HOW TO COUNT ILLNESS?

Basic epidemiological concepts for understanding the COVID-19 epidemic

May 2020

The Association of Schools of Public Health in the European Region (ASPHER)

TABLE OF CONTENTS

FOREWORD

There are over a hundred definitions of epidemiology. The one I use is "the study of disease in populations". It's simple and easy to remember…. Epidemiologists will probably question if it's right…

There has never been a greater interest in epidemiology than right now in the COVID-19 pandemic. There are have-a-go epidemiologists from all walks of life – people who use numbers for a living – mathematicians, statisticians, geographers, philosophers computer programmers, even accountants and quantity surveyors can be found showing their insights on the twitter sphere. There is some brilliant stuff out there, and new ways of presenting data hopefully giving us all new knowledge to keep people safe and stop the spread of this terrible virus. Our major newspapers have built up extensive repositories of data often shared for free, sometimes ahead of academic institutions and national governments. And in the common parlance, who would have imagined three months ago we would all be talking "epidemiology", "R_0", "R_t", "prevalence", "incidence", "predictive value" and many more terms. But we must also encourage our politicians and public to get beyond a superficial understanding of the terms they are using and recognise some of the pitfalls, misconceptions and potential errors inherent in what we do.

It is necessary for us all to understand what we mean by these terms. Colleagues in the Association of Schools of Public Health in the European Region (ASPHER) – the oldest Association of Public Health – represent the great teaching engines of public health in Europe and beyond. This rapidly constructed compendium will hopefully help journalists, business consultants, other stakeholders and also members of the general public to develop their knowledge and expand the power of citizen science. We are all citizens of the world now, and we must all play our part in controlling and preventing the further spread of this pandemic.

I commend this glossary to epidemiology, translated in five languages, to you all.

John Middleton
President ASPHER

1. Numbers, proportions, ratios and rates

Standard definitions:

ABSOLUTE NUMBERS: Quantification of a phenomenon not dependent on other figures (i.e. mere counting).

RELATIVE NUMBERS: Values which are dependent on other figures or numbers.

PROPORTIONS: A type of fraction in which the numerator is included in the denominator. A proportion's values range from 0 to 1, and it can be expressed in decimals or percentage (0% to 100%).

RATIOS: A fraction in which the numerator is not included in the denominator.

RATE: A measure of the frequency of occurrence of a phenomenon in a defined population, in a given period. The components of a rate are the numerator (i.e. number of cases), the denominator (reflecting the defined population – explicit or implicit place, region, or country – and the specified time-frame in which the events occurred), and usually a multiplier (as 100, 1 000, 100 000 etc.).

$$Rate = \frac{Number\ of\ events\ in\ specified\ period}{Person-time\ (Time\ each\ person\ was\ observed, totaled\ for\ all\ persons)}\ x\ 10^n$$

Development of the concepts and examples:

The **absolute number** of cases satisfies general administrative needs such as number of hospitalization or number of deaths. To have a clearer idea of a health phenomenon, the number of cases should be divided by the reference population. The example in Table 1 refers to notified COVID-19 cases in five countries with different populations.

Table 1. COVID-19 cases as of May 25th 2020

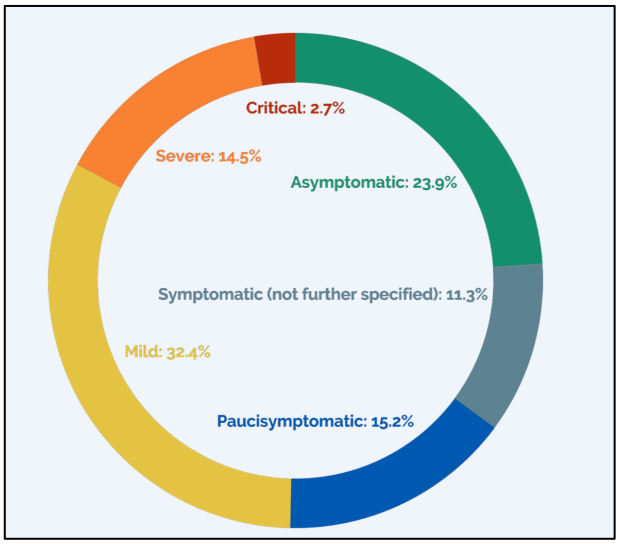
Country	Absolute number of cases	Total population (millions)	N. of cases per 100,000 population
U.S.A.	1 592 599	328 200 000	485.3
Italy	229 858	60 400 000	380.6
United Kingdom	259 563	66 600 000	389.7
Iceland	1 804	360 000	501.1
Andorra	763	77 000	991

(Source: https://www.who.int/ Retrieved on 25 May 2020)

An example of a **ratio** is the male to female ratio of mortality for COVID-19.In Italy this is 3:2 according to data available on May 21th. (Epicentro, Istituto Superiore di Sanità)

The **proportion** of asymptomatic cases of SARS-CoV-2 infection is the number of asymptomatic individuals with a positive test result, divided by the total number of individuals with a positive test: the numerator is included in the denominator. Figure 1 shows the proportion of Italian cases which were asymptomatic, critical, severe, mild, paucisymptomatic, and not further specified.

Figure 1. Clinical presentation of COVID-19 cases in Italy



(Source: Italian National Institute of Health (ISS); Available at epicentro.iss.it)

The **rate** introduces the variable "time". Table 2 shows the comparison of the cumulative mortality rate of six countries, which is the proportion of a population that dies over a specified time, i.e. from the start of the epidemic to mid-May 2020.

Table 2. Cumulative COVID-19 mortality rate of selected countries (as of 15th May 2020)

Country	Confirmed COVID-19 Deaths	Population (million)	COVID-19 Mortality Rate (deaths per million)
Belgium	8 843	11.42	774.20
France	27 045	66.99	403.73
Italy	31 106	60.43	514.73
Spain	27 104	46.72	580.09
Sweden	3 460	10.18	339.78
UK	33 186	66.49	499.12

(Source: https://www.statista.com/statistics/1104709/coronavirus-deaths-worldwide-per-million-inhabitants/; Retrieved on 14 May 2020)

2. Crude and adjusted epidemiological measures

Standard definition

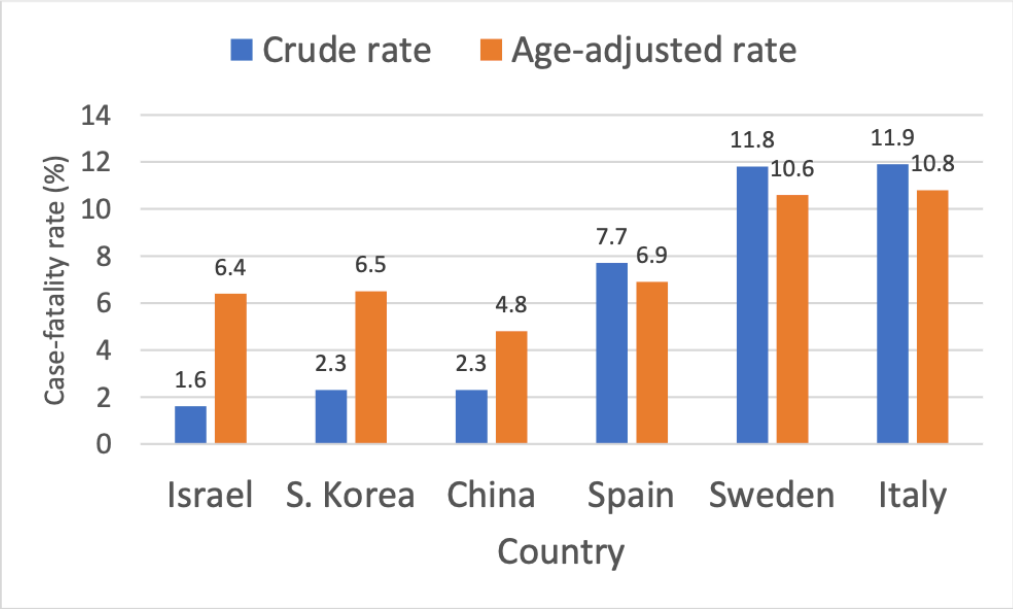
CRUDE: A crude measure consists of the "raw" data (i.e. cases divided by population), not adjusted for any factor that may interfere with the final interpretation.

ADJUSTED: The adjusted measure is standardized to take into account factors that might condition the results, and therefore distort our direct interpretation of it. We may need to adjust for age, gender, race, or any other key confounding factor.

Development of the concepts and examples:

The **crude** mortality rate (explained later) is the proportion of the number of all deaths during the year to the average population in that year. It's easy to understand that the older the population, the higher the mortality rate will be. Instead, the age-**adjusted** measures (mortality rate in the example) takes into account the differences in population age distribution. In the example in Figure 2, the difference between Israel and Spain in crude case fatality rates for COVID-19 is reduced after adjusting for age, as the population is older in Spain than in Israel.

Figure 2. Crude and age-adjusted COVID-19 case fatality rates for six countries



(Source: Green MS et al., The confounded crude case-fatality rates for COVID-19 hide more than they reveal - a comparison of age-specific and age-adjusted rates between six countries. Preprint https://doi.org/10.1101/2020.05.09.20096503)

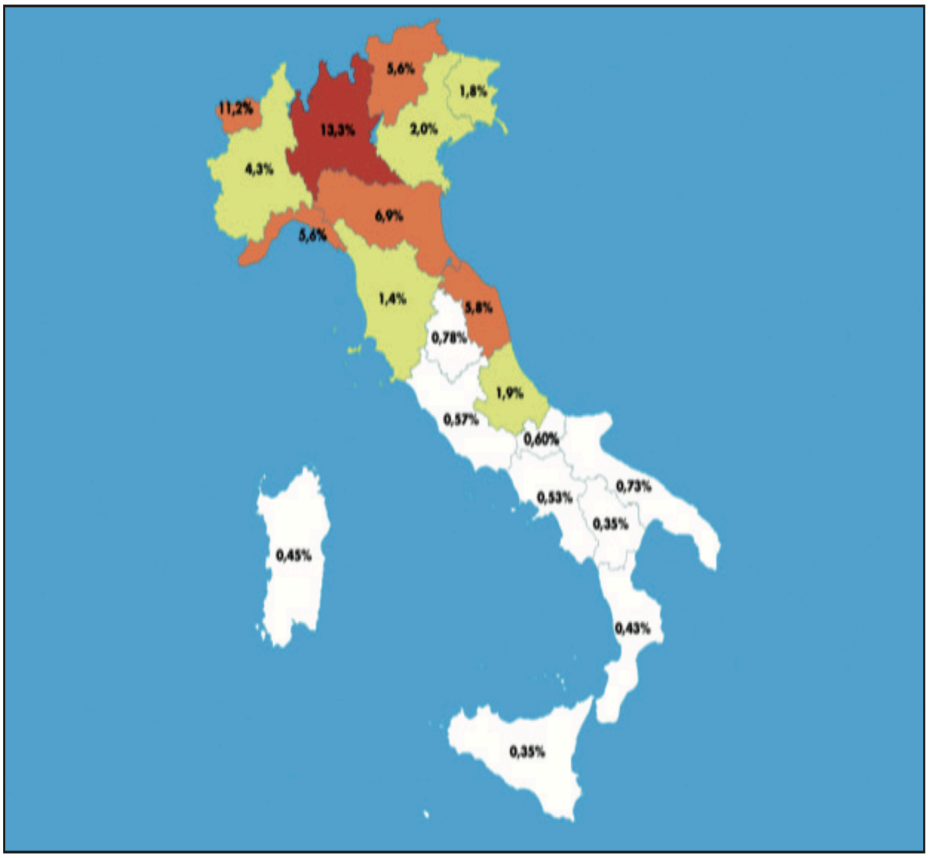
3. Point and period prevalence of a disease

Standard definition

PREVALENCE OF A DISEASE: A measure of disease occurrence: the total number of individuals who have a disease at a particular time, divided by the population at risk of having the disease at that time. It gives a snapshot of the population at a certain point in time (**point prevalence**).

PERIOD PREVALENCE OF A DISEASE: The proportion of individuals with a disease during a defined period of time. To calculate a period prevalence, the most appropriate denominator for the period must be found. Prevalence differs from incidence in that prevalence includes all cases, both new and preexisting, in the population at the specified time, whereas incidence is limited to new cases only.

Development of the concepts and examples:



Normally it makes more sense to calculate the point prevalence (at a certain time) such as the number of people affected by a disease (i.e. 5% of the EU population is affected by diabetes). In the case of an epidemic of a new disease such as COVID-19 it might make more sense to calculate the period prevalence (how many people have become infected since the beginning of the epidemic to date). Note that for non-communicable diseases the prevalence is more stable than for infectious disease where the recovery can be quick. Figure 3 shows the estimated period prevalence of COVID-19 in Italian regions, which is the prevalence of the disease estimated in the period starting from the beginning of the epidemic to date.

Figure 3. The estimated period prevalence of COVID-19 in Italy (Data update 7.04.20)

(Source: Signorelli C et al., COVID-19 in Italy: impact of containment measures and prevalence estimates of infection in the general population, Acta Biomed 2020)

4. Incidence of a disease, cumulative incidence and attack rate

Standard definition

INCIDENCE OF A DISEASE: The number of new cases of a disease occurring during a given period in a specified population. It could be measured in terms of **incidence proportion** (when the people in the numerator, those who develop disease, are all included in the denominator, i.e.: the entire population) or in terms of **incidence rate or person-time incidence** (when time is directly incorporated into the denominator, see above for the definition of *rate*).

Synonyms of incidence proportion are two very important terms in outbreaks research:

CUMULATIVE INCIDENCE: The proportion of the population at risk for a disease and that develops the disease during a specified interval of time.

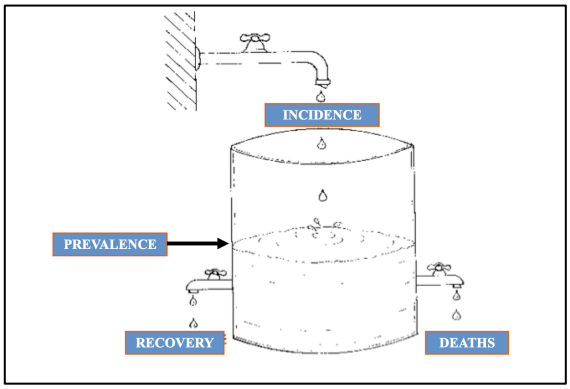
ATTACK RATE: The proportion of a group that experiences the outcome under study over a given period, generally very short (e.g., the incubation period during an outbreak).

Development of the concepts and examples:

Normally, incidence is calculated per year per 1 000 or 100 000 population depending on the frequency of the disease. In the case of an epidemic of a new disease such as COVID-19 it makes more sense, at least initially, to present data considering the cumulative incidence.

The underlying concepts representing incidence and prevalence are interrelated. Prevalence measures how much of a disease or a condition is spread in a population at a given time, and is a function of the incidence (the rate of occurrence of new cases) and the average duration of the condition (the length of the process or disease). Thus, incidence conveys information about the risk of contracting the disease, whereas prevalence indicates how widespread the disease is (Figure 4).

Figure 4. Relationship between incidence and prevalence



(Source: Signorelli C, Elementi di metodologia epidemiologia, Società Editrice Universo, 7th edition)

5. Case fatality rate and infection fatality rate

Standard definitions

CASE FATALITY RATE (CFR): The proportion of persons with a specific condition (e.g., a disease), i.e. cases, that die from that condition. The numerator is the number of cause-specific deaths and the denominator is the number of diagnosed cases (incident cases) of that condition. It measures the severity of the condition. These are some examples of CFR for renown diseases:

- *Rabies:* 100%
- *Pancreatic cancer*: 90%
- *Meningococcal disease*: 10%
- *Influenza*: 0.1%

CRUDE CFR: The CFR without adjustment. The formula is:

$$CFR(\%) = \frac{Number\ of\ disease\text{–}specific\ deaths\ among\ the\ incident\ cases}{Number\ of\ incident\ cases\ during\ a\ specified\ period\ of\ time} \text{x}100$$

ADJUSTED CFR: The CFR is adjusted to take into account confounding factors that might alter the results, e.g., age, under-reporting or delay from hospitalization to death. Statistical techniques are used to adjust the rates among the populations to be compared.

ESTIMATED CFR: When the total number of cases is not completely known, it can be estimated for example from the number of deaths. If there is a high number of undiagnosed cases, the CFR would be overestimated. According to the latest estimates, the crude CFR of COVID-19 varies between 1.6% and 11% (Green MS et al., 2020) while the estimated CFR varies between 0.5% and 1.1% (Russel TW, et al. 2020).

INFECTION FATALITY RATE (IFR): The proportion of persons with an infection that die from that infection. The numerator is the number of infection-specific deaths and the denominator is the number of infections. It measures the severity of the condition. The formula is:

$$IFR(\%) = \frac{Number\ of\ infection\text{–}specific\ deaths\ among\ the\ incident\ infections}{Number\ of\ incident\ infections} x100$$

It is not used very much during a pandemic, during which we only account for the diagnosed cases. It will be more useful when wide serological studies will be performed.

Development of the concepts and examples:

The CFR and IFR are not true rates, but in fact proportions, i.e. the numerator is restricted to deaths among the cases included in the denominator.

Considering the data from WHO on 25th of May of 2020, since the beginning of the epidemic, there were 5,463,392 cases worldwide and 344,533 deaths.

So, the CFR would be calculated as follows:

$$CFR = \frac{278\,892}{4\,006\,257}\ x\ 100 = 7.0\%$$

CFR is a poor indicator of mortality risk in an ongoing pandemic, since the denominator refers only to a part of the cases (those who have been diagnosed and notified) and depends on the case definition used, the testing criteria and the capacity of testing across countries, making data hard to compare.

Because nucleic acid testing is limited and currently available primarily to people with significant indications of and risk factors for COVID-19, and because a large number of infections with SARS-CoV-2 result in mild or even asymptomatic disease, the IFR is likely to be significantly lower than the CFR.

6. Recovery rate

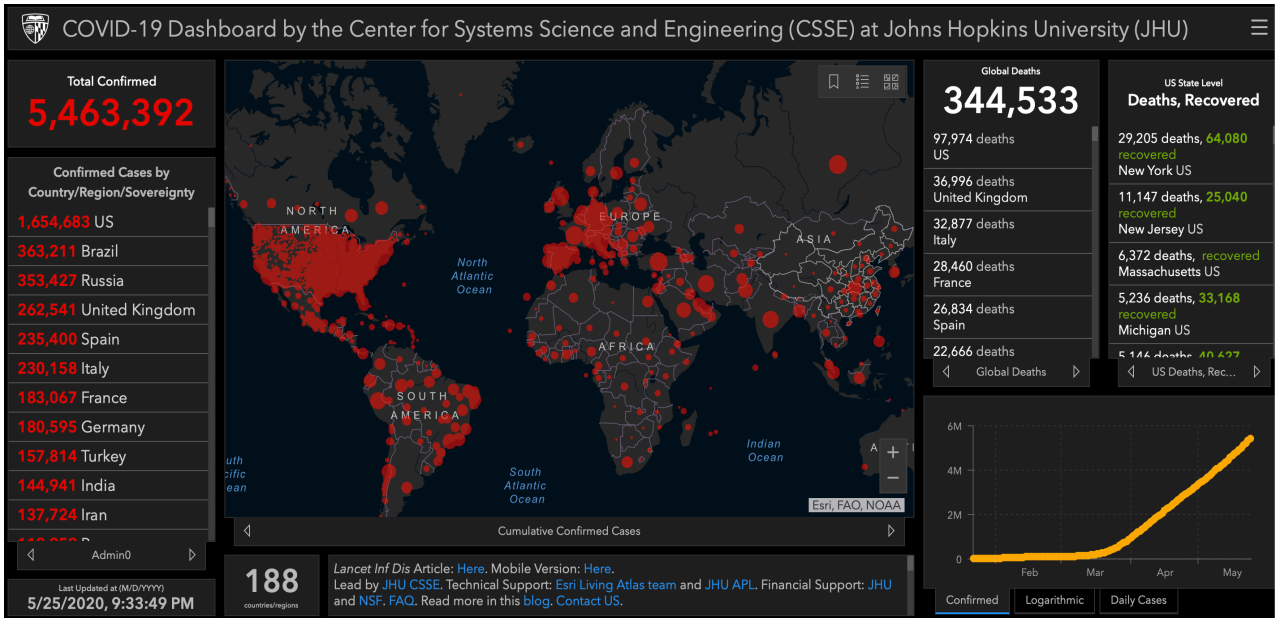
Standard definition

RECOVERY RATE: The rate of transition from the state of being infected to the state of absence of disease.

Development of the concepts and examples:

Recovery rate is one of the most frequently disseminated pieces of data during the COVID-19 epidemic compared to the number of those newly infected. In the first phase of the epidemic, the number of patients recovered was less than the new cases (recovery rate less than incidence rate), after the peak epidemic was reached, the patients who recovered exceeded new cases.

Figure 5 Johns Hopkins University dashboard of the world's situation of COVID-19 cases



(Source: https://coronavirus.jhu.edu/map.html, accessed on 25.05.20)

On the right side of the dashboard the cumulative number of deaths and of recovered cases can be found.

There is a delay in the confirmation of recovered cases that is due to two factors. First,countries have different criteria to define a case as recovered; for example, in Italy a case can be considered recovered only after there is evidence of two negative swab tests done 48 hours apart. Second, infected individuals can remain contagious and shed the virus for a relatively long time even after they have recovered from the COVID-19 clinical illness.

7. Mortality rate, cumulative death rate, excess mortality

Standard definition

MORTALITY RATE: is a measure of the number of deaths (in general, or due to a specific cause) in a particular population, in relation to the size of that population, per unit of time.

The numerator is the number of persons dying during the given time period; the denominator is usually expressed as the size of the population among which the deaths occurred (usually estimated as the midyear population).

$$\frac{Number\ of\ deaths\ during\ a\ given\ period}{Number\ of\ persons\ at\ risk\ of\ dying\ during\ the\ period} x\ 10^n$$

We may speak about **crude death rates** (total number of deaths during a given time interval divided by mid-interval population per 1,000 or 100,000), or **cause-specific death rate** (number of deaths assigned to a specific cause during a given time interval).

CUMULATIVE DEATH RATE: The proportion of a group that dies over a specified time interval. It is the incidence proportion of death.

EXCESS MORTALITY: Mortality that is above what would be expected based on the non-crisis mortality rate in the population of interest (i.e. in "normal conditions"). Excess mortality is thus mortality that is attributable to the crisis conditions.

Excess Mortality = Observed Mortality in Crisis – Expected Mortality in Non-crisis

Development of the concepts and examples:

The mortality rate of a country is the number of deaths divided by the population, usually expressed in deaths per million inhabitants. During the COVID-19 epidemic the definition *death toll* was used, especially in the US to indicate the number of people who die because of an event such as a war or an accident.

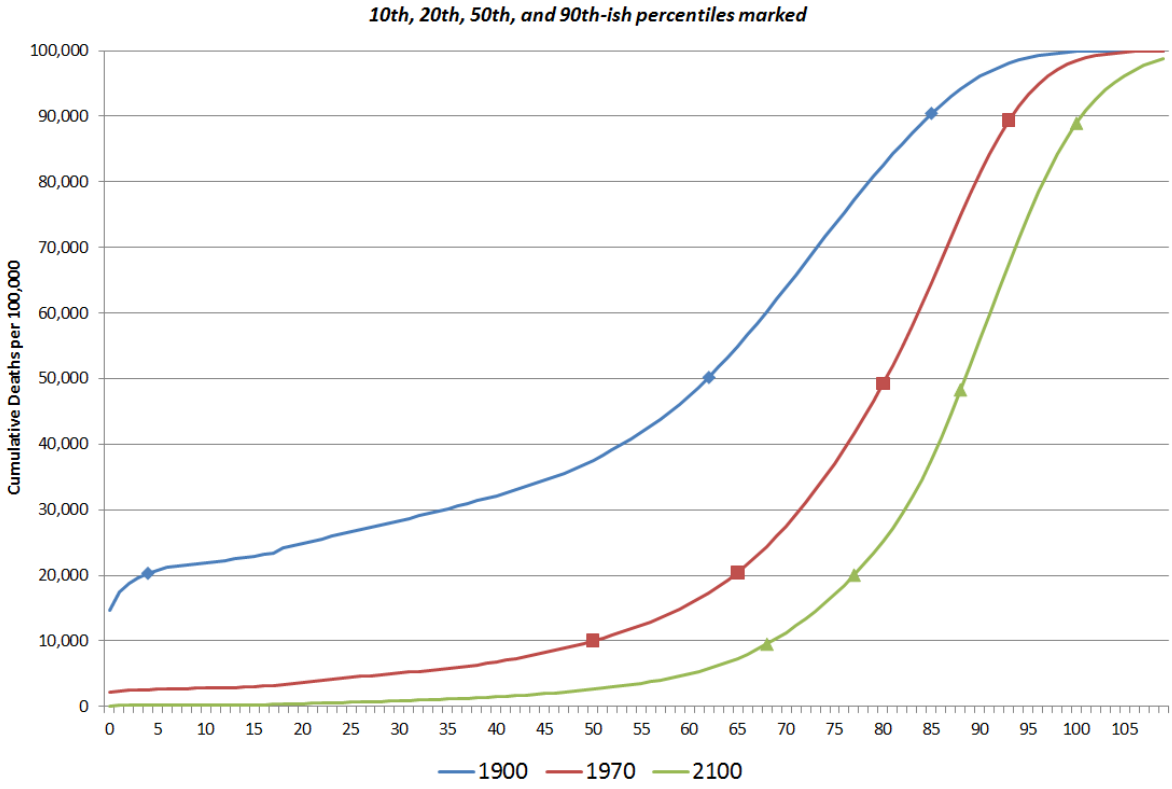
Cumulative death rate refers to the proportion of individuals alive at the start of a specific period of time that dies over that period.

An example of cumulative mortality rate can be found on page 5 (Part 1: Absolute numbers, proportions and rates), where Table 2 shows the comparison among the cumulative mortality rate of some countries.

The concept of cumulative death rate is illustrated by the graph in Figure 6, which shows 3 groups of people: born in 1900, 1970, and 2100 (projected data). At the beginning of life, deaths per

100,000 were low for all three groups. As time goes by, people die and the cumulative deaths increase. At around 100-105 years old, the cumulative death rates are approaching 100% for all three groups. When we compare the curves of the 1900 cohort and the 1970 cohort, we can see that cumulative death rate was higher for the 1900 cohort than the 1970 cohort at all ages, meaning that throughout a life time, people born in 1970 survived better than those born back in 1900.

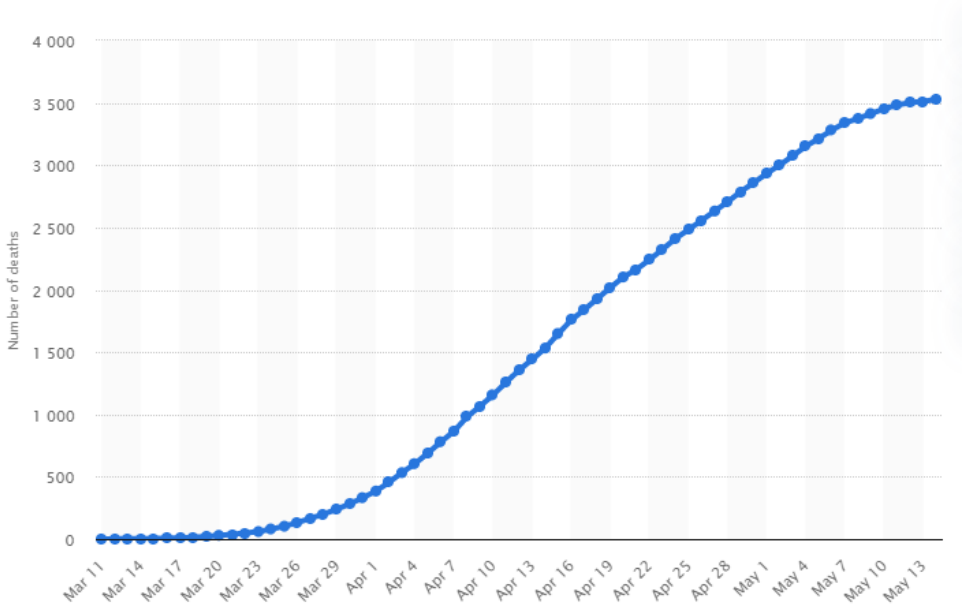
Figure 6. Male cumulative mortality curves, by cohort, actual and projected.



(Source: Meep. Mortality Monday: How young is "So young to die"?; Retrieved from: https://stump.marypat.org/article/676/mortality-monday-how-young-is-so-young-to-die)

Cumulative death rate is not widely used in the reporting of COVID-19 burden but cumulative number of COVID-19 deaths is often used as a descriptive measure. Figure 7 presents an example from Sweden while in Figure 8 the estimated excess of deaths in NY City are illustrated.

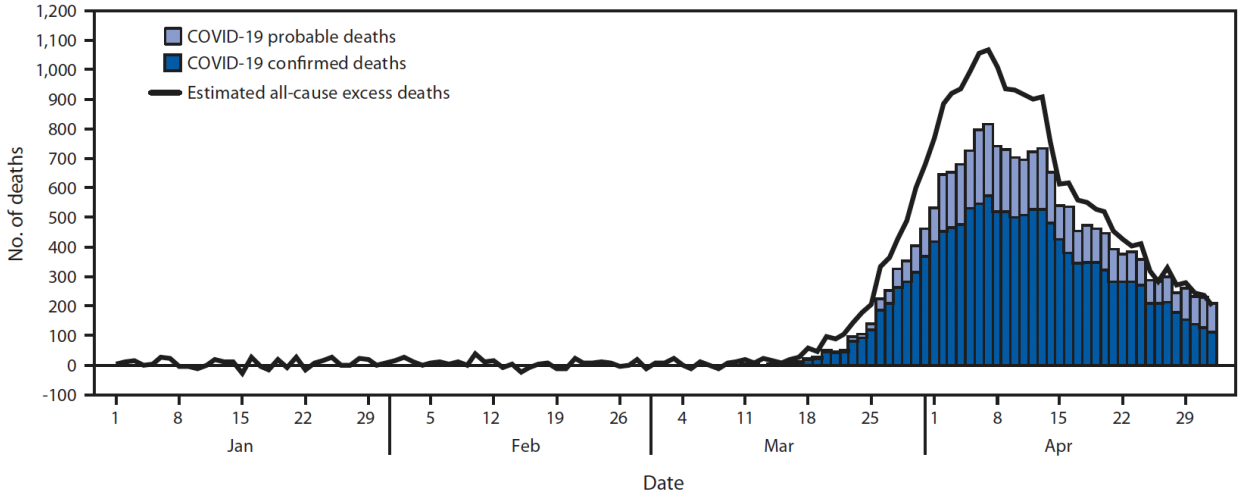
Figure 7. Cumulative number of COVID-19 deaths in Sweden (as of mid-May 2020)



(Source: Statista. Cumulative number of coronavirus (COVID-19 deaths in Sweden since March 11, 2020; Retreived from: https://www.statista.com/statistics/1105753/cumulative-coronavirus-deaths-in-sweden/)

Figure 8. Total estimated excess deaths in NY City (as of 2nd-May 2020)

FIGURE. Number of laboratory-confirmed* and probable† COVID-19–associated deaths and total estimated excess deaths§ — New York City, March 11–May 2, 2020



* Death in a person with a positive laboratory test for SARS-CoV-2 RNA.
† Death in a person without a positive test for SARS-CoV-2 RNA but for whom COVID-19, SARS-CoV-2, or a related term was listed as an immediate, underlying, or contributing cause of death on the death certificate.
§ Total excess all-cause deaths were calculated as observed deaths minus expected deaths as determined by a seasonal regression model using mortality data from the period January 1, 2015–May 2, 2020.

(Source: MMWR, 15 May 2020)

The accuracy of excess mortality projected based on modeling depends largely on the assumptions of the projection method. Since COVID-19 is an ongoing outbreak and the data are evolving continuously, assumptions that are true today may not be true after a certain period when new data emerge.

8. Standardized Mortality Ratio

Standard definition

STANDARDIZED MORTALITY RATIO (SMR): The ratio of the number of deaths observed in the population over a given period, to the number that would be expected over the same period if the study population had the same age-specific rates as the standard population. If the ratio is greater than one, it is interpreted as excess mortality in the study population. If less than one, the study population is interpreted as having lower than expected mortality. The ratio can be directly expressed as the result of that quotient, or expressed by a factor of 100 (in other words, multiplied by 100).

Development of the concepts and examples:

During the COVID-19 epidemic, the SMR was often used (with its confidence intervals) to evaluate the potential excess mortality of the populations affected by the epidemic considering the age distribution of the population, because older populations naturally have a tendency to have higher observed total mortality.

The most frequently used standardization is age standardization because age is an important risk factor for health outcomes. It can be misleading if we compare the mortality of two countries with very different age structure. For many diseases, mortality tends to be higher in an older population. Table 3 compares mortality adjusted by age profile in three countries.

Table 3. Mortality and age structure in England, Belgium and France

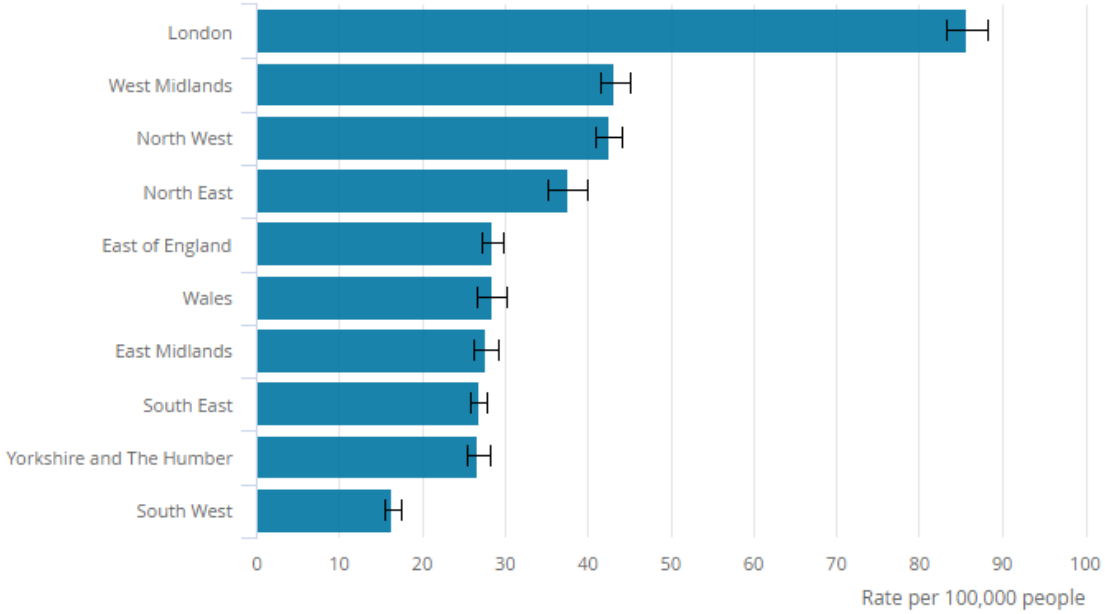
	ENGLAND				BELGIUM			FRANCE		
	Deaths	Population (000)	Deaths per million pop		Deaths	Population (000)	Deaths per million pop	Deaths	Population (000)	Deaths per million pop
80+	533	2439	219	75+	534	1042	512	1444	6231	232
60-79	261	9394	28	65-74	119	1190	100	320	7315	44
40-59	271	14161	19	45-64	45	3102	15	151	16991	9
20-39	66	14304	5	18-44	5	3642	1	16	19325	1
0-19	1	6290	0	0-17	1	2615	0	0	15411	0

(Source: Neil Monnery. Adjusting Covid-19 expectations to the age profile of deaths; Retrieved from: https://blogs.lse.ac.uk/businessreview/2020/04/09/adjusting-covid-19-expectations-to-the-age-profile-of-deaths/)

After age-standardisation, the SMR can then be compared directly and age can no longer explain the apparent difference, instead, other demographic factors, such as gender and socioeconomic status, or health system differences might play a role in the difference in SMR.

Figure 9 is an example, comparing the COVID-19 SMR in different regions in the UK.

Figure 9. Age-standardised mortality rates for deaths involving the coronavirus (COVID-19), per 100 000 population, England and Wales, by country and region (Mar-Apr 2020)



(Source: Office for National Statistics. Deaths involving COVID-19 by local area and socioeconomic deprivation: deaths occurring between 1 March and 17 April 2020; Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsinvolvingcovid19bylocalareasanddeprivation/deathsoccurringbetween1marchand17april)

9. Sensitivity and Specificity

Standard definition

SENSITIVITY OF A TEST: The probability that a diseased person (case) in the population tested will be identified as having the disease by the test. Sensitivity is thus the probability of correctly diagnosing a case, or the probability that any given case will be identified by the test (synonym: true-positive rate).

SPECIFICITY OF A TEST: The probability that a person without the disease (non-case) will be correctly identified as not having the disease by the test. It is thus the probability of correctly identifying a non-diseased person with a test (synonym: true-negative rate).

The relationships are shown in Table 4.

Table 4. Contingency table (2 entrance table) used to calculate sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) (see section 10 for explanation of PPV and NPV).

		True status		
		Diseased	Not diseased	Total
Screening test results	Positive	*a*	b	a+b
	Negative	c	d	c+d
	Total	a+c	b+d	a+b+c+d

a. Diseased individuals detected by the test (true positives)
b. Non-diseased individuals who tested positive (false positives)
c. Diseased individuals not detectable by the test (false negatives)
d. Non-diseased individuals who tested negative (true negatives)

$$\boldsymbol{Sensitivity} = \frac{a}{a+c}$$

$$\boldsymbol{Specificity} = \frac{d}{b+d}$$

Development of the concepts and examples:

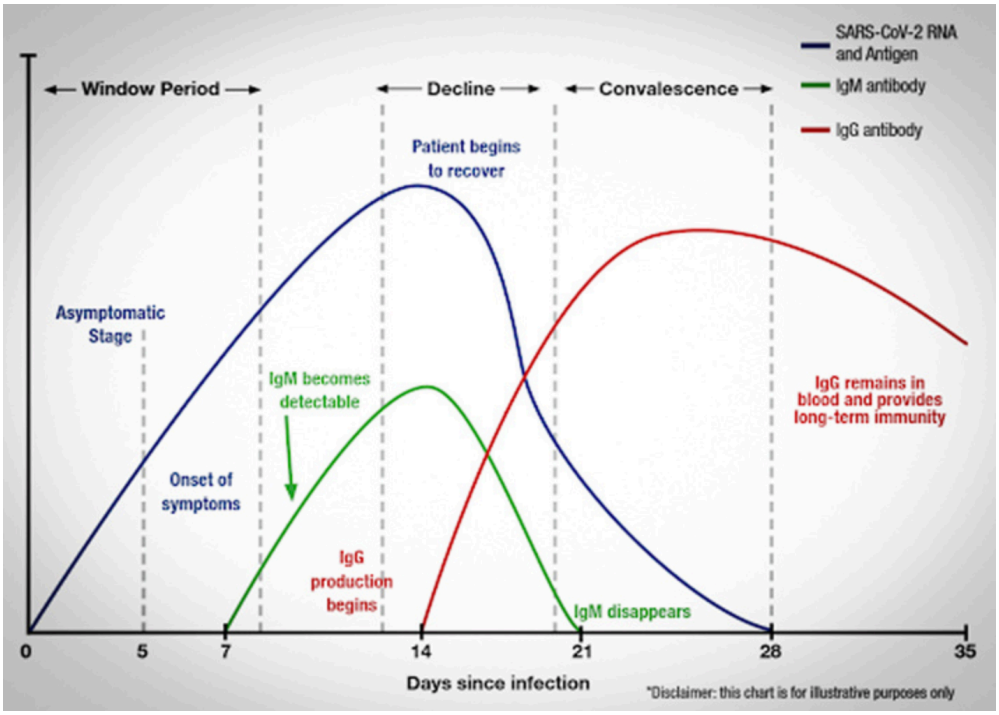
No test is perfect and there is often a trade-off between test performance and time or cost of the test. It is important to know when to use what type of test. Various screening and testing methods are employed in COVID-19 and how a specific test is used hinges on its sensitivity and specificity. Mass screening aims at testing a large population and individuals with a positive result will receive another test for confirmation; therefore, it is important to use a highly sensitive test to minimise

the probability of missing any case and it is less of a concern even if you have some false positive. For confirmatory purpose, you would prefer a highly specific test to exclude the non-diseased.

COVID-19 can be tested by detecting the viral RNA in the nasopharynx or by detecting the antibodies against the virus in blood.

Viral RNA detection is highly specific and is therefore used in many countries to confirm a case in COVID-19. However, the timing of the test and how the sample is collected may affect the sensitivity. It is best to test an individual around the onset of symptoms as the concentration of virus is thought to be highest around this time point in the course of disease. Nasopharyngeal swab is recommended because the virus concentration is the highest in this area in most patients, whereas other swabs or saliva may give lower sensitivity. That means if a person is tested too early (before symptom onset) or if the sample is not collected in the best way, the likelihood of false negative increase and you are more likely to miss a case. When exposed to COVID-19, IgM is the earliest antibody produced, which is followed by a large amount of IgG. Therefore, it takes 3-7 days for an individual infected by SARS-CoV-2 to produce detectable levels of IgM and most patients have detectable IgG by 14 days following onset of symptoms (see Figure 10). This means, such tests have low sensitivity in the early phase of infection. Due to the time lag, the antibody test is not used for identifying cases for isolation and treatment but it can be useful in mass screening when one is interested in finding out the regional or nation-wide disease burden, including the asymptomatic cases. It is worth noting that the antibodies remain in the body for a period of time and thus can be used to check for previous infection.

Figure 10. Trend analysis of SARS-CoV-2 RNA, antigen and antibodies



(Source: Diazyme Laboratories. Why do we need antibody tests for COVID-19 and how to interpret test results; Retrieved from: https://www.diazyme.com/covid-19-antibody-tests)

Sensitivity and specificity of antibody tests can vary greatly depending on the manufacturers. Table 5 shows the sensitivity and specificity of some commercially available SARS-CoV-2 antibody tests.

Table 5. Sensitivity and specificity of some commercial tests

COMMERCIAL TEST	*SENSITIVITY*	*SPECIFICITY*
ARTON LABORATORIES	42.2%	97.9%
ACRO BIOTECH	83.3%	100%
AUTOBIO DIAGNOSTIC	93.3%	100%
DYNAMIKER	90.0%	100%
CTK BIOTECH	90.0%	100%

(Source: Ricco M et al., 2020)

10. Positive predictive value, negative predictive value and overall efficacy of a screening programme

Standard definition

SCREENING: The presumptive identification of unrecognized disease or defect by the application of tests, examinations or other procedures which can be rapidly applied. Screening tests sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic. Persons with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment. The characteristics of a screening test must include accuracy, estimates of yield, precision, reproducibility, sensitivity and specificity, and validity.

ACCURACY: The ability of a diagnostic test to correctly classify the presence or absence of the disorder. The diagnostic accuracy of a test is usually expressed by its sensitivity and specificity.

PREDICTIVE VALUE OF A SCREENING TEST: The probability of the disease given the results of the test. Predictive values of a test are determined by the sensitivity and specificity of the test and by the prevalence of the condition for which the test is used.

POSITIVE PREDICTIVE VALUE (PPV): The probability that a person with a positive test result is a true positive (e.g., does have the disease).

NEGATIVE PREDICTIVE VALUE (NPV): The probability that a person with a negative test result is a true negative (e.g., does not have the disease).

Taking into account Table 5 (in the previous section), the PPV and NPV formulas are the following:

$$\boldsymbol{PPV} = \frac{a}{a+b}$$

$$\boldsymbol{NPV} = \frac{d}{c+d}$$

PRECISION: Relative lack of random error.

REPRODUCIBILITY: A test that gives results that are identical or closely similar each time it is conducted.

VALIDITY: Relative absence of bias or systematic error.

ADHERENCE: Usually expressed as the proportion of people who undergo the screening test on all target population; A measure of participation in a screening program

Development of the concepts and examples:

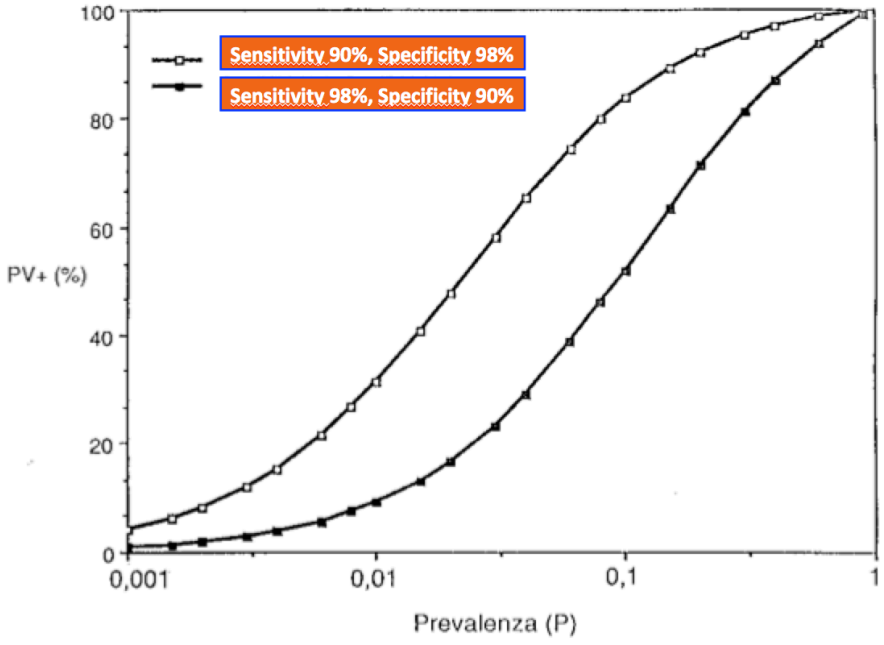
A significant proportion of COVID-19 cases result from the transmission of the virus from asymptomatic or pre-symptomatic cases. **Screening** is a widely employed strategy that consists of testing large populations to find these unrecognized infections. Their aim is to identify as many cases as possible and estimate the spread in the population; a high participation rate in the screenings is therefore essential.

A screening test must meet high quality standards to be efficient: it must be able to correctly detect the presence of the virus, accurately identify cases and be precise to ensure minimal error. Additionally, the test must be reproducible, meaning that it gives consistent results each time it is used.

However, a test almost never correctly diagnoses everyone tested. Sometimes they return a *false positive*, a test result that wrongly identifies a person as being infected or a *false negative*, a test result that fails to identify a person who is infected. To ascertain the likelihood of a false positive or a false negative, **predictive values** of these tests are calculated. The predictive values are determined by the specificity and sensitivity of the test (*see section 9*) but are influenced by the prevalence of the disease in the population considered (see Figure 11).

Many viral tests and antibody tests for COVID-19 are currently being developed. However, they vary in quality and predictive value, which influences the efficiency of screening programmes and can be variable in different populations.

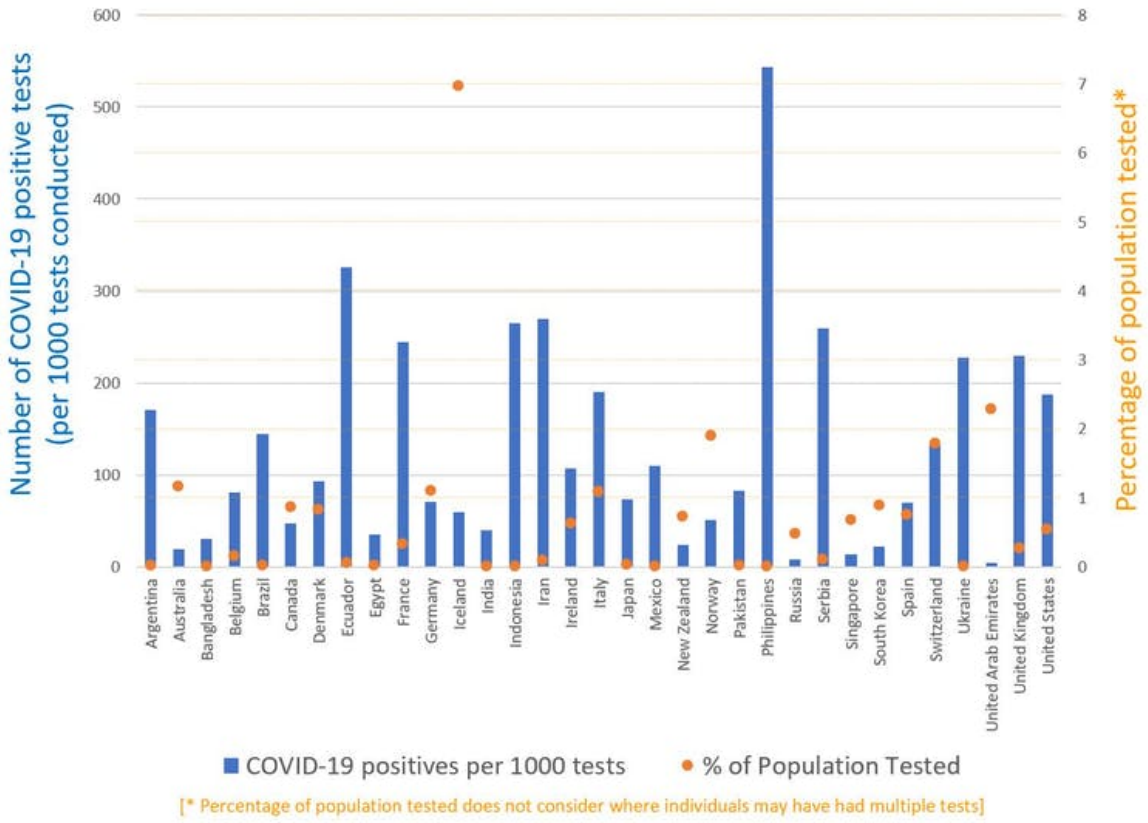
Figure 11. Relationship between positive predictive value and prevalence (log_{10} scale) of a disease in a population screened



(Source: Signorelli C, Elementi di metodologia epidemiologia, Società Editrice Universo, 2011)

When interpreting data on case numbers, it is important to compare these results to the total number of tests conducted and to the proportion of the population that has been tested. As has been observed for COVID-19, screening can vary hugely, both between countries and over time (see Figure 12).

Figure 12. Relationship between number of positive tests and percentage of population tested



(Source: Osborn M. Available at https://theconversation.com/the-bar-necessities-5-ways-to-understand-coronavirus-graphs-135537)

11. Random error, bias, sample, iceberg phenomenon

Standard definition

RANDOM ERROR: Error occurs because of random variations in observation or measurement. Increasing the sample size of a study can reduce random error, but cannot reduce bias.

BIAS: Systematic deviation of results from the truth. An error in the conception and design of a study (or in the collection, analysis, interpretation, reporting, publication, or review of data) leading to results or conclusions that are systematically different from truth.

SELECTION BIAS: A bias caused by the modality in which the sample was selected. E.g., when the study sample is not representative of the population because some characteristics are over- or under-represented in the study population.

INFORMATION BIAS: A bias caused by misclassification of the status of subjects included in the study (e.g., symptoms, risk factors).

SAMPLE: A subset of the population that is included in the study.

ICEBERG PHENOMENON: That portion of disease which remains unrecorded or undetected despite physicians' diagnostic endeavors and community disease surveillance procedures is referred to as the "submerged portion of the iceberg." Detected or diagnosed disease is the "tip of the iceberg." The submerged portion comprises disease not medically attended, medically attended but not accurately diagnosed, and diagnosed but not reported.

Development of the concepts and examples:

When epidemiological studies about COVID-19 are conducted, researchers chose a group of individuals that they want to study in order to answer their research question, the **population**. From this target population, a number of individuals is selected to participate in the study. This is called a **sample**. This sample should be representative of the population so that the findings allow researchers to draw conclusions about various aspects of COVID-19 in the target population.

The data collection process of a study can be flawed by random error and bias.

Random errors can occur because of unknown and unexpected changes in observation and measurement. Having a larger sample could minimize the effect of such errors on the study results.

Bias is a systematic error which results in misleading study results. It can occur in a number of ways:

1. **Selection bias** refers to issues with the way the sample for a study is selected, making it non-representative for the target population. The wide differences in studies of COVID-19 deaths across countries can be attributed to selection bias because each country has a different way of recording their deaths.

   Selection bias is clearly present when using reported cases for the denominator of rates for COVID-19. If only those with more severe symptoms are tested this will affect the denominator

of the incidence rates and case-fatality rates. It will thus depend on the testing strategy of each country. If more mild cases are identified, this is likely to reduce the incidence and case-fatality rates.

Selection bias may also affect the numerator if only deaths occurring in hospital are reported.

2. **Information bias** arises from the misclassification of symptoms or risk factors of study participants. This is often the result of incomplete medical records, testing errors or the misinterpretation of records. This is a pitfall for COVID-19 studies because exposed/infected individuals could be classified as non-exposed/non-infected and vice versa.

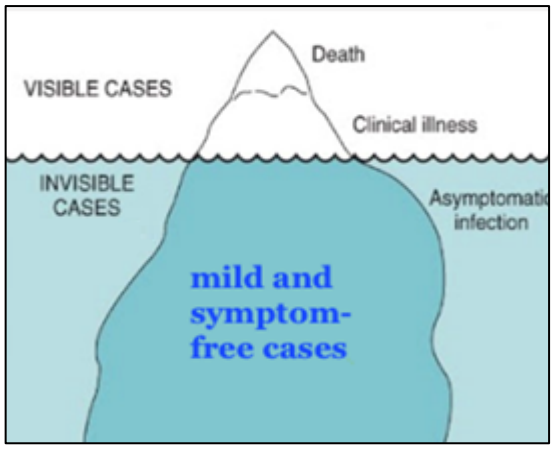
   Information bias can be present in the numerator of the COVID-19 incidence and case-fatality rates, due to the way in which the cause of death is coded. This could be particularly problematic in elderly people with multiple co-morbidities, leading to difficulties in assigning the true cause of death.

   Information bias may also occur in the denominator of incidence and case-fatality rates. The inclusion and exclusion of COVID-19 cases will depend on the sensitivity and specificity of the diagnostic procedures.

3. **Lag time bias** occurs since there is a lag time between the reporting of the case and the death, which can occur up to weeks later. In country reports, cases and deaths are usually reported at the same time, so the cases in the denominator are usually an overestimate of the true denominator, which should be the number of cases reported sometime earlier. This will have a more dramatic effect when the number of cases is rising rapidly.

The **"iceberg phenomenon"** is a metaphor that can be used to explain that a health phenomenon is not always observed and reported. This is quite evidently true for COVID-19 where only a small proportion of cases is known (the tip of the iceberg) (see Figure 13). The submerged part below water represents all cases that remain undetected or unrecorded. This comprises asymptomatic or mild cases, but also cases which are not medically attended or properly diagnosed. This number can be 10 to 25 times higher than the reported cases of COVID-19, highly dependent upon the number of tests performed.

Figure 13. Visualization of the iceberg phenomenon



(Source: Reddy D. et al, 2017)

12. R_0, R_t and the epidemic curve

Standard definition

BASIC REPRODUCTIVE NUMBER (R_0): A measure of the number of infections produced, on average, by an infected individual in the early stages of an epidemic, when virtually all contacts are susceptible.

Table 7. Values of R_0 of selected infectious diseases

Disease	Transmission	R_0
Measles	Aerosol	12–18
Chickenpox (varicella)	Aerosol	10–12
Mumps	Respiratory droplets	10–12
Polio	Fecal–oral route	5–7
Rubella	Respiratory droplets	5–7
Pertussis	Respiratory droplets	5.5
Smallpox	Respiratory droplets	3.5–6
COVID-19	Respiratory droplets	1.94–5.7
HIV/AIDS	Body fluids	2–5
SARS	Respiratory droplets	0.19–1.08
Common cold	Respiratory droplets	2–3
Diphtheria	Saliva	1.7–4.3
Influenza (1918 pandemic strain)	Respiratory droplets	1.4–2.8
Ebola (2014 Ebola outbreak)	Body fluids	1.5–1.9
Influenza (2009 pandemic strain)	Respiratory droplets	1.4–1.6
Influenza (seasonal strains)	Respiratory droplets	0.9–2.1
MERS	Respiratory droplets	0.3–0.8

(Source: Wikipedia with scientific references)

EFFECTIVE REPRODUCTIVE NUMBER (R_t): The value of the R_0 index can be changed as a result of the introduction of preventive measures (i.e. physical distancing, use of masks, etc.) or following a reduction in the number of susceptible people due to post-infection acquired immunity or to vaccinations. This reproduction number is defined as R_t, that is the actual transmission rate of the virus at a given time *t*. This appropriately denotes the effective reproduction number during an evolving epidemic such as COVID-19.

EPIDEMIC CURVE: A graphic plotting of the distribution of cases by time of onset, in a linear or logarithmic scale. When presented in a logarithmic scale, the vertical axis is graduated by orders of magnitude (1, 10, 100, 1,000), and this is the preferred method to plot an epidemic that is growing exponentially, so that large numbers do not skew the entire graph.

Development of the concepts and examples:

An **epidemic curve** of an outbreak is a statistical graph that visualizes the number of cases and their temporal progression. It commonly shows the number of new cases on the vertical axis and the corresponding date on the horizontal axis. Figure 14 presents an example of the global epidemic curve of COVID-19.

Figure 14. Total cases of COVID-19 worldwide in linear (left) and logarithmic (right) scales (as of May, 19 2020)

(Source: Retrieved from https://www.worldometers.info/coronavirus/worldwide-graphs/)

The progression of the epidemic curve of COVID-19 depends on the **basic reproduction number R_0** (pronounced *R nought*), which measures the potential for the virus to spread in the population. R_0 can be defined as the average number of new cases generated by an infectious case in a totally susceptible population. As the virus that causes COVID-19, SARS-CoV-2, is a novel virus, the world population has not been exposed before, effectively making everybody susceptible.

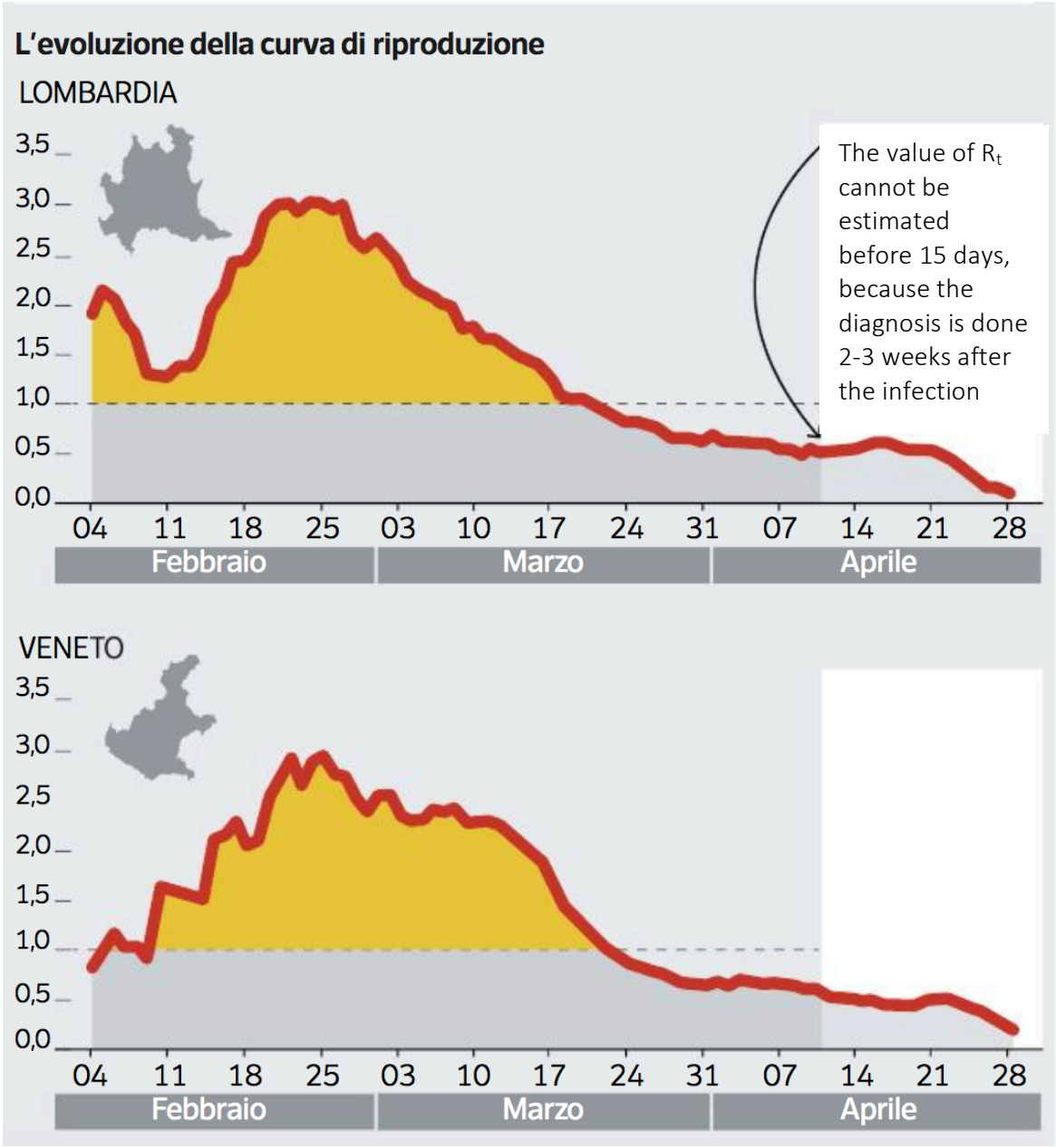
Generally speaking, R_0 depends on the number of days people are infectious, the number of susceptible people they interact with and the chance of transmission during such an interaction.

An epidemic only develops if R_0 is greater than 1. This means that every infected person on average infects more than one new person. Modelling studies currently estimate the R_0 of COVID-19 at between 2 and 3, but this is subject to change.

A crucial point for the calculation of R_0 and R_t is to have reliable information on the total number of infected people in the various geographical areas and on the date of infection or onset of symptoms, data not easy to obtain in the case of the COVID-19 epidemic. Therefore in this context

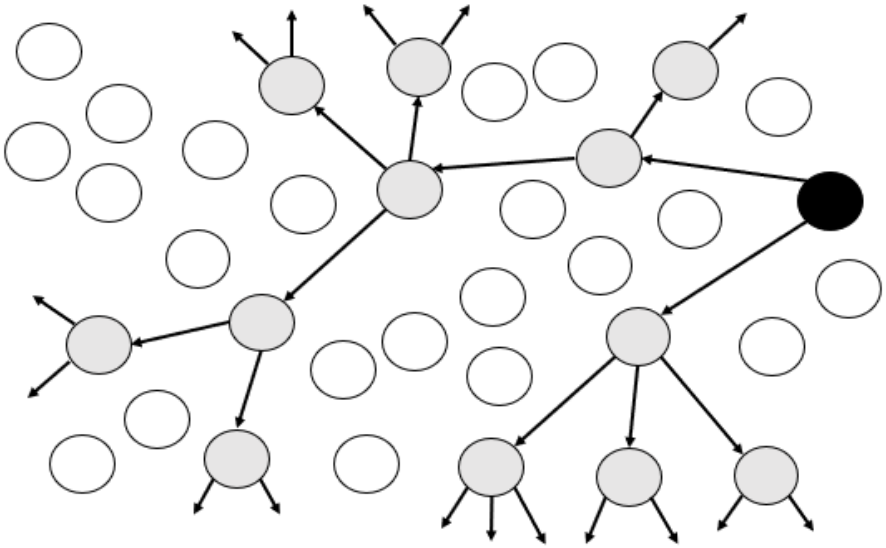
R_0 and R_t were estimated only at a later time (see Figure 15) and the usefulness of using the R_t index to predict the evolution of the epidemic – as it was proposed to do in phase 2 of the epidemic – does not appear supported by sufficient scientific evidence, also due to the frequent changes in external conditions (reopening of some business activities, resumption of social contacts). In Figure 15 the reproduction number in two Italian regions were estimated. Official news of the beginning of the epidemic was February 22nd; after that date progressive prevention measures were put into place and the index shifted from R_0 to R_t.

Figure 15. Evolution of reproduction number in Italy (Lombardy and Veneto Regions)



(Source: Corriere della sera, 2020)

Figure 16. Schematic spread of COVID-19 in a group



(Source: ASPHER original work)

The black dot at the right border (Figure 16) represents the person who introduced the virus to the group. They infect two other persons, the grey dots, who then in turn infect 5 other persons, and so on.

The goal of the current mitigation strategies, such as social distancing, is to push R_0 below 1. This would mean that one infected person on average infects less than one other person, leading to the epidemic petering out.

Since COVID-19 may confer some immunity, the potential for the virus to spread changes as the epidemic develops. More people become immune after their infection and the susceptible population decreases. This is measured by the **effective reproduction number**, denoted as $\mathbf{R_t}$.

However, one needs to be mindful that various contextual factors, such as behavior or living conditions, can influence the spread. This results in varying R_t depending on the setting.

13. Epidemiological surveillance

Standard definition

CASE DEFINITION: Establishing unified standard criteria for categorizing for person, place, time, and clinical features (CDC 2020).

CRITERIA FOR CASE DEFINITION:

I. **SUSPECT CASE:** Unspecified initial sign and symptom
II. **PROBABLE CASE:** Description of clinical criteria and epidemiological link
III. **CONFIRMED CASE:** Laboratory confirmation

CASE FINDING: First identify the primary source, the person that public health authorities suspect as the index case. After that the goal is to identify and trace as many cases as possible in order to establish the magnitude of the outbreak.

CONTACT TRACING: "Contacts" are subjects that have come into contact with an infected person during the incubation period or the symptomatic stage of the disease, thus having the potential of being infected. An important part of the process of epidemiological surveillance consists in tracing the contacts of infected people, collecting information on their present infection status and following up with them to record the onset of any symptoms. Subsequently, they might be quarantined by health authorities. During the COVID-19 pandemic the use of digital contact tracing has been implemented by some countries; despite its efficiency, this method may raise important privacy issues which have to be balanced with the public health imperative.

INCUBATION PERIOD: The incubation period is essentially the time between exposures to the causative agent until the onset of symptoms for each disease agent. For example, the incubation period for COVID-19 is thought to extend to 14 days, with a median time of 4-5 days from exposure to symptoms onset

ISOLATION: separates sick people with a contagious disease from people who are not sick.

QUARANTINE: separates and restricts the movement of people who were exposed to a contagious disease to see if they become sick.

Development of the concepts and examples:

The World Health Organization released an interim guidance to perform an accurate contact tracing. They state that contact tracing can only be effective if countries have adequate capacity to test suspect cases in a timely manner. Otherwise, testing and contact tracing strategies can focus on specific high-risk settings with vulnerable individuals, such as hospitals and care homes.

The terms ***quarantine*** and ***isolation*** are strictly related to the plague and date back to the year 1377. The chief physician of Ragusa, Jacob of Padua, established a place outside the city walls for the treatment of sick (or suspected to be infected) citizens for 40 days to land travelers. Furthermore, in 1423 Venice set up one of the first known 'lazzaretto' (quarantine station) on an island near the city, and the Venetian system became a model for other European countries. (Source: Cosmacini G. et al., 2001; Sehdev P.S. et al., 2002)

That being said, quarantine doesn't necessarily last for 40 days: its duration depends on the maximum incubation period of a disease. For example, the incubation period of measles lasts 9 to 15 days, for MERS-CoV the incubation period lasts 5 to 7 days; finally influenza has an incubation period that lasts from a few hours to a couple of days.

An estimation of the maximum duration of the incubation period as precise as possible is necessary to plan public health interventions, including active surveillance, infection control and modeling of the epidemic.

According to a study by Johns Hopkins Bloomberg School of Public Health, published on Annals of Internal Medicine, COVID-19 has a median incubation period estimated between 2 to 14 days. 97.5% of people develop symptoms within 11.5 days from exposure, hence the recommended quarantine period of 14 days is a reasonable amount of time.

Quarantine measures have not been used for a long time, but it's included in the International Health Regulations (adopted by WHO) and it's been employed for COVID-19 due to its relatively long incubation period, in particular for contacts of confirmed cases and areas with high concentration of cases.

Figure 17. Historic depiction of a quarantine area

(Source: Malta: view of the quarantine area. Etching by M-A. Benoist, c. 1770, after J. Goupy, c. 1725.)

Figure 18. Example of quarantine life in 2020 during COVOD-19 pandemic

14. Epidemiological trend

Standard definition

EPIDEMIOLOGICAL TREND: is the branch of epidemiology that deals with causes and distribution of diseases in the general population over time, to assess if there have been significant changes in disease patterns throughout the world. It applies statistics to explain present disease patterns but also to help predict how they may change in the future.

EPIDEMIC: The occurrence in a community or region of cases of an illness clearly in excess of normal expectancy.

OUTBREAK: An epidemic limited to a localized increase in the incidence of a disease, e.g., in a village or town.

PANDEMIC: An epidemic occurring worldwide or over a very wide area, crossing international borders, and usually affecting a large number of people.

SPORADIC: An infectious disease occurring irregularly, from time to time, and generally infrequently.

ENDEMICITY: The constant presence of a disease or infectious agent within a given geographic area or population group.

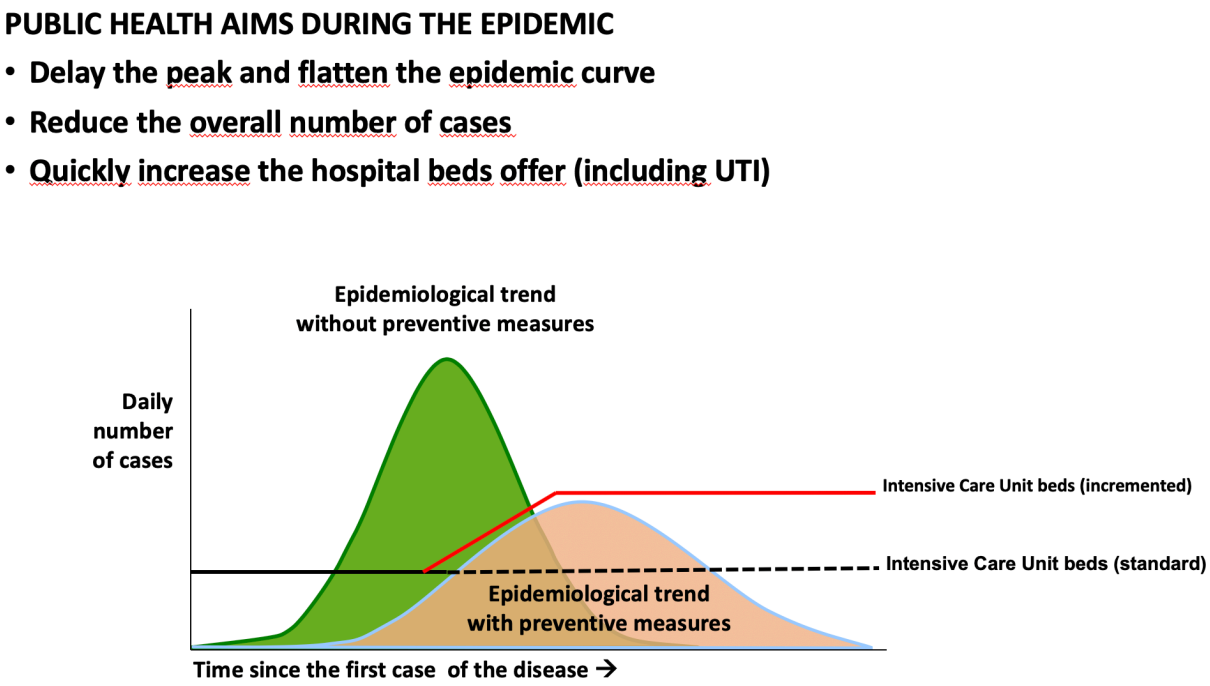
Development of the concepts and examples:

COVID-19 is considered to have started as an outbreak limited to the province of Wuhan, China. Then the number of reported cases started increasing rapidly, marking an epidemic. It was declared by the WHO Public Health Emergency of International Concern (PHEIC) on January 30th 2020. On March 11th 2020, the WHO declared the COVID-19 a pandemic, spread over several countries and continents.

Flattening the curve: The commonly used phrase "Flattening the Curve" is a public health strategy to reduce the number of new COVID-19 infections to a level within the capacity limits of a healthcare system. This is particularly important for intensive care unit (ICU) beds that patients with severe illness from the virus need (red line in Figure 19). The faster the epidemic curve rises, the quicker a healthcare system can be overloaded and reach its capacity limits (the part of the green curve above the red line in Figure 19). To avoid this, a flatter epidemic curve is needed. This can be achieved by interventions, such as containment and mitigation measures (social distancing, use of masks, personal hygiene behaviour, lockdown, etc.), that slow the spread of the virus (brown curve). The same number of people may still become sick but the number of cases spreads over a longer time period. This reduces the number of people requiring care at the same time and allows hospitals to treat everyone. In Figure 19 the standard way to illustrate this phenomenon graphically

is integrated with a possible increase of hospital beds in other to satisfy the demand, as happened in many countries during the first phase of COVID-19 epidemic.

Figure 19. "Flattening the Curve"



(Source: Signorelli C, et, 2020.)

15. Herd immunity

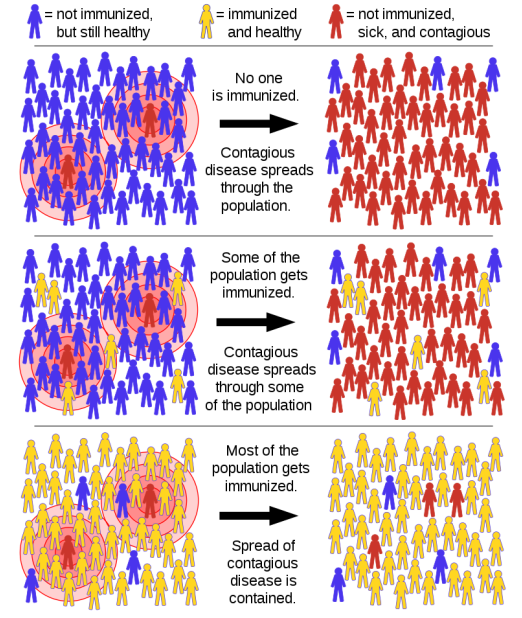
Standard definition

HERD IMMUNITY: Resistance of a population to invasion and spread of an infectious agent, based on the agent-specific immunity of a high proportion of the population reducing the likelihood that an infected person will come in contact with a susceptible one among human populations, also called community immunity. The rationale is that if a large proportion of the population is immune to a virus, many people who encounter someone with the disease won't get sick (and won't spread the disease any further), greatly curtailing the transmission of the disease. The proportion of the population required to be immune varies according to the agent, its transmission characteristics, the distribution of immune and susceptible individuals, and other (e.g., environmental) factors.

Development of the concepts and examples:

Herd immunity can be achieved with the infection of a relevant part of a population or through vaccination campaigns. The proportion of the population not susceptible to the achievement of the herd immunity varies according to the mode of transmission and the contagiousness of the infective agent. For many infectious childhood diseases, it is between 90% and 95%. for COVID-19 it could be even lower (60-70%) (Randolph HE, et al. 2020).
. See Figure 20 for three different infectious disease spread scenarios based on the immunity proportion of the population.

Figure 20. Three different scenarios – with different community immunity proportions – as an example of herd immunity.



(Source: Tkarcher - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=56760604)

References:

- Banerjee A, Pasea L, Harris S, Gonzalez-Izquierdo A, Torralbo A, Shallcross L, et al. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. Lancet. 2020;395(10238):1715-1725. doi:10.1016/S0140-6736(20)30854-0
- Beaglehole R, Bonita R, Kjellström T. Basic epidemiology, WHO 1993
- Benois MA . Malta: view of the quarantine area., c. 1770, after J. Goupy, c. 1725.
- Catalogue of OECD indicators, OECD 2016
- Center for Disease Control and Prevention. Principles of Epidemiology in Public Health Practices, 3rd Ed
- Center for Systems Science and Engineering at Johns Hopkins University. COVID-19 Dashboard available at https://coronavirus.jhu.edu/map.html. Accessed on 25.05.20
- Chowell G, Hyman JM. Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases, Springer 2016
- Coronavirus deaths worldwide per million inhabitants. Available at https://www.statista.com/statistics/1104709/coronavirus-deaths-worldwide-per-million-inhabitants/; Retrieved on 14 May 2020
- Corriere della sera. L'evoluzione della curva di riproduzione
- Cosmacini G. L'arte lunga. Storia della medicina dall'antichità a oggi. Editori Laterza; Bari: 2001
- Cumulative number of coronavirus (COVID-19 deaths in Sweden since March 11, 2020); Available at https://www.statista.com/statistics/1105753/cumulative-coronavirus-deaths-in-sweden/
- Diazyme Laboratories. Why do we need antibody tests for COVID-19 and how to interpret test results; Retrieved from: https://www.diazyme.com/covid-19-antibody-tests
- Epicentro – Istituto Superiore di Sanità, Caratteristiche dei pazienti deceduti positivi all'infezione da SARS-CoV-2 in Italia. Available at https://www.epicentro.iss.it/coronavirus/bollettino/Report-COVID-2019_21_maggio.pdf
- EuroMoMo. Graphs and maps; Retrieved from: https://www.euromomo.eu/graphs-and-maps
- Glossary of Humanitarian Terms, ReliefWeb 2008
- Green MS, Peer V, Nitzan D. The confounded crude case-fatality rates for COVID-19 hide more than they reveal - a comparison of age-specific and age-adjusted rates between six countries. Preprint https://doi.org/10.1101/2020.05.09.20096503)
- Istituto Superiore di Sanità. Integrated surveillance of COVID-19 in Italy, available at epicentro.iss.it. Accessed on 15.05.20
- Meep. Mortality Monday: How young is "So young to die"? Retrieved from: https://stump.marypat.org/article/676/mortality-monday-how-young-is-so-young-to-die
- Monnery N. Adjusting Covid-19 expectations to the age profile of deaths; Retrieved from: https://blogs.lse.ac.uk/businessreview/2020/04/09/adjusting-covid-19-expectations-to-the-age-profile-of-deaths/
- Office for National Statistics. Coronavirus Main Figures. Available at www.ons.gov.uk
- Osborn M. The bar necessities: 5 ways to understand coronavirus graphs. Available at https://theconversation.com/the-bar-necessities-5-ways-to-understand-coronavirus-graphs-135537
- Porta M. A Dictionary of Epidemiology, Oxford University Press, 5th edition

- Randolph HE, Barreiro LB. Herd Immunity: Understanding COVID-19. Immunity. 2020;52(5):737-741. doi:10.1016/j.immuni.2020.04.012
- Reddy D, Kalyani G, Pradeep K, Asif MD Kartheek D, Gangabhavani M. The Survey Of Cancer Patients In The Region Of Guntur: Based On Hospital Registry. International Journal of Pharmacy and Pharmaceutical Sciences. 9. 288. 10.22159/ijpps.2017v9i2.16026.
- Reported Cases and Deaths by Country, Territory, or Conveyance, available at https://www.worldometers.info/coronavirus/#countries
- Riccò M, Ferraro P, Gualerzi G, Ranzieri S, Henry BM, Said YB, Pyatigorskaia NV, Nevolina E, Wu J, Bragazzi NL, Signorelli C. Point-of-Care diagnostic for detecting SARS-CoV-2 antibodies: a systematic review and meta-analysis of real-world data. Journal of Clinical Medicine 2020
- Russel TW, Hellewell J, Jarvis CI, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. Euro Surveill. 2020;25(12):pii=2000256. https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000256
- Sehdev PS. The origin of quarantine. Clin Infect Dis. 2002;35:1071–1072
- Signorelli C. Elementi di metodologia epidemiologia, Società Editrice Universo, 7th edition
- Signorelli C, Odone A, Gianfredi V, Bossi E, Bucci D, Oradini-Alacreu A, Frascella B, Capraro M, Chiappa F, Blandi L, Ciceri F. The spread of COVID-19 in six western metropolitan regions: a false myth on the excess of mortality in Lombardy and the defense of the city of Milan. Acta Bio Med. 2020May11;91(2):23-0.
- Signorelli C, Scognamiglio T, Odone A. COVID-19 in Italy: impact of containment measures and prevalence estimates of infection in the general population . Acta Bio Med. 2020 Apr.10;91(3-S):175-9
- The Public Health Textbook, available at https://www.healthknowledge.org.uk/public-health-textbook. Public Health Action Support Team (PHAST) 2020
- Tkarcher - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=56760604
- Values of selected infectious diseases, available at https://en.wikipedia.org/wiki/Basic_reproduction_number
- Ward H, Toledano MB, Shaddick G, Davies B, Elliot P. Oxford Handbook of Epidemiology for Clinicians, Oxford University Press 2012
- World Health Organization. Coronavirus disease (COVID-19) Situation Report – 112. 2020.
- World Health Organization. Contact tracing in the context of COVID-19. Interim guidance. 10 May 2020