I numeri del SARS-CoV-2 (COVID-19)

Yinon M. Bar-On, Avi Flamholz, Rob Phillips, and Ron Milo

1Weizmann Institute of Science, Rehovot 7610001, Israel 2University of California, Berkeley, CA 94720, USA 3California Institute of Technology, Pasadena, CA 91125, USA 4Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

*Corresponding author: ron.milo@weizmann.ac.il

Commenti sono ben accetti; questo è articolo aggiornato costantemente a: https://bit.ly/2WOeN64

**Taglia e Contenuto**

Diametro: ≈100 nm
Volume: ≈10^6 nm^3 = 10^3 fL
Massa: ≈10^3 Da = 1 fg

**Genoma**

Identità nucleotidica con SARS-CoV-2

- Pipistrello CoV
- Pangolino CoV
- SARS-CoV-1
- MERS-CoV
- Raftreldoce CoV

- SARS-CoV-2

Lunghezza =30kb; β-coronavirus con 10-14 ORFs (24-27 proteine)

Tasso di evoluzione: ≈10^-2 nt/yr^-1 (misurato per SARS-CoV-1)
Tasso di mutazione: ≈10^-6 nt/cycle^-1 (misurato per MHV coronavirus)

**Tempi di replicazione**

In coltura tessutale
Entrata del virione nella cellula: ≈10 min (misurato per SARS-CoV-1)
Periodo di eclisse: ≈10 ore
Numero di virus rilasciati per cellula: ≈10^2 Virioni (misurato per MHV coronavirus)

**Cellule infette**

- Pneumociti di Tipo I e Tipo II (≈10^11 cellule)
- Macrophagi alveolari (≈10^10 cellule)
- Cellule mucose della cavità nasale (≈10^9 cellule)

Volume delle cellule infette: ≈10^5 μm^3 = 10^3 fL

Virioni non in scala

**Risposta anticolpore - Sieroconversione**

Anticorpi appaiono nel sangue dopo: ≈10-20 giorni
Mantenimento della risposta anticolpore: ≈2-3 anni (misurato per SARS-CoV-1)

**Stabilità del virus nell’ambiente**

Rilevanza per la sicurezza personale non è nota

- Emivita: 1000 volte
- Tempo di decespennamento: 1-2 ore
- Aerosol: ≈1 h
- Superficie: ≈4-24 h
- E. gplastiche, cartoncino

Basato sulla quantificazione di virioni infettivi. Testato a 21-23°C
Con umidità relativa di 40-65%. I valori variano a seconda delle Condizioni e il tipo di superficie (Doran et al. 2020).

RNA virale è osservato su superfici anche dopo alcune settimane (Morarish et al. 2020).

Nota la differenza nella notazione tra il simbolo «» che indica "approssimativamente" e denota un’incertezza entro un fattore di 2, e il simbolo «» che indica "ordine di magnitudine" o accuratezza entro un fattore di 10.

**“Tipica” progressione dell’infezione in un singolo paziente**

- Tasso netto di riproduzione R0: tipicamente 2-4
- Varia ulteriormente nel tempo e nello spazio (Li et al. 2020; Park et al. 2020)
- (numero di nuovi casi direttamente generati da un singolo paziente)

**Infezione da virus**

- **Periodi di incubazione:** <5 giorni (99% ≤ 14 giorni a meno di essere asintomatico)
- **Diagnosi dopo <5 giorni**
- **Sintomatico**
- **Esposto**
- **Infettivo**
- **Periodo Latente** <3 giorni
- **Periodo di contagio (infettivo)»4 giorni**

**Tasso di letalità:** ECDC 2020
(1-5%) = 0.3-10% (non corretto)
Tasso di infezione (IFR) = 0.3-1.3%

La variabilità interindividuale è sostanziale e non ben caratterizzata. Le età sono parametri che si adattano alla mediana della popolazione in Cina e non tengono conto di questa variabilità (Li et al. 2020; He et al. 2020).

La quantità di RNA può marcatamente sovrastimare i virioni infettivi.
Taglia e Contenuto

**Diametro**: ≈100 nm  
**Volume**: ~10^{6} nm^{3} = 10^{-3} fL  
**Massa**: ~10^{3} MDa ≈ 1 fg

**Membrana**: ≈2000 copie (misurate per SARS-CoV-1)  
**Envelope** (Involucro): ≈20 copie (100 monomeri, misurati per il coronavirus TGEV)  
**Nucleoproteine**: ≈1000 copie (misurate per SARS-CoV-1)  
**Trimeri di spikes** (spine):  
  - **Lunghezza**: ≈10 nm  
  - **Copie per virione**: ≈100 (misurate per SARS-CoV-1) (300 monomeri)  
  - **Affinità per recettori ACE2**: {K_{c}} = 1-30 nM  
  - **Facilitato da TMPRSS2**

**Genoma**  
**Lunghezza genoma**: ≈30kb  
**Numero di geni**: 10-14  
**Numero di proteine**: 24-27  
**Tasso di evoluzione**: ~10^{-3} nt·yr^{-1} (misurato per SARS-CoV-1)  
**Tasso di mutazione**: ~10^{-6} nt·cycle^{-1} (misurato per coronavirus MHV)  
**Identità nucleotidica con SARS-CoV-2**: pipistrello CoV - 96%; pangolino CoV 91%; SARS-CoV-1 80%; MERS 55%; raffreddore CoV 50%  

**Tempi di replicazione**  
In coltura tissutale  
**Entrata del virione nella cellula**: ~10 min (misurato per SARS-CoV-1)  
**Periodo di eclisse**: ~10 ore  
**Numero di virus rilasciati per cellula**: ~1000 virioni (misurata per coronavirus MHV)

**Cellule infettate**  
( lista approssimativa; numero di cellule per persona)  
- Pneumociti di Tipo I e Tipo II: ~10^{11} cellule  
- Macrophagi alveolari: ~10^{10} cellule  
- Cellule mucose della cavità nasale: ~10^{9} cellule  
**Volume delle cellule infettate**: ~10^{3} pm^{3} = 10^{-3} fL

**Concentrazione**  
Valori massimi osservati dopo la diagnosi  
- Nasofaringe: 10^{6}-10^{9} RNAs/tampone  
- Gola: 10^{4}-10^{8} RNAs/tampone  
- Feci: 10^{4}-10^{8} RNAs/g  
- Sputo: 10^{6}-10^{11} RNAs/mL  
La quantità di RNA può marcatamente sovrastimare i virioni infettivi

**Risposta anticorpale - Sieroconversione**  
- **Anticorpi appaiono nel sangue dopo**: ≈10-20 giorni  
- **Mantenimento della risposta anticorpale**: ≈2-3 anni (misurato per SARS-CoV-1)

**Stabilità del virus nell’ambiente**  
Rilevanza per la sicurezza personale non è nota  
**emivita**:  
- **Aerosoli**: 1 h  
- **Superfici**: 1-7 h  
- **E.g. plastica, cartone, E metalli**: 4-96 h  

Basato sulla quantificazione di virioni infettivi. Testato a 21-23°C  
Con umidità relativa di 40-65%. I valori variano a seconda delle Condizioni e al tipo di superfici (rif).  
RNA virale è osservato sulle superfici anche dopo alcune settimane (rif)

**“Tipica” progressione dell’infezione in un singolo paziente**  
**Tasso netto di riproduzione, R_{0}**: tipicamente 2-4, ma varia ulteriormente nel tempo e nello spazio  
(numbero di nuovi casi direttamente generati da un singolo paziente)  
**Periodi di incubazione (mediana)**: ≤5 giorni (99% ≤ 14 giorni a meno di essere asintomatici)  
**Diagnosi dopo**: ≤5 giorni  
**Periodo latente**: ≤3 giorni  
**Periodo di contatto (infettivo)**: ≤4 giorni  
**Guarigione**: casi leggeri: ≤2 settimane  
- casos severi: ≤6 settimane  
**Tasso di letalità**: 0.8%-10% (non corretto)  
**Tasso di infezione**: 0.3%-1.3%  
La variabilità inter-individuale è sostanziale e non ben caratterizzata. Le stime sono parametri che si adattano alla mediana della popolazione in Cina e non tengono conto di questa variabilità (rif, rif).

Nota la differenza nella notazione tra il simbolo ≈, che indica "approssimativamente" e denota un'accuratezza entro un fattore di 2, e il simbolo ~, che indica "ordine di magnitudine" o accuratezza entro un fattore di 10.

**Abstract**  
The current SARS-CoV-2 pandemic is a harsh reminder of the fact that, whether in a single human host or a wave of infection across continents, viral dynamics is often a story about the numbers. In this snapshot, our aim is to provide a one-stop, curated graphical source for the key numbers that help us understand the virus driving our current global crisis. The discussion is framed around two broad themes: 1) the biology of the virus itself and 2) the characteristics of the infection of a single human host. Our one-page summary provides
the key numbers pertaining to SARS-CoV-2, based mostly on peer-reviewed literature. The numbers reported in summary format are substantiated by the annotated references below. Readers are urged to remember that much uncertainty remains and knowledge of this pandemic and the virus driving it is rapidly evolving. In the paragraphs below we provide "back of the envelope" calculations that exemplify the insights that can be gained from knowing some key numbers and using quantitative logic. These calculations serve to improve our intuition through sanity checks, but do not replace detailed epidemiological analysis.

## Sommario

L’attuale pandemia di SARS-CoV-2 ci ricorda duramente che, sia che si tratti di una singola persona infetta o di un’ondata di infezione che varca i continenti, le dinamiche virali sono spesso una storia fatta di numeri. In questa illustrazione, il nostro obiettivo è quello di fornire un’immediata e accurata rappresentazione che fosse comunemente chiave per aiutarsi a capire il virus che sta trainando l’attuale crisi globale. La discussione è organizzata attorno a due temi principali: 1) la biologia stessa del virus e 2) le caratteristiche dell’infezione in una singola persona infetta. Il nostro breve compendio fornisce i numeri chiave del SARS-CoV-2, basati per la maggior parte su letteratura scientifica (peer-reviewed). I numeri riportati in breve nella grafica sono valutati dalle referenze indicate sotto. I lettori devono ricordare che c’è ancora molta incertezza e che le conoscenze su questa pandemia e sul virus responsabile di essa sono in rapida evoluzione. Nei paragrafi sotto, forniamo calcoli “oltre l’involucro (envelope)” che esemplificano le intuizioni che si possono cogliere conoscendo alcuni numeri chiave e usando la logica quantitativa. Questi calcoli servono a migliorare la nostra percezione (del virus e della pandemia) acquistata attraverso i controlli sanitari, ma non possono sostituire analisi epidemiologiche dettagliate.

### 1. How long does it take a single infected person to yield one million infected people?

If everybody continued to behave as usual, how long would it take the pandemic to spread from one person to a million infected victims? The basic reproduction number, $R_0$, suggests each infection directly generates $2-4$ more infections in the absence of countermeasures like social distancing. Once a person is infected, it takes a period of time known as the latent period before they are able to transmit the virus. The current best-estimate of the median latent time is $3$ days followed by $4$ days of close to maximal infectiousness (Li et al. 2020, He et al. 2020). The exact durations vary among people, and some are infectious for much longer. Using $R_0=4$, the number of cases will quadruple every $4$ days or double every $3$ days.

1000-fold growth (going from one case to $10^3$) requires $10$ doublings since $2^5 = 10$; $3$ days $\times$ $10$ doublings $= 30$ days, or about one month. So we expect $1000x$ growth in one month, million-fold ($10^6$) in two months, and a billion fold ($10^9$) in three months. Even though this calculation is highly simplified, ignoring the effects of “super-spreaders”, herd-immunity and incomplete testing, it emphasizes the fact that viruses can spread at a bewildering pace when no countermeasures are taken. This illustrates why it is crucial to limit the spread of the virus by social distancing measures. For fuller discussion of the meaning of $R_0$, the latent and infectious periods, as well as various caveats, see the “Definitions” section.

### 2. What is the effect of social distancing?

A highly simplified quantitative example helps clarify the need for social distancing. Suppose that you are infected and you encounter $50$ people over the course of a day of working, commuting, socializing and running errands. To make the numbers meaningful, let’s consider an average person encounter with $300$ people over a year. The viral load per person encounter is estimated to be $10^6$ viral particles per person (Lauer et al. 2020). If you instead meet $5$ people each day (preferably fewer) because of social distancing, then you will infect $0.1$ people per day, or $0.4$ people before you become less infectious. The desired effect of social distancing is to make each current infection produce $<1$ new infections. An effective reproduction number ($R_0$) smaller than $1$ will ensure the number of infections eventually dwindles. It is critically important to quickly achieve $R_0 < 1$, which is substantially more achievable than pushing $R_0$ to near zero through public health measures.

### 3. Why is the quarantine period two weeks?

The period of time from infection to symptoms is termed the incubation period. The median SARS-CoV-2 incubation period is estimated to be roughly $5$ days (Lauer et al. 2020). Yet there is much person-to-person variation. Approximately $99\%$ of those showing symptoms will show them before day $14$, which explains the two week confinement period. Importantly, this analysis neglects infected people who never show symptoms. Since asymptomatic people are not usually tested, it is still not clear how many such cases there are or how long asymptomatic people remain infectious for.

### 4. How do N95 masks block SARS-CoV-2?

N95 masks are designed to remove more than $95\%$ of all particles that are at least $0.3$ microns ($\mu$m) in diameter (NIOSH 42 CFR Part 84). In fact, measurements of the particulate filtration efficiency of the N95 masks are often quoted as greater than $99\%$ of particles $< 1 \mu m$ (Regnasamy et al. 2017). SARS-CoV-2 is an enveloped virus $0.1 \mu m$ in diameter, so N95 masks are capable of filtering most free virions, but they do more than that. How so? Viruses are often transmitted through respiratory droplets produced by coughing and sneezing. Respiratory droplets are usually divided into two size bins, large droplets ($> 5 \mu m$) that fall rapidly to the ground and are thus transmitted only over short distances, and small droplets ($\leq 5 \mu m$ in diameter). Small droplets can evaporate into “droplet nuclei,” remain suspended in air for significant periods of time and could be inhaled. Some viruses, such as measles, can be transmitted by droplet nuclei (Teller et al. 2019). At present there is no direct evidence showing SARS-CoV-2 transmission by droplet nuclei. Rather, larger droplets are believed to be the main vector of SARS-CoV-2 transmission, usually by settling onto surfaces that are touched and transported by hands onto mucosal membranes such as the eyes, nose and mouth (CDC 2020). The characteristic diameter of large droplets produced by sneezing is $\approx 100\mu m$ (Han J. R. Soc. Interface 2013), while the diameter of droplet nuclei produced by coughing is on the order of $\approx 1 \mu m$ (Yang et al. 2007). Therefore, N95 masks likely protect against several modes of viral transmission.

### 5. How similar is SARS-CoV-2 to the common cold and flu viruses?

SARS-CoV-2 is a beta-coronavirus whose genome is a single $\approx 30$ KB strand of RNA. The flu is caused by an entirely different family of RNA viruses called influenza virus. Influenza viruses have smaller genomes ($\approx 10^4$ nucleotides) encoded in $8$ different strands of RNA, and they infect human cells in a different manner than coronaviruses. The “common cold” is caused by a variety of viruses, including some coronaviruses and rhinoviruses. Cold-causing coronaviruses (e.g. OC43 and 229E strains) are quite similar to SARS-CoV-2 in genome length (within $10\%$) and gene content, but different from SARS-CoV-2 in sequence ($\approx 50\%$ nucleotide identity) and infection severity. One interesting facet of coronaviruses is that they have the largest genomes of any known RNA viruses ($\approx 30$ KB). These large genes led researchers to suspect the presence of a “proofreading mechanism” to reduce the mutation rate and stabilize the genome. Indeed, coronaviruses have a proofreading exonuclease called Exonuc, which explains their very low mutation rates ($\approx 10^{-6}$ per site per cycle) in comparison to influenza ($\approx 3 \times 10^{-6}$ per site per cycle) (Sanjuan et al. 2010). This relatively low mutation rate will be of interest for future studies predicting the speed with which coronaviruses can evade our immunization efforts.

### 6. How much is known about the SARS-CoV-2 genome and proteome?

SARS-CoV-2 has a single-stranded positive-sense RNA genome that codes for $10$ genes ultimately producing $26$ proteins according to an NCBI annotation (NC_045512). How is it that $10$ genes code for $>20$ proteins? One long gene, orf1ab, encodes a polyprotein that is cleaved into $16$ proteins by proteases that are predominantly code for structural components of the virus: (i) the spike protein which binds the cognate receptor on a human or animal cell; (ii) a nucleoprotein that packages the genome; and (iii) two membrane-bound proteins. Though much current work is centered on understanding the role of “accessory” proteins in the viral life cycle, we estimate that it is currently possible to ascribe clear biochemical or structural functions to only about half of SARS-CoV-2 gene products.

### 7. What can we learn from the mutation rate of the virus?

Studying viral evolution, researchers commonly use two measures describing the rate of genomic change. The first is the evolutionary rate, which is defined as the average number of substitutions that become fixed per year in strains of the virus, given in units of mutations per site per year. The second is the mutation rate, which is the number of substitutions per site per replication cycle. How can we relate these
two values? Consider a single site at the end of a year. The only measurement of a mutation rate in a β-coronavirus suggests that this site will accumulate $\sim 10^6$ mutations in each round of replication. Each round of replication cycle takes $\sim 10$ hours, and so there are $10^7$ cycles/year. Multiplying the mutation rate by the number of replications, and neglecting the potential effects of evolutionary selection and drift, we arrive at $10^7$ mutations per site per year, consistent with the evolutionary rate inferred from sequenced coronavirus genomes. As our estimate is consistent with the measured rate, we infer that the virus undergoes near-continuous replication in the wild, constantly generating new mutations that accumulate over the course of the year. Using our knowledge of the mutation rate, we can also draw inferences about single infections. For example, since the mutation rate is $10^7$ mutations/cycle and an mL of sputum might contain upwards of $10^6$ viral RNAs, we infer that every site is mutated more than once in such samples.

8. How stable and infectious is the virus on surfaces? SARS-CoV-2 RNA has been detected on various surfaces several weeks after they were last touched (Moriarty et al. 2020). In the definitions we clarify the difference between detecting viral RNA and active virus. The probability of human infection from such exposure is not yet characterized as experiments to make this determination are very challenging. Nevertheless, caution and protective measures must be taken. We estimate that during the infectious period an uninfected person touches surfaces tens of times. These surfaces will subsequently be touched by hundreds of other people. From the basic reproduction number $R_0 > 2$, we can infer that not everyone touching those surfaces will be infected. More detailed bounds on the risk of infection from touching surfaces urgently awaits study.

**Glossary**

**Clinical Measures**

- **Incubation period**: time between exposure and symptoms.
- **Seroscversion**: time between exposure to virus and detectable antibody response.

**Epidemiological Inferences**

- **Latent period**: time between exposure and becoming infective.
- **Infectious period**: time for which an individual is infective.
- **Interval of half-maximum infectiousness**: the time interval during which the probability of viral transmission is higher than half of the peak infectiousness. This interval is similar to the infectious period, but applies also in cases where the probability of infection is not uniform in time.

**Viral Species**

- **SARS-CoV-1**: β-coronavirus that caused the 2002 SARS outbreak in China.
- **MERS**: The Middle East Respiratory Syndrome coronavirus outbreak beginning in Jordan in 2012.

**Viral Life-Cycle**

- **Eclipse period**: time between viral entry and appearance of intracellular virions.
- **Latent period (cellular level)**: time between viral entry and appearance of extracellular virions. Not to be confused with the epidemiological latent period described below.
- **Burst size**: the number of virions produced from infection of a single cell. More appropriately called "per-cell viral yield" for non-lytic viruses like SARS-CoV-2.
- **Viron**: a viral particle.
- **Polyprotein**: a long protein that is proteolytically cleaved into a number of distinct proteins. Distinct from a polypeptide, which is a linear chain of amino acids making up a protein.

**Human Biology**

- **Alveolar Macrophage**: immune cells found in the lung that engulf foreign material like dust and microbes ("professional phagocytes").
- **Pneumocytes**: the non-immune cells in the lung.
- **ACE2**: Angiotensin-converting enzyme 2, the mammalian cell membrane receptor that SARS-CoV-2 binds.
- **TMPRSS2**: Transmembrane protease, serine 2, a mammalian membrane-bound serine protease that cleaves the viral spike trimer after it binds ACE2, revealing a fusion peptide that participates in membrane fusion which enables subsequent injection of viral DNA into the host cytoplasm.
- **Nasopharynx**: the space above the soft palate at the back of the nose which connects the nose to the mouth.

**Notation**

Note the difference in notation between the symbol $\approx$, which indicates "approximately" and connotes accuracy to within a factor 2, and the symbol $\sim$, which indicates "order of magnitude" or accuracy to within a factor of 10.

More on definitions and measurement methods

**What are the meanings of $R_0$, "latent period" and "infectious period"?** The basic reproduction number, $R_0$ estimates the average number of new infections directly transmitted by an infectious person to other susceptible people. This number connotes that this refers to early stages of an epidemic, when everyone in the region is susceptible (i.e. there is no immunity) and no counter-measures have been taken. As geography and culture affect how many people we encounter daily, how much we touch them and share food with them, estimates of $R_0$ can vary between locales. Moreover, because $R_0$ is defined in the absence of countermeasures and immunity, we are usually only able to assess the effective $R (R_e)$. At the beginning of an epidemic, before any countermeasures, $R_e$ can be very large. Several days after severe cases become infectious themselves. This "latent period" is typically followed by several days of infectivity called the "infectious period." It is important to understand that reported values for all these parameters are population averages inferred from epidemiological models fit to counts of infected, symptomatic, and dying patients. Because testing is always incomplete and model fitting is imperfect, and data will vary between different locations, there is substantial uncertainty associated with reported values. Moreover, these median or average fit values only describe person-to-person variation. For example, viral RNA was detectable in patients with moderate symptoms for $> 1$ week after the onset of symptoms, and more than 2 weeks in patients with severe symptoms (ECDC 2020). Though detectable RNA is not the same as active virus, this evidence calls for caution in using uncertain, average parameters to describe a pandemic. Why aren't detailed distributions of these parameters across people published? Direct measurement of latent and infectious periods at the individual level is extremely challenging, as accurately identifying the precise time of infection is usually very difficult.

**What is the difference between measurements of viral RNA and infectious viruses?** Diagnosis and quantification of viruses utilizes several different methodologies. One common approach is to quantify the amount of viral RNA in an environmental (e.g. surface) or clinical (e.g. sputum) sample via quantitative reverse-transcription polymerase chain reaction (RT-qPCR). This method measures the number of copies of viral RNA in a sample. The presence of viral RNA does not necessarily imply the presence of infectious virions. Virions could be defective (e.g. by mutation) or might have been deactivated by environmental conditions. To assess the concentration of infectious viruses, researchers typically measure the "50% tissue-culture infectious dose" (TCID$_{50}$). Measuring TCID$_{50}$ involves infecting replicate cultures of susceptible cells with dilutions of the virus and noting the dilution at which half the replicate dishes become infected. Viral counts reported by TCID$_{50}$ tend to be much lower than RT-qPCR measurements, which could be one reason why studies relying on measurements (Moriarty et al. 2020) report the persistence of viral RNA for much longer times than studies relying on TCID$_{50}$ (van Doremalen et al. 2020). It is important to keep this caveat in mind when interpreting data about viral loads, for example a report measuring viral RNA in patient stool samples for several days after recovery (We et al. 2020). Nevertheless, for many viruses even a small dose of virions can lead to infection. For the common cold, for example, $0.1$ TCID$_{50}$ are sufficient to infect half of the people exposed (Couch et al. 1966).

**What is the difference between the case fatality rate and the infection fatality rate?** Global statistics on new infections and fatalities are pouring in from many countries, providing somewhat different views on the severity and progression of the pandemic. Assessing the severity of the pandemic is critical for policy making and thus much effort has been put into quantification. The most common measure for the severity of a disease is the fatality rate. One commonly reported measure is the case fatality rate (CFR), which is the proportion of fatalities out of total diagnosed cases. The CFR reported in different countries varies significantly, from 0.3% to about 10%. Several key factors affect the CFR. First, demographic parameters and practices associated with increased or decreased risk differ greatly across societies. For example, the prevalence of smoking, the average age of the population, and the capacity of the healthcare system. Indeed, the majority of people dying from SARS-CoV-2 have a preexisting condition such as cardiovascular disease or smoking (China CDC 2020). There is also potential bias in estimating the CFR. For example, a tendency to identify more severe cases (selection bias) will tend to overestimate the CFR. On the other hand, there is usually a delay between the onset of symptoms and death, which can lead to an underestimate of the CFR early in the progression of an epidemic. Even when correcting for these factors, the CFR does not give a complete picture as many cases with mild or no symptoms are not tested. Thus, the CFR will tend to overestimate the rate of fatalities per infected person, termed the infection fatality rate (IFR). Estimating the total number of infected people is usually accomplished by testing a random sample for anti-viral antibodies, whose presence indicates that the patient was previously infected. As of writing, such assays are not widely available, and so researchers resort to surrogate datasets generated by testing of foreign citizens returning home from infected countries (Verity et al. 2020), or epidemiological models estimating the number of undocumented cases (Li et al. 2020). These methods provide a first glimpse of the true severity of the disease.

**What is the burst size and the replication time of the virus?** Two important characteristics of the viral life cycle are the time it takes them to produce new infectious progeny, and the number of progeny each infected cell
produces. The yield of new virions per infected cell is more clearly defined in lytic viruses, such as those infecting bacteria (bacteriophages), as viruses replicate within the cell and subsequently lyse the cell to release a “burst” of progeny. This measure is usually termed “burst size.” SARS-CoV-2 does not release its progeny by lysing the cell, but rather by continuous budding (Park et al. 2020). Even though there is no “burst”, we can still estimate the average number of virions produced by a single infected cell. Measuring the time to complete a replication cycle or the burst size in vivo is very challenging, and thus researchers usually resort to measuring these values in tissue-culture. There are various ways to estimate these quantities, but a common and simple one is using “one-step” growth dynamics. The key principle of this method is to ensure that only a single replication cycle occurs. This is achieved by infecting the cells with a low MPN of virions, so that each infected cell produces a burst of virions that every cell gets infected, thus leaving no opportunity for secondary infections.

Assuming entry of the virus to the cells is rapid (we estimate 10 minutes for SARS-CoV-2), the time it takes to produce progeny can be estimated by quantifying the lag between inoculation and the appearance of new intracellular virions, also known as the “eclipse period.” This eclipse period does not account for the time it takes to release new virions from the cell. The time from cell entry until the appearance of the first extracellular virions, known as the “latent period” (not to be confused with the epidemiological latent period, see Glossary), estimates the duration of the full replication cycle. The burst size can be estimated by waiting until virion production saturates, and then dividing the total virion yield by the number of cells infected. While both the time to complete a replication cycle and the burst size may vary significantly in an animal host due to factors including the type of cell infected or the action of the immune system, these numbers provide us with an approximate quantitative view of the viral life-cycle at the cellular level.

References and excerpts
Note that for about 10 out of 45 parameters, the literature values are from other coronaviruses. We await corresponding measurements for SARS-CoV-2.

Size & Content
Diameter: (Zhu et al. 2020) - “Electron micrographs of negative-stained 2019-nCoV particles were generally spherical with some pleomorphism (Figure 3). Diameter varied from about 60 to 140 nm.”
Volume: Using diameter and assuming the virus is a sphere: 
\[ \text{Volume} = \frac{4}{3} \pi \times \left( \frac{\text{diameter}}{2} \right)^3 \]
Number of surface spikes trimers: 
\[ \text{Number of spikes trimers} = \frac{4}{3} \times \frac{\text{Volume}}{\pi \times (15 \text{ nm})^2} \]
Volume- Using diameter and assuming the virus is a sphere: 
\[ \text{Volume} = \frac{4}{3} \pi \times \left( \frac{\text{diameter}}{2} \right)^3 \]
Number of genes: 
\[ \text{Number of genes} = 30 \text{ for SARS}, 31 \text{ for SARS-CoV}, 29 \text{ for SARS-CoV-2} \]
Type: 
\[ \text{Type} = \frac{\text{Membrane (M; 222 aa)}}{\text{Envelope (E; 257 aa)}} \]

Genome
Type: VirusZone: SARS-CoV-2: SARS-CoV-2: Monopartite, linear ss(+)RNA genome
Genome length: (Wu et al. 2020) - Figure 2
Number of genes: (Wu et al. 2020) - “SARS-CoV-2 genome has 10 open reading frames (Fig. 2A)” or (Wu et al. 2020) - “The 2019-nCoV genome was annotated to possess 14 ORFs encoding 27 proteins”
Number of proteins: (Wu et al. 2020) - “By aligning with the amino acid sequence of SARS PPIa and analyzing the characteristics of restriction cleavage sites recognized by 3CLpro and PRo, we speculated 14 proteolytic sites of 3CLpro and PRo in SARS-CoV-2 PPIa (Fig. 2B).”

Table 1 and Figure 2: Number of surface spikes trimers: 
\[ \text{Number of spikes trimers} = \frac{4}{3} \times \frac{\text{Volume}}{\pi \times (15 \text{ nm})^2} \]

Replication Timelines
Virion entry into cell: (Schneider et al. 2012) - “Previous experiments had revealed that virus is internalized within 15 min” and “Within 10 min, some virus particles were internalized into vacuoles (arrow) that were just below the plasma membrane surface (Fig. 2, arrows).”
The observation at 15 min postinfection (p.i.), did not differ much from 10 min p.i. (Fig. 4a)

Eclipse period: (Schneider et al. 2012) - “SARS-CoV replication cycle from adsorption to release of infectious progeny takes about 7 to 8 h (data not shown)” and (Hammond et al. 2020) - “We observed the first burst of virions at 6 to 8 h postinfection.”
Acknowledgements
We thank Uri Alon, Niv Antonovsky, David Baltimore, Rachel Banks, Arren Bar Even, Naama Barkai, Molly Bassette, Menalu Berihoon, Biana Bernshtein, Pamela Bjorkman, Cecilia Blikstad, Julia Borden, Bill Burkholder, Griffin Chure, Lillian Cohn, Bernadeta Dadonaité, Emmie De wit, Ron Diskin, Ana Duarte, Tal Einav, Avigdor Eldar, Elizabeth Fischer, William Gelbart, Alon Gildoni, Britt Glausinger, Shmuel Gleizer, Dani Gluck, Soichiro Hirokawa, Greg Huber, Christina Hueschen, Amit Huppert, Shalev Itzkovitz, Martin Jonikas, Leeat Keren, Gilmor Keshet, Marc Kirschner, Roy Kishony, Amy Kistler, Liad Levi, Sergei Maslov, Adi Millman, Amir Milo, Elad Noor, Gal Offir, Alan Perelson, Steve Quake, Itai Raveh, Andrew Rennekamp, Tom Roeschinger, Daniel Rokhsar, Alex Rubinsteyn, Gabriel Salmon, Maya Schuldiner, Eran Segal, Ron Sender, Alex Sigal, Maya Shamir, Ark Shams, Mike Springer, Adi Stern, Noam Stern-Ginossar, Lubert Stryer, Dan Tawfik, Boris Veytsman, Aryeh Wides, Tali Wiesel, Anat Yarden, Yossi Yovel, Dudi Zeevi, Mushon Zer Aviv, and Alexander Zlokapa for productive feedback on this manuscript. Figure created using Biorender.