involved. It represents a set of essential variables that should be collected, at a minimum, in an outbreak situation so that basic descriptive epidemiology analysis (i.e. case counting, case classification, description of the time, place and populations affected by the outbreak) can be conducted. This set of Epi Core variables should support the description of the outbreak in time (when), place (where) and persons (who) with the necessary details to guide the initial response activities. The method to identify the Epi Core variables is described below.

The T0 (T-zero) initial case investigation form represents the minimum set of variables to answer: the "when", "where" and "who", and the "what" and "how" questions about the outbreak or epidemic; that is to describe the medical signs of the disease or syndrome to develop the case definition and verify the case classification of each patient, and to draw hypotheses regarding the origin or source of the outbreak. The form also captures the results of a laboratory investigation to help disease identification (biochemistry) and to classify the case as confirmed, probable or suspect. The initial case investigation form integrates the Epi Core variables augmented with a list of signs and symptoms covering 80% of the signs and symptoms of the most frequent epidemic prone diseases; questions exploring the main sources and transmission patterns; and laboratory test results (figure 1). Expert assessment, performed by the working group, was used to validate the questionnaire. Discordant opinions were resolved by consensus.

The data dictionary for the T0 initial case investigation form provides definitions for each variable to ensure a common understanding of each variable. The data dictionary also includes standardized variable names, formats and where applicable, common coding and data categorization. The objective of the data dictionary for outbreak investigations is to improve comparability of data and information between outbreaks and the interoperability with data analysis and visualization software – better leveraging automated analysis and reporting programs. It will enable field epidemiologists to collect the same information with the same names and same formats for different outbreaks.

Figure 1: A Venn diagram, showing the relationship between "Epi Core variables" and the "Initial case investigation form".



Method for the identification of the EPI Core Variables:

The process of identifying the minimum set of core variables is based upon a combination of literature review and an experts working group (with members from World Health Organisation, US-Center for Disease Control, Médecins Sans Frontières, European Centre for Disease Control, the RECON group, Helmholtz-Zentrum fur Infektionsforschung, Epidemic Data Kit of the UK Rapid Support Team of London School of Hygiene and Tropical Medicine, Global Outbreak Alert and Response Network, Infectious Disease Data Observatory group, the WHO Rapid Response Team Knowledge Network).

The process began with the collection and review of a large number of existing case investigation forms which variables were systematically transferred into an Excel database for gathering and classification. Questions were classified into six categories: Notification and interview, case information, clinical information, exposure, laboratory confirmation and case classification.

In each category, the questions were ranked as 1) common to any outbreak, 2) disease specific and 3) context specific. The list of variables, classification and categories were refined upon circulation in the working group.

A total of eighty-two variables were selected as the set of variables common to all outbreak investigation of any cause and settings. That set was used for the identification of the Epi Core variables. To develop consensus amongst the group of experts to designate the Epi Core variables, a two-stage electronic Delphi process was conducted from January-June 2019 (article describing the process is pending). Participants were tasked to rank the variables as "essential", "high priority", "medium priority" or "low priority".

All variables which reached a 60% agreement among participants as being ranked as essential variables were included in the set of Epi Core variables for outbreak investigation.

A total of 25 variables were designed as Epi Core variables (table 1).

Table 1: List of variables designated as Epi Core variables

Classification	**Variables**
Notification interview	Unique case ID number, Date case was first reported, Name of reporting facility, Interview date

Case information	Case's surname, Case's first and second name, Date of birth Age at illness onset, Age in months, Biological sex, Neighborhood of residence, Address of residence, Town or city of residence, Administrative level 4 of residence, Administrative level 3 of residence, Administrative level 2 of residence, Administrative level 1 of residence, Administrative level 0 of residence
Clinical information	Date of first signs or symptoms, Hospitalized, Outcome of illness, Date of death
Exposure	Direct contact with symptomatic people Travelled outside since symptom onset, Locations travelled to

Perspectives

1. Work in progress:

The list of core variables and the minimum information to be collected are subject to refinement after real life experiences.

2. Enhancing the use of mobile devices

Our objective is to stimulate the use of electronic data capture tools in the field. The data dictionary has been developed in a format that can be easily operated in software applications for field data capture such as Kobo-collect, Go data, Epi-info, and Excel. These applications can run on mobile phones, tables, laptops or any other electronic device.

The WHO outbreak toolkit team
outbreaktoolkit@who.int