

WHO Glossary of Health Data, Statistics and Public Health Indicators

June 2024

A.....	1
ACCESSIBILITY	1
ACCOUNTABILITY	1
ACCURACY	1
ADMINISTRATIVE AREA.....	1
AGE GROUPINGS RECOMMENDED BY WHO.....	1
AGE-SPECIFIC MORTALITY RATE.....	1
ANONYMITY	1
B.....	1
BIOSPECIMEN.....	1
BIRTH REGISTRATION DATA	1
BIRTH REGISTRATION FACILITY	2
C.....	2
CAUSES OF DEATH	2
CHI-SQUARE(χ^2) TEST	2
CIVIL REGISTRATION AND VITAL STATISTICS (CRVS)	2
COMPLETENESS OF REPORTING	2
COMPOSITE INDICATOR.....	2
CONCURRENT VALIDITY	2
CONFIDENTIALITY	2
CONFOUNDING	2
CONSTRUCT VALIDITY	2
CONTENT VALIDITY	3
CONTINUUM OF CARE	3
CONVERGENT VALIDITY	3
CORE INDICATOR	3
CORRELATION ANALYSIS.....	3
COUNT	3
COVARIATES.....	3
COVERAGE	3
CREDIBILITY.....	3
CRITERION VALIDITY	3
D	3
DATA COLLECTION LEVEL.....	3
DATA COLLECTION METHOD.....	3
DATA CUSTODIAN	4
DATA INFORMATION PYRAMID	4
DATA INPUTS	4
DATA LIFE CYCLE	4
DATA PROVIDER.....	4
DATA QUALITY	4
DATA QUALITY ASSESSMENT	4
DATA QUALITY ASSURANCE (DQA)	4
DATA SOURCE	4
DATA TRIANGULATION	5

DEATH REGISTRATION DATA.....	5
DEATH REGISTRATION FACILITY.....	5
DEMOGRAPHIC SURVEILLANCE SYSTEM.....	5
DENOMINATOR.....	5
DESCRIPTIVE ANALYSIS	5
DIGITAL HEALTH.....	5
DISAGGREGATION.....	5
DISEASE SURVEILLANCE SYSTEM.....	5
DOMAIN.....	5
E.....	5
ECOLOGICAL ANALYSIS	5
ECOLOGICAL FALLACY	6
EPIDEMIC INTELLIGENCE.....	6
ESTIMATION METHOD	6
EVALUATION	6
EVENT BASED SURVEILLANCE	6
EXTERNAL CONSISTENCY OF DATA	6
EXTERNAL RESPONSIVENESS.....	6
F.....	6
FEASIBLE	6
FOCAL POINT.....	6
G.....	6
GENDER	6
GEOCODED	7
GEOGRAPHIC INFORMATION SYSTEMS (GIS)	7
GEOSPATIAL ANALYSIS.....	7
GEOSPATIAL DATA	7
GLOBAL DATABASE	7
GLOBAL HEALTH ESTIMATES (GHE).....	7
GOLD STANDARD	7
GRANULARITY	7
H.....	7
HEALTH CARE ADMINISTRATIVE DATA.....	7
HEALTH ESTIMATES	7
HEALTH FACILITY CENSUS	7
HEALTH FACILITY SURVEY	7
HEALTH IMPACT ASSESSMENT.....	8
HEALTH INDICATOR	8
HEALTH INEQUALITY	8
HEALTH INEQUITY	8
HEALTH INFORMATION SYSTEM	8
HEALTH RECORD	8
HEALTH SURVEY	8
HEAPING OF DATA	8
I.....	8
IMPACT INDICATOR	8
IMPUTATION.....	9
INCIDENCE RATE	9
INDICATOR-BASED SURVEILLANCE.....	9

INDICATOR CLASSIFICATION	9
INDICATOR DEFINITION	9
INPUT INDICATOR	9
INTEGRITY OF DATA	9
INTERACTION	9
INTERNAL CONSISTENCY OF DATA	9
INTERNAL RESPONSIVENESS	9
INTERNATIONAL CLASSIFICATION OF DISEASES (ICD)	9
INTERNATIONAL HEALTH REGULATIONS (IHR).....	10
INTERNATIONAL STANDARDS	10
INTEROPERABILITY	10
J.....	10
JOINT EXTERNAL EVALUATION (JEE)	10
K.....	10
KAPPA STATISTIC (K)	10
L	10
LINKAGE	10
M	10
MASTER FACILITY LIST (MLF).....	10
MATHEMATICAL OR STATISTICAL MODELS.....	10
MEAN	10
MEASURE	10
MEASUREMENT	10
MEASUREMENT LEVEL	11
MEASUREMENT METHOD.....	11
MEDIAN	11
MEDICAL CERTIFICATION OF CAUSE OF DEATH (MCCD)	11
METADATA.....	11
METHOD OF AGGREGATE ESTIMATION	11
METHODOLOGICAL SOUNDNESS	11
MICRODATA.....	11
MINISTRY OF HEALTH (MoH)	11
MONITORING.....	11
MONITORING & EVALUATION (M&E) FRAMEWORK.....	11
MORBIDITY DATA.....	11
MORTALITY CODER	11
MORTALITY DATA	12
N	12
NATIONAL HEALTH STRATEGIC PLAN	12
NATIONAL STATISTICS OFFICE (NSO)	12
NATIONALLY REPRESENTATIVE	12
NEGATIVE PREDICTIVE VALUE (NPV).....	12
NOTIFIABLE CONDITIONS.....	12
NUMERATOR.....	12
O	12
ODDS.....	12
OUTCOME INDICATOR	12
OUTPUT INDICATOR.....	12
P.....	13

PERCENTAGE.....	13
PERIODICITY	13
POPULATION BASED SURVEY	13
POPULATION CENSUS	13
POSITIVE PREDICTIVE VALUE (PPV)	13
POST ENUMERATION SURVEY (PES).....	13
PREDICTIVE VALIDITY	13
PREFERRED DATA SOURCES	13
PREVALENCE RATE	13
PRIMARY DATA	13
PRIMARY DATA SOURCES.....	14
PROCESS INDICATOR.....	14
PROCESS OF VALIDATION	14
PROCESSED HEALTH DATA	14
PROPORTION	14
PROXY HEALTH INDICATOR.....	14
PUBLIC HEALTH SURVEILLANCE SYSTEM.....	14
PUBLICLY AVAILABLE.....	14
PUNCTUALITY.....	14
P VALUE	14
R.....	15
RATE.....	15
RATIO	15
RATIONALE.....	15
RAW HEALTH DATA.....	15
RECORD LINKAGE.....	15
REGISTRAR	15
REGISTRATION FORM	15
REGRESSION.....	15
REGULAR ASSESSMENT	15
RELEVANCE	15
RELIABILITY OF DATA	15
REPORTING SITE.....	15
REPRESENTATIVE	16
RESPONSE RATE	16
RESPONSIVENESS.....	16
ROUTINE HEALTH INFORMATION SYSTEM (RHIS)	16
S.....	16
SAMPLE	16
SAMPLE SIZE	16
SAMPLING ERROR	16
SECONDARY DATA SOURCES.....	16
SENSITIVITY ANALYSIS.....	16
SEX	16
STAKEHOLDER.....	16
STANDARD OPERATING PROCEDURES (SOP)	17
STATISTICAL DATA.....	17
STATISTICAL METHODS.....	17
STATISTICAL OUTPUT	17
STATISTICAL SIGNIFICANCE	17
STRATIFICATION.....	17

STRUCTURAL VALIDITY	17
SUB-NATIONAL	17
SURVEILLANCE	17
SURVEY	17
SURVEY DATA.....	17
SUSTAINABILITY	17
SUSTAINABLE DEVELOPMENT GOAL (SDG)	18
T	18
TARGET POPULATION	18
TIMELINESS	18
TRACER INDICATOR.....	18
TREND	18
U	18
UNCERTAINTY MEASURE	18
UNDERCOUNT RATE.....	18
UNDERSTANDABLE/SIMPLICITY	18
UNIT OF MEASURE	18
USEFULNESS/UTILITY	18
V.....	19
VALIDITY.....	19
VERBAL AUTOPSY.....	19
VITAL EVENT	19
VITAL STATISTICS	19
W	19
WEIGHTING.....	19

WHO Glossary of Health Data, Statistics and Public Health Indicators

These definitions are considered in the context of WHO's technical work on health data, statistics and public health indicators.

A

ACCESSIBILITY

Definition: The ease with which users can find, retrieve, understand, and use data.

ACCOUNTABILITY

Definition: Answerability or legal responsibility for identifying and removing obstacles and barriers to health services. This should include responding to findings from monitoring and evaluation.

ACCURACY

Definition: The degree of closeness estimates are to the true values.

ADMINISTRATIVE AREA

Definition: Administrative area refers to the clusters that are administered by a departmental area. For example: county, district, province, state, sub-national, national.

AGE GROUPINGS RECOMMENDED BY WHO

Definition: Age groups for data analysis that capture a time interval representing a developmental stage in the life course of a human. WHO proposes age groups for analysis as follows:

- 0-6 days (early neonates)
- 7-27 days (late neonates)
- 28-364 days (post-natal infants)
- 1-4 years (young children)
- 5-9 years (older children)
- 10-14 years (young adolescents)
- 15-19 years (older adolescents)
- 20-24 years (young adults)
- 25-59 years (adults) in five year age groups
- 60-99 years (older adults) in five year age groups
- 100+ (older adults)

AGE-SPECIFIC MORTALITY RATE

Definition: A mortality rate limited to a particular age group. The numerator is the number of deaths in that age group; the denominator is the number of persons in that age group in the population.

ANONYMITY

Definition: The condition of being anonymous.

B

BIOSPECIMEN

Definition: Biological sample.

BIRTH REGISTRATION DATA

Definition: Data collected during birth registration, including place of birth, sex, etc.

BIRTH REGISTRATION FACILITY

Definition: Registration desk that receives birth notification, validates information, and enters it into central registration IT system.

C

CAUSES OF DEATH

Definition: Cause of death as all those diseases, morbid conditions, or injuries which either resulted in or contributed to death and the circumstances of the accident or violence which produced any such injuries. It does not include symptoms or modes of dying such as cardiac arrest.

CHI-SQUARE(χ^2) TEST

Definition: The Chi-square test of independence (also known as the Pearson Chi-square test, or simply the Chi-square) is one of the most useful statistics for testing hypotheses when the variables are nominal, as often happens in clinical research. Unlike most statistics, the Chi-square (χ^2) can provide information not only on the significance of any observed differences, but also provides detailed information on exactly which categories account for any differences found.

CIVIL REGISTRATION AND VITAL STATISTICS (CRVS)

Definition: The continuous, permanent, compulsory, and universal recording of the occurrence and characteristics of vital events pertaining to the population, as provided through decree or regulation in accordance with the legal requirements in each country. Vital statistics include events like birth, marriage, divorce, adoption, death, and cause of death. Civil registration is the universal recording of the occurrence and characteristics of those vital events pertaining to a specific population.

COMPLETENESS OF REPORTING

Definition: Reflects the percentage of reporting units that have provided data. This could be the completeness of facility reporting, district reporting or globally the completeness of reporting from countries.

COMPOSITE INDICATOR

Definition: An index composed of several indicators within a health topic to represent that topic; a composite indicator may combine indicators from across several health topics to represent a broader concept, such as universal health coverage.

CONCURRENT VALIDITY

Definition: Concurrent validity is one approach of criterion validity that estimates individual performance on different tests at approximately the same time.

CONFIDENTIALITY

Definition: Refers to the obligation to maintain confidentiality agreements that limit access or restrict the disclosure of certain types of information.

CONFOUNDING

Definition: The distortion of a measure of the effect of an exposure on an outcome due to the association of the exposure with other factors that influence the occurrence of the outcome. An additional variable related to the independent and dependent variables and that distorts the relationship between them.

CONSTRUCT VALIDITY

Definition: The extent to which the measure 'behaves' in a way consistent with theoretical hypothesis and represents how well scores on the measurements indicative of the theoretical construct, providing confidence in the meaningfulness and relevance of the results obtained using that measurement. Construct validity is often used for process indicators.

CONTENT VALIDITY
Definition: The degree to which an assessment instrument is relevant to, and representative of, the targeted construct it is designed to measure.
CONTINUUM OF CARE
Definition: Describes the pathway for measurement across the life course including reproductive health, pregnancy, childbirth, postnatal care for mothers and newborns and prevention and treatment measures for children, adolescents, adults, and older people.
CONVERGENT VALIDITY
Definition: how closely the indicator is related to other variables and other measures of the same construct. This approach is utilized when a gold standard does not exist. Convergent validity is often used for impact indicators.
CORE INDICATOR
Definition: Core indicators may be defined in collaboration with all key stakeholders (e.g., ministry of health (MoH), national statistics office (NSO), other relevant ministries, professional organizations, subnational experts, and major disease-focused programmes), and depends on the priority monitoring requirements related to health and health-related SDGs, among other health priorities.
CORRELATION ANALYSIS
Definition: Examination of the strength and direction of linear relationships between two continuous variables.
COUNT
Definition: a count gives the number of occurrences of the event(s) being studied, within a specified time and at a specified place. This is the absolute frequency and indicates the impact of a disease in precise numerical terms.
COVARIATES
Definition: Data, including non-health data, which are used in a statistical model to improve the estimation of the health indicator of interest. These data are population-specific and are available for every population included in the analysis. A common covariate is gross domestic product per capita.
COVERAGE
Definition: the extent to which the real, observed population matches the ideal or normative population.
CREDIBILITY
Definition: Confidence that users place in the statistics.
CRITERION VALIDITY
Definition: The extent to which the measurement correlates with an external criterion of the phenomenon under study; ideally, a GOLD STANDARD.

D

DATA COLLECTION LEVEL
Definition: The specific setting that the indicator is designed to measure/monitor, e.g., national, sub-national, facility, or community.
DATA COLLECTION METHOD
Definition: Description of all methods used for data collection. This description should include, when applicable, the sample frame used, the questions used to collect the data, the type of interview, the dates/duration of fieldwork, the sample size, and the response rate.

DATA CUSTODIAN

Definition: Data custodians are agencies responsible for managing the use, disclosure and protection of source data used in a statistical data integration project. Data custodians collect and hold information on behalf of a data provider (defined as an individual, household, business, or other organisation which supplies data either for statistical or administrative purposes). The role of data custodians may also extend to producing source data, in addition to their role as a holder of datasets.

DATA INFORMATION PYRAMID

Definition: A schematic way of looking at the number of data items to be collected at each level of the health system allowing each level to gather data of importance and relevance to their daily work while avoiding excessive data collection where no action is taken. The pyramid illustrates how most data are collected at the base of the pyramid in the health facility, where most health service action takes place. Data are processed, filtered, and streamlined as data sets that are then passed up the health system.

DATA INPUTS

Definition: All numerical inputs to mathematical or statistical models that are used to generate global health estimates. Model inputs may include raw health data, processed health data, covariates, and other parameters.

DATA LIFE CYCLE

Definition: The main steps of the data life cycle include data collection, entry and recording, storage, processing and analysis, presentation and visualization, interpretation, sharing and dissemination, retention and archiving, maintenance and quality assurance, and disposal.

DATA PROVIDER

Definition: Data providers consist of the individuals and organizations who are responsible, whether formally or informally, for making data accessible to others. Sometimes a data provider may be simply the producer of those data. In other cases, data may be deposited in a repository, centre, or archive that has the responsibility of disseminating the data.

DATA QUALITY

Definition: Data quality is a set of standards that data should reach to be usable. Quality data must encompass the following characteristics: Relevance, Credibility, Accuracy, Timeliness, Punctuality, Methodological soundness, Coherence and Accessibility.

DATA QUALITY ASSESSMENT

Definition: Is the analysis or evaluation of data to determine its accuracy, completeness, consistency, and other quality attributes against predefined criteria. The primary goal of DQ assessment is to evaluate the quality of data that already exists within a system, database, or dataset. It helps identify issues and areas for improvement in the current state of data quality.

DATA QUALITY ASSURANCE (DQA)

Definition: is a proactive and systematic process that focuses on preventing errors and ensuring that data meets predefined quality standards throughout its lifecycle. The main goal of DQA is to guarantee the overall quality of data from the point of its creation or acquisition to its eventual use.

DATA SOURCE

Definition: Description of all actual and recommended sources of data. This description should include, when applicable, any changes of the data source over time, details of denominator (if from a different source) and any other relevant information related to the origin of the source or indicator. Similar details should be given for administrative sources. Definition of primary, secondary, and preferred data sources are provided separately.

DATA TRIANGULATION

Definition: The analysis of data from three or more sources obtained by different methods. Findings can be corroborated, and the weakness or bias of any of the methods or data sources can be compensated for by the strengths of another, thereby increasing the validity and reliability of the results.

DEATH REGISTRATION DATA

Definition: Data collected during death registration, including cause of death, sex, occupation, etc.

DEATH REGISTRATION FACILITY

Definition: Registration desk that receives death notification, validates information, and enters it into central registration IT system

DEMOGRAPHIC SURVEILLANCE SYSTEM

Definition: Demographic surveillance systems are longitudinal data collection platforms that track births, deaths, migrations, and socioeconomic and health circumstances over time in established geographic areas.

DENOMINATOR

Definition: Total population of interest in a specified population. The lower portion of a fraction used to calculate a rate or ratio. In a rate, the denominator is usually the population (or population experience, as in person-years, etc.) at risk.

DESCRIPTIVE ANALYSIS

Definition: is the process of using and analyzing summary statistics that quantitatively describe or summarize features from a specific population.

DIGITAL HEALTH

Definition: Digital health is the systematic application of information and communication technologies, computer science, and data, to support informed decision-making by individuals, the health workforce and health systems, in order to strengthen resilience to disease and improve health and wellness.

DISAGGREGATION

Definition: The breakdown of observations to a more detailed level when finer details are required and made possible by the codes given to the primary observations. In health this often includes, sex, age, wealth quintile, education level, place of residence and occupation.

DISEASE SURVEILLANCE SYSTEM

Definition: is the infrastructure for ongoing systematic collection, analysis, and interpretation of outcome specific disease data for use in planning, implementing, and evaluating public health policies and practices. A communicable disease surveillance system serves two key functions; early warning of potential threats to public health and programme monitoring functions which may be disease specific or multi-disease in nature.

DOMAIN

Definition: Categorization of an indicators-based factors such as health status, risk factors, service coverage and health systems. The list includes a selection of priority indicators relating to 4 domains that includes health status, risk factors, service coverage and health systems.

E

ECOLOGICAL ANALYSIS

Definition: analysis of the relationship between a health indicator and a health determinant or exposure at a population level.

ECOLOGICAL FALLACY

Definition: an erroneous inference that may occur because an association observed between variables on an aggregate level does not necessarily represent or reflect the association that exists at an individual level; a causal relationship that exists on a group level or among groups may not exist among the group individuals.

EPIDEMIC INTELLIGENCE

Definition: System to detect, verify, investigate, and respond to early warning signals.

ESTIMATION METHOD

Definition: The estimation method is the steps taken to generate estimates. Explanation of how the indicator is calculated, including mathematical formulas and descriptive information of computations made on the source data to produce the indicator (including adjustments and weighting). This explanation should also highlight cases in which mixed sources are used or where the calculation has changed over time (i.e., discontinuities in the series). [synonyms: method of computation]

EVALUATION

Definition: A process that attempts to determine as systematically and objectively as possible the relevance, effectiveness, and impact of activities in the light of their objectives.

EVENT BASED SURVEILLANCE

Definition: Event-based surveillance is the organized and rapid capture of information about events that are a potential risk to public health. This information can be rumours and other ad-hoc reports transmitted through formal channels (i.e., established routine reporting systems) and informal channels (i.e., media, health workers and nongovernmental organizations reports).

EXTERNAL CONSISTENCY OF DATA

Definition: An assessment of the level of agreement between two sources of data measuring the same health indicator. The two sources of data that are usually compared are data flowing through the HMIS or the programme-specific information system and data from a periodic population-based survey.

EXTERNAL RESPONSIVENESS

Definition: Reflects the extent to which change in a measure relates to corresponding change in a reference measure of clinical or health status.

F

FEASIBLE

Definition: refers to the availability of data to measure the indicator. If the data is available from existing health data at a reasonable cost and/or will not add additional burden to collect data.

FOCAL POINT

Definition: The designated person to respond on behalf of a Member State or technical group during data collection or country consultation process.

G

GENDER

Definition: Gender refers to the characteristics of women, men, girls, and boys that are socially constructed. This includes norms, behaviors and roles associated with being a woman, man, girl, or boy, as well as relationships with each other. As a social construct, gender varies from society to society and can change over time.

GEOCODED

Definition: Geocoding is the process of transforming a description of a location-such as a pair of coordinates, an address, or the name of a place-to a location on the earth's surface.

GEOGRAPHIC INFORMATION SYSTEMS (GIS)

Definition: A geographic information system (GIS) is a system that creates, manages, analyzes, and maps all types of data.

GEOSPATIAL ANALYSIS

Definition: The use of geospatial data and statistical techniques to uncover patterns, relationships, and trends.

GEOSPATIAL DATA

Definition: data about objects, events or other features that have a location on the surface of the earth.

GLOBAL DATABASE

Definition: For health and health-related indicators, the Global Health Observatory (GHO) is the preferred global database, presenting the latest available data at global, regional, and national level. The GHO data repository is WHO's gateway to health-related statistics, providing access to over 1000 indicators on priority health topics including mortality and burden of diseases.

GLOBAL HEALTH ESTIMATES (GHE)

Definition: Are WHO's estimates on death and disability globally, by region and country, available by age, sex, and cause. These provide key insights on mortality and morbidity trends and are a powerful tool to support informed decision-making on health policy and resource allocation.

GOLD STANDARD

Definition: A reference measure or criterion against which other assessments are compared.

GRANULARITY

Definition: Granularity is the level of detail of the data.

H

HEALTH CARE ADMINISTRATIVE DATA

Definition: Information primarily collected for the purpose of record-keeping, which is subsequently used to produce statistics. These data are generated at every encounter with the health care system, whether through a visit to a clinic, a diagnostic procedure, an admission to a hospital or receipt of a prescription.

HEALTH ESTIMATES

Quantitative population-level estimates (including global, regional, national, or subnational estimates) of health indicators, including indicators of health status such as estimates of total and cause-specific mortality, incidence and prevalence of diseases, injuries, and disability and functioning; and indicators of health determinants, including health behaviours and health exposures.

HEALTH FACILITY CENSUS

Definition: periodic enumeration of all public and private healthcare facilities within a country about the facilities and the services they provide.

HEALTH FACILITY SURVEY

Definition: periodic enumeration of a representative sample of public and private healthcare facilities within a country about the facilities and the services they provide.

HEALTH IMPACT ASSESSMENT

Definition: Health impact assessment is a combination of procedures, methods, and tools by which a policy, programme, product, or service may be judged concerning its effects on the health of the population and the distribution of those effects within the population.

HEALTH INDICATOR

Definition: A measurable quantity that can be used to describe a population's health or its determinants. Health indicators can be categorized into domains: health status (e.g., life expectancy, HIV prevalence), risk factors (e.g., childhood stunting, prevalence of smoking), service coverage (e.g., immunization coverage rate), or health systems (e.g., hospital bed density, death registration coverage).

HEALTH INEQUALITY

Definition: A measured difference in health between population subgroups. Health inequalities can be measured and monitored.

HEALTH INEQUITY

Definition: unfair, avoidable, or remediable differences in health among groups of people. In some cases, the absence of a difference between groups (that is, a situation of equality) might be considered inequitable. Health inequity is rooted in the unfair distribution of, and access to, power, wealth, and other social resources, and is linked to forms of disadvantage that are socially produced, such as poverty, discrimination, and lack of access to services or goods.

HEALTH INFORMATION SYSTEM

Definition: A system that collects data from health and other relevant sectors, analyses the data and ensures their overall quality, relevance, and timeliness, and converts the data into information for health-related decision-making. It has four key functions: (i) data generation, (ii) compilation, (iii) analysis and synthesis, and (iv) communication and use. A solid health information system will be capable of generating reliable data from hospitals, outpatients, reportable diseases registries, cancer registries and other relevant data for health.

HEALTH RECORD

Definition: Records that contain diagnoses and treatment, medications, allergies, immunizations, as well as radiology images and laboratory results. Health records contribute to tracking a patient's medical history.

HEALTH SURVEY

Definition: A survey that is designed to gather information about health (physical and mental) and health-related factors. Health surveys generally include measures of risk factors, health behaviors, and non-health determinants or correlates of health such as socioeconomic status. The range of measures that can be included is wide and varies by survey. Age, sex/gender, and race/ethnicity are the basic demographic variables that are included in health surveys. Socioeconomic determinants of health include education, income, geographic region, and urbanicity of residence.

HEAPIING OF DATA

Definition: A measure of the of value falling on specific values such as age of disease diagnosis or date of other events, rounded up or down to the nearest integer, (e.g., Birthweight on 2000g or 2500g) or rounded (i.e., ending in "00" or "50").

I

IMPACT INDICATOR

Definition: Measures long-term outcomes that programmes are designed to affect, including decreases in mortality and morbidity.

IMPUTATION

Definition: Data imputation is a method for retaining the majority of the dataset's data and information by substituting missing data with a different value.

INCIDENCE RATE

Definition: A new event or case of a disease (or a death or other health condition) that occurred in a specified time period.

INDICATOR-BASED SURVEILLANCE

Definition: Routine reporting of cases of disease, including notifiable disease surveillance systems, sentinel surveillance, and laboratory-based surveillance. Indicator-based surveillance commonly comes from health care facilities and can be regularly reported.

INDICATOR CLASSIFICATION

Definition: The level of measurement provided by the indicator; this can be one of five levels starting with input (lowest) and moving through process, output, outcome and finally impact (highest)

INDICATOR DEFINITION

Definition: How the indicator is measured, including numerators, denominators, data type and disaggregation in common use. The indicator definition should be unambiguous and be expressed in universally applicable terms.

INPUT INDICATOR

Definition: Measures human and financial resources, physical facilities, equipment, and operational policies that enable program activities to be implemented. This includes health financing, health workforce, health infrastructure, and health information and governance.

INTEGRITY OF DATA

Definition: Data have integrity when the system used to generate them is protected from deliberate bias or manipulation for political or personal reasons.

INTERACTION

Definition: when the relationship between two variables depends on the value of another variable (also referred to as effect modification).

INTERNAL CONSISTENCY OF DATA

Definition: An assessment of the level of agreement between two variables in the same source documents or document flow. Typically, this involves assessing reporting accuracy of selected indicators through the review of source documents in health facilities and district offices. This element of internal consistency is measured by a data verification exercise. (*e.g., Answers the question: Is the observed relationship between the indicators, as reflected in the reported data, that which we would expect?*).

INTERNAL RESPONSIVENESS

Definition: Characterises the ability of a measure to change over a prespecified time frame.

INTERNATIONAL CLASSIFICATION OF DISEASES (ICD)

Definition: WHO's International Statistical Classification of Diseases and Related Health Problems (ICD) is a structured translation of each medical condition into an alphanumeric code, which allows for the harmonization and comparison of mortality statistics across time and location. ICD is a gold standard coding system for reporting cause of death data.

INTERNATIONAL HEALTH REGULATIONS (IHR)

Definition: The IHR is an instrument of international law that is legally-binding on 196 countries, including the 194 WHO Member States. The IHR grew out of the response to deadly epidemics that once overran Europe. They create rights and obligations for countries, including the requirement to report public health events.

INTERNATIONAL STANDARDS

Definition: Set of standards defined and agreed at international level on ethics, design, implementation, confidentiality, analysis and dissemination of information, surveys, services, etc.

INTEROPERABILITY

The ability of different applications to access, exchange, integrate and use data in a coordinated manner through the use of shared application interfaces and standards, within and across organizational, regional, and national boundaries, to provide timely and seamless portability of information and optimize health outcomes.

J

JOINT EXTERNAL EVALUATION (JEE)

Definition: Voluntary, collaborative, multisectoral process to assess country capacities to prevent, detect and rapidly respond to public health risks.

K

KAPPA STATISTIC (K)

Definition: Measures of agreement between categorical measures.

L

LINKAGE

Definition: Linkage is the process of combining information from different sources or datasets.

M

MASTER FACILITY LIST (MLF)

Definition: The unique, complete, up-to-date, and uniquely coded list of all the active and prior health facilities in the country officially that are officially curated by the mandated agency. At a minimum, the HFML includes a unique ID, location, type, and name of each facility.

MATHEMATICAL OR STATISTICAL MODELS

Definition: A statistical model is a mathematical model that embodies a set of statistical assumptions.

MEAN

Definition: The average of a set of values.

MEASURE

Definition: refers to the procedure of applying a reference scale to a variable or set of variables.

MEASUREMENT

Definition: refers to the extent, dimension, quantity, etc. of an attribute.

MEASUREMENT LEVEL
Definition: The specific setting that the indicator is designed to measure/monitor; e.g., global, national, sub-national, facility, household, community, patient level.
MEASUREMENT METHOD
Definition: How the data from data sources are used; this can process or other types of analyses that make use of the indicator.
MEDIAN
Definition: middle point of a set of ordered numbers; half of the values are higher than the median, and half of the values are lower.
MEDICAL CERTIFICATION OF CAUSE OF DEATH (MCCD)
Definition: Medical certification of cause of death involves confirmation of death, external examination of the body and ascertainment of the circumstances and cause of death.
METADATA
Definition: Data that define or describe other data. They are the information needed to explain and understand the data or values being presented.
METHOD OF AGGREGATE ESTIMATION
Description of the methodology, including any mathematical formulas, used for the calculation of the regional/global aggregates from the country values.
METHODOLOGICAL SOUNDNESS
Definition: The application of the available international standards, guidelines, and good practices in the production of data.
MICRODATA
Definition: Granular data that may include individual level information (anonymized or not)
MINISTRY OF HEALTH (MoH)
Definition: The government institution that is in charge of all aspects of health. This may have different names at country level, but the main function is to safeguard the health of the population.
MONITORING
Definition: The systematic and routine collection of information to assess performance and progress towards specific targets and over an established period of time.
MONITORING & EVALUATION (M&E) FRAMEWORK
Definition: Necessary structure to support analysis and perform monitoring and evaluation activities. A framework can establish and maintain a set of global and country indicators to support strategic thinking, operational tracking, real-time.
MORBIDITY DATA
Definition: Is the information registered on the state of being symptomatic or unhealthy due to a disease or health-related condition.
MORTALITY CODER
Definition: Mortality coder is the trained person that registers medical conditions and events reported in MCCD forms to determine the underlying cause of death and assign mortality codes using ICD rules and principles.

MORTALITY DATA

Definition: Is the information registered when a death occurs.

N

NATIONAL HEALTH STRATEGIC PLAN

Definition: A national health strategic plan is the set of priorities to achieve key milestones that will have impact beyond the health sector. There will be medium-term and long-term expected outcomes and concrete and realistic allocation of resources to implement the activities within a clear timing. The national health strategic plan will concretize priorities; keep focus on medium and long term without deviating from an optimal path; integrate the health sector; help focus the policy dialogue on health priorities and guide operational planning, resource allocation and health sector monitoring and evaluation.

NATIONAL STATISTICS OFFICE (NSO)

Definition: Government agency or institution responsible for collecting, analyzing, using, and disseminating statistical data related to a country.

NATIONALLY REPRESENTATIVE

Definition: A survey that will use design methods and standardized criteria scalable to national context using a sub-sample that represents the target population in terms of age, sex, urban/rural and other categories of interest.

NEGATIVE PREDICTIVE VALUE (NPV)

Definition: Negative predictive value is the proportion of the cases giving negative test results who are already healthy. It is the ratio of subjects truly diagnosed as negative to all those who had negative test results (including patients who were incorrectly diagnosed as healthy). This characteristic can predict how likely it is for someone to truly be healthy, in case of a negative test result.

NOTIFIABLE CONDITIONS

Definition: A disease that, when diagnosed, requires health providers (usually by law) to report to state or local public health officials. Notifiable diseases are of public interest by reason of their contagiousness, severity, or frequency.

NUMERATOR

Count of values captured by the indicator in a specified population. The upper portion of a fraction used to calculate a rate or ratio.

O

ODDS

The numerator is the proportion of the event of interest, and the denominator is the proportion of the non-event. The numerator and denominator are thus complementary proportions ($p/1-p$).

OUTCOME INDICATOR

Definition: Measures whether the program is achieving the expected effects/changes in the short, intermediate, and long term. Some programs refer to their longest-term/most distal outcome indicators as impact indicators. This usually includes coverage of interventions and risk factors and behaviours.

OUTPUT INDICATOR

Definition: Measures the results of the processes in terms of service access, availability, quality and safety and health security.

P

PERCENTAGE

Definition: Number or ratio that can be expressed as a fraction of 100.

PERIODICITY

Definition: Data can be compiled continuously in systems such as civil registries, cancer registries and surveillance systems of reportable diseases. Data can also be compiled periodically, which is to say at regular intervals or without predefined periodicity, and at a particular point in time.

POPULATION BASED SURVEY

Definition: descriptive cross-sectional epidemiological study that is useful for calculating the prevalence of self-reported events or events measured during the investigation, generally employing a representative sample from the population of interest.

POPULATION CENSUS

Definition: A population census is the total process of planning, collecting, compiling, evaluating, disseminating, and analyzing demographic, economic and social data at the smallest geographic level pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country.

POSITIVE PREDICTIVE VALUE (PPV)

Definition: Positive predictive value is the proportion of cases giving positive test results who are already patient. It is the ratio of patients truly diagnosed as positive to all those who had positive test results (including healthy subjects who were incorrectly diagnosed as patient). This characteristic can predict how likely it is for someone to truly be patient, in case of a positive test result.

POST ENUMERATION SURVEY (PES)

Definition: The purpose of the Post-Enumeration Survey is to measure the accuracy of the census by independently surveying a sample of the population. The survey estimates the proportion of people and housing units potentially missed or counted erroneously in the census.

PREDICTIVE VALIDITY

Definition: the degree to which predictions are confirmed by facts expressed in terms of its ability to predict future outcomes or events. Predictive validity is often used for impact indicators.

PREFERRED DATA SOURCES

Definition: Recommended sources of data. This includes civil registration and vital statistics system, national population-based surveys; routine facility information systems; health facility assessments; • administrative data sources; human resources information systems; key informant surveys; indicators from other sources, including modelling.

PREVALENCE RATE

Definition: number of existing cases of a disease or other health event divided by the number of persons in population at that specified time. Each individual is observed on a single occasion, at which time the individual's status with respect of the event in question is ascertained.

PRIMARY DATA

Definition: Primary data comes from country health information systems (including administrative reporting, household surveys, etc.). Data is reported as is, or with modest adjustment.

PRIMARY DATA SOURCES
Definition: Primary source data provide direct evidence about an event. Data collection can take different forms, whether through a population census, national or local research (typically, sample-based, or non-sample-based surveys). The creation of an information system to achieve specific objectives will generate primary data. In other words, indicators are said to be based on primary data if the data source was created to achieve a specific purpose.
PROCESS INDICATOR
Definition: Measures the program's activities and outputs (direct products/deliverables of the activities). Together, measures of activities and outputs indicate whether the program is being implemented as planned. (e.g. health workforce training, constructing a health facility, the process of registering births and deaths.)
PROCESS OF VALIDATION
Definition: A description of how the indicator was validated to assess how accurately it measures what it is intended to do. This glossary includes some methods for validation.
PROCESSED HEALTH DATA
Definition: Health statistics that have been calculated from raw health data, but which are not the result of synthesizing multiple data sources. Examples of processing raw health data include cleaning data by removing implausible values, calculating an indicator with an algorithm, or adjusting a statistic for bias.
PROPORTION
Definition: The size, number, or amount of one thing or group as compared to the size, number, or amount of another. When the numerator is a subset of the denominator. A proportion tends to be expressed as a percentage (%). It is the observed relative frequency of an event and provides an estimate of probability. It should be noted that, according to the frequentist approach, the probability of an event occurrence is given by the relative frequency of the event over the long term (in infinite attempts or repetitions of the experiment).
PROXY HEALTH INDICATOR
Definition: an indicator that stands in for another indicator or topic that is difficult to measure or for which data are limited.
PUBLIC HEALTH SURVEILLANCE SYSTEM
Definition: Is the system that systematically collects, analyzes, and interprets health-related data essential to planning, implementation and evaluation of public health practice.
PUBLICLY AVAILABLE
Definition: information in any form that is generally accessible, without restriction, to the public.
PUNCTUALITY
Definition: The time lag between the release date of data and the target date on which they were scheduled for release as announced in an official release calendar.
P VALUE
Definition: The probability that a test statistic would be as extreme as or more extreme than observed if the null hypothesis were true.

R

RATE

Definition: A rate is an expression of the frequency with which an event occurs in a defined population, usually in a specified period of time. The components of a rate are the numerator, the denominator, the specified time in which events occur, and usually a multiplier, a power of 10, that converts the rate from a fraction or decimal to a whole number. The numerator is the absolute number of occurrences of the event being studied in a specified time. The denominator is the reference population (or population being studied) at the same time.

RATIO

Definition: The result of dividing one quantity by another without regard for details such as a time dimension. Rates, proportions, and percentages are all types of ratios. The distinction between a proportion and a ratio is that, whereas the numerator of a proportion is included in the population defined by the denominator, this is not necessarily so for a ratio, which expresses the relationship of two separate and distinct quantities, neither of which is included in the other.

RATIONALE

Definition: Importance of the indicator for public health response.

RAW HEALTH DATA

Definition: Measurements derived from primary data collection with no adjustments or corrections.

RECORD LINKAGE

Definition: The methodology of bringing together corresponding records from two or more files or finding duplicates within files.

REGISTRAR

Definition: The local civil registrar is the official authorized by law to register the occurrence of vital events and to represent the legal authority of government in the field of civil registration.

REGISTRATION FORM

Definition: The registration form is the paper or electronic format that is used to register a vital event.

REGRESSION

Definition: a statistical technique that relates a dependent variable to one or more independent (explanatory) variables.

REGULAR ASSESSMENT

Definition: Frequent evaluation that may be done in established periods (weekly, monthly, yearly, etc.).

RELEVANCE

Definition: The degree to which the data meet the user needs. indicators must provide information that is appropriate and useful for guiding policies and programmes as well as for decision-making.

RELIABILITY OF DATA

Definition: The degree to which the results obtained by a measurement/procedure can be replicated. Consistency of the data when collected repeatedly using the same procedures and under the same circumstances. [synonym as replicability]

REPORTING SITE

Definition: Health facilities designated by authorities to mandatory report cases of diseases. If available, public, and private reporting sites should be identified.

REPRESENTATIVE

Definition: the ability for the indicator to accurately describe the occurrence of a health-related event over time and its distribution in the population by place and person. This involves the absence of selection bias with respect to the population that the indicator is intended to represent.

RESPONSE RATE

Definition: In survey research, response rate, also known as completion rate or return rate, is the number of people who answered the survey divided by the number of people in the sample.

RESPONSIVENESS

Definition: Is the indicator's ability to detect changes over time in response to interventions, treatments, or natural progression of the condition. [consult External responsiveness and Internal responsiveness]

ROUTINE HEALTH INFORMATION SYSTEM (RHIS)

Definition: Systems that record data generated at public and private health facilities, institutions and community-level, healthcare posts and clinics, use these data for analysis, monitoring and reporting at regular intervals. The data give a picture of health services, health interventions and health resources. Most of the data are gathered by healthcare providers as they go about their work. The sources of those data are generally records from these institutions.

S

SAMPLE

Definition: Sample is a subset of a population.

SAMPLE SIZE

Definition: the number of people upon which a disaggregated (subgroup) estimate is based; that is, the denominator used to calculate a disaggregated estimate.

SAMPLING ERROR

Definition: The difference between the sample statistics and population parameter. Sampling errors can be estimated by general methods such as bootstrapping or specific methods incorporating assumptions from the true population distribution.

SECONDARY DATA SOURCES

Definition: data that was originally collected for other purposes. The data from these existing sources are considered secondary. Although these sources were not created for the purpose at hand, they facilitate the development of the required indicators. Data from a census, research, information system, etc. are secondary source data.

SENSITIVITY ANALYSIS

Definition: Sensitivity analysis is a systematic approach to evaluate how the variation in the output of a system or model can be attributed to different sources of variation in its inputs. It involves examining the sensitivity of the model's outcomes or outputs to changes in individual input parameters, providing insights into the relative importance of each parameter in influencing the overall results.

SEX

Definition: Biological sex (male/female).

STAKEHOLDER

Definition: Interested parties, group or organization who may affect, be affected by, or perceive itself to be affected by a decision, activity or outcome of a project or programme.

STANDARD OPERATING PROCEDURES (SOP)
Definition: Procedures that are documented to guarantee that these are standardized and followed.
STATISTICAL DATA
Definition: Data that has been organized for analysis, interpretation, and representation through statistical analysis.
STATISTICAL METHODS
Definition: Various techniques employed in data analysis for validation studies.
STATISTICAL OUTPUT
Definition: numerical data relating to an aggregate of individuals or entities.
STATISTICAL SIGNIFICANCE
Definition: mathematical measure of the probability that a result is likely due to chance or another factor (that is, the null hypothesis).
STRATIFICATION
The process of sorting data into defined segments or groups. This method can be used when sampling a population for conducting a survey or this can be used for analysis to control for confounding.
STRUCTURAL VALIDITY
Definition: The degree to which scores of an Indicator are an adequate reflection of the dimensionality of the construct to be measured and involves examining the construct's underlying structure or dimensionality. A measurement instrument with strong structural validity should demonstrate that its items align logically with the theoretical framework of the construct, ensuring that it effectively captures the intended concepts or traits.
SUB-NATIONAL
Definition: County, district, state levels.
SURVEILLANCE
Definition: The continuous, systematic collection, analysis, interpretation, and dissemination of data needed for the planning, implementation, and evaluation of public health actions. Some examples are public health surveillance system, indicator-based surveillance, disease surveillance system, demographic surveillance system, etc.
SURVEY
Definition: Survey is a structured list of questions that collect data on a specific population.
SURVEY DATA
Definition: A survey is an investigation about the characteristics of a given population by means of collecting data from a sample of that population and estimating their characteristics through the systematic use of statistical methodology.
SUSTAINABILITY
Definition: the source's potential to remain relevant and be of the quality needed to generate information over time. This depends not only on the periodicity of the data collection, but on the availability of the financial resources needed to sustain that source of data, the presence of a legal framework, political will, among other factors.

SUSTAINABLE DEVELOPMENT GOAL (SDG)

Definition: The Sustainable Development Goals (SDGs) are 17 global objectives that were agreed by Members of the United Nations and aim to transform our world. They are a call to action to end poverty and inequality, protect the planet, and ensure that all people enjoy health, justice, and prosperity. It is critical that no one is left behind.

T

TARGET POPULATION

Definition: A set of elements about which information is wanted and estimates are required.

TIMELINESS

Definition: When the data is quickly available and accessible for use. In the context of data quality, the degree to which reports are submitted on time according to established deadlines. Timeliness involves the availability and reliability of the data at the time it is needed to construct the indicators. Thus, timely produced indicators provide better opportunities for making health-related decisions.

TRACER INDICATOR

Definition: a highly specified indicator chosen as an example to represent a broader health topic.

TREND

Definition: Multiple standardized measurements in time demonstrating how the values have increased, remained the same or decreased.

U

UNCERTAINTY MEASURE

Definition: Measures that indicate the level of certainty around a point estimate and quantify the imprecision. Common measures of certainty include confidence interval (CI), standard deviation (SD), credible interval (CrI), uncertainty interval (UI), standard error, etc.

UNDERCOUNT RATE

Definition: Net undercount of the census is the difference between the number of persons counted in the census and the number of people who should have been counted.

UNDERSTANDABLE/SIMPLICITY

Definition: if the indicator is presented in a clear, concise, and easily comprehensible way. Whether the indicator is presented in a clear, concise, and easily comprehensible way. The indicator must be understood by those responsible for taking action, and, specifically, by those responsible for decision-making. [synonyms comprehensible]

UNIT OF MEASURE

Definition: Unit of measure in which the indicator is presented; e.g., deaths per 1000 live births; US\$; liters per person per year.

Note: Percentage is not considered a unit of measure and indicators that are presented as percentages should have the unit of measure field filled in as "N/A" (not applicable)

USEFULNESS/UTILITY

Definition: if the indicator is useful for program improvement and policy issues.

V

VALIDITY

Definition: Ability of an indicator to measure what it is intended to measure (absence of distortions, bias, or systematic errors). The most relevant biases are those related to selection of the study population and the quality of the information compiled. The data source should include the variables needed to develop the indicator.

VERBAL AUTOPSY

Definition: Verbal autopsy (VA) is a method used to determine the cause of death through interviews with the deceased person's next of kin or caregivers. These interviews involve a standardized questionnaire to gather details on symptoms, medical history, and the circumstances leading to death. Healthcare professionals or algorithms then analyze this information to identify the likely cause of death. The primary goal of verbal autopsy is to describe the causes of death at the community level or population level in areas where there is no medical certification of deaths, or it is not yet well-established.

VITAL EVENT

Definition: Some vital events that are captured through civil registration and vital statistic systems are: birth, adoption, marriage, divorce, migration, death.

VITAL STATISTICS

Definition: The systematic record of vital events such as birth, marriage, divorce, adoption, death, and cause of death to generate data and statistics.

W

WEIGHTING

Definition: Statistics weighting is a technique used to adjust data to reflect the known population profile. It is used to balance out any significant variance between actual and target profile. Weighting is commonly used. Weighted analyses of population-based surveys allow us to generalize findings to a larger or more general population. This approach aims to provide unbiased estimates of descriptive statistics or model parameters of the population of interest, which may be a general population or major population subgroup. Incorporating weights in the analyses can be crucial to achieve statistically valid, representative population-based findings in surveys to make adjustments for sampling errors.