The total number and mass of SARS-CoV-2 virions in an infected person

Ron Sender1*, Yinon M. Bar-On1*, Avi Flamholz2, Shmuel Gleizer1, Biana Bernstein1,5, Rob Phillips3,4, Ron Milo1

1 Weizmann Institute of Science, Rehovot 7610001, Israel
2 University of California, Berkeley, CA 94720, USA
3 California Institute of Technology, Pasadena, CA 91125, USA
4 Chan Zuckerberg Biohub, 499 Illinois Street, SF CA 94158, USA
5 Present address: Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02138, USA

* Equal contribution

Quantitatively describing the time course of the SARS-CoV-2 infection within an infected individual is important for understanding the current global pandemic and possible ways to combat it. Here we integrate the best current knowledge about the abundance of potential SARS-CoV-2 host cells and typical concentrations of virions in bodily fluids to estimate the total number and mass of SARS-CoV-2 virions in an infected person. We estimate that each infected person carries \(10^9\)-\(10^{11}\) virions during peak infection, with a total mass of about 1 \(\mu\)g-0.1 mg, which curiously implies that all SARS-CoV-2 virions currently in the world have a mass of only 0.1-1 kg. Knowledge of the absolute number of virions in an infected individual can put into perspective parameters of the immune system response, minimal infectious doses and limits of detection in testing.

Estimating key biological quantities such as total number and mass of cells in our body or the biomass of organisms in the biosphere in absolute units helps develop a deep quantitative perspective and improves our intuition and understanding of the living world (Moran et al. 2010; Sender et al. 2016a; Sender et al. 2016b; Bar-On et al. 2018). Such a quantitative perspective could help the current intensive effort to study and bring under control the COVID-19 pandemic by clarifying a myriad of key numbers such as minimal infectious doses and limits in testing. We have recently compiled quantitative data at the virus level as well as at the community level to help communicate state-of-the-art knowledge to both the public and researchers and provide them with a quantitative toolkit to think about the pandemic (Bar-On et al. 2020). Here we leverage such quantitative information to estimate the total number and mass of SARS-CoV-2 virions present in an infected individual during the peak of the infection.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Viral concentrations and viral particles are being measured in several different ways resulting in great differences in values and their meanings. Here, we report our estimates in two ways. First via viral RNA copies, as they are measured by RT-PCR, and representing the amount of viral RNA produced, including defective and deactivated virions. The second way is by infectious virions, excluding the defective and deactivated virions, measured in units of fifty-percent tissue culture infective dose (TCID50) by virus titer (endpoint dilution assay quantifying the amount of virus required to kill 50% of infected host cells). Thus, we try to estimate both the total amount of viral particles through RNA copies and the number of infectious virions through TCID50 measurements.

To estimate the total number of virions present in an infected individual at the peak of the infection, we rely on several studies which measured the concentration of virions (measured in genome copies per gram tissue) in tissues of rhesus macaques, after infections with SARS-CoV-2 (Munster et al. 2020; Williamson et al. 2020). The viral concentrations were analyzed for samples of all the relevant tissues of the respiratory, digestive and immune systems, and were given per gram of tissue. An estimate for the total number of virions can be obtained from these measurements by multiplying the viral concentration of each tissue by the total tissue mass (ICRP 2002; Snyder et al. 1975). The lungs are the largest tissue in terms of mass ($M_{\text{lungs}} \approx 1\, \text{kg}$) and had the highest viral concentration ($C_{\text{lungs}}^{\text{virions}}$) and thus contribute the most to the overall estimate, with

$$C_{\text{virions}}^{\text{lungs}} \times M_{\text{lungs}} = N_{\text{virions}}$$

Other tissues, like the nasal mucosa, larynx, bronchial tree and adjacent lymph nodes all have a combined mass of $\approx 100\, \text{g}$ and maximal concentrations of $10^6$-$10^7$ RNA copies/ml and hence contribute at most an additional 10% to the estimate based on the lungs (Figure 1).

Another study (Rockx et al. 2020) measured the viral concentrations in tissues taken from infected rhesus macaques, a few days after inoculation. However, this study reports its values in units of TCID$_{50}$ which give an assessment of the concentration of infectious viruses. The study reports much smaller maximal values of $10^3$-$10^4$ TCID$_{50}$/ml for lung tissue. Combining their concentration with the volume of an adult human lungs, we get an estimate of $10^5$-$10^7$ TCID$_{50}$ virions in an adult, compared with $10^9$-$10^{11}$ RNA copies, estimated from (Munster et al. 2020)(Figure 1). This large difference appears to
be representative of the use of viral titers instead of viral loads, as a similar difference of 4-5 orders of magnitude is observed in bronchoalveolar lavage fluid measurements in rhesus macaques by (Williamson et al. 2020) and in nasopharyngeal swabs taken from 454 human participants (Quicke et al. 2020).

Based on this line of evidence, we estimate the total number of virions in an infected individual during peak infection at $10^{9}$-$10^{11}$ RNA copies, or $10^{5}$-$10^{7}$ TCID$_{50}$.

**Calculating the total number of cells infected with SARS-CoV-2**

We can leverage our estimate of the total number of virions in the body of an infected individual to estimate the size of the population of cells which are infected with the virus during peak infection. As shown in Figure 2, to estimate the total number of infected cells, we need to know how many virions are found within each infected cell.
In order to determine the number of virions within an infected cell at any given time we rely on several lines of evidence. The first is data regarding the burst size of other betacoronaviruses (as we are not aware of measurements for SARS-CoV-2). The burst size is defined as the total number of viruses produced by an infected cell throughout its lifetime, in analogy with bacterial lytic infections. Two previous studies suggested virion yield of either 10-100 or 600-700 infectious virions (in Plaque Forming units - PFUs) (Robb and Bond 1979; Hirano et al. 1976). The time scale that is relevant for this yield is a few days (Chu et al. 2020). Thus, we can assume that by the first estimate at any given moment there are few to tens infectious virions that reside in each infected cell, and tens to hundreds of infectious virions by the second. The second line of evidence concerns the density of viral particles, within a single cell. Several studies have used transmission electron microscopy (TEM) to characterize the dynamics of viral infection within cells (Imai et al. 2020; Kim et al. 2020; Klein et al. 2020; Ogando et al. 2020). Using seven TEM scans taken from those studies we estimated that the density of viral particles within infected cells is $10^5$ viral particles per 1 pL. As the human cells targeted by SARS-CoV-2 have a volume of $\approx 1$ pL (resulting in a cellular mass of $\approx 1$ ng) (Stone et al. 1992; Crapo et al. 1982), the TEM scans indicate that at a given moment there are $\approx 10^5$ viral particles within a single cell. Those viral particles include defective virions, and as we saw earlier, it is reasonable to assume that only 1 in $10^4$ of them is infective. Thus TEM scans indicate that $\approx 10$ infectious virions reside inside a cell in a given moment. We can perform a sanity check using mass considerations to see that our estimate of the number of viral particles is not beyond the maximal amount that is feasible. Each viral particle has a mass of $\approx 1$ fg (Bar-On et al. 2020). Hence, $10^5$ viral particles weigh $\approx 0.1$ ng, about 10% of the total mass of the cell (and more than a third of its dry weight). That means that it is not reasonable to assume that there are more than $10^5$ viral particles or 10 infectious virions within a cell at any given moment. Following those lines of evidence we conclude that at a given moment there are $\approx 10$ infectious virions residing inside a cell, and that the overall yield of an infected cell is 10-100 infectious virions over a few days.

$$\frac{N_{\text{infectious virions}}}{10^5 \text{ virions per cell}} = \frac{N_{\text{infected cells}}}{10^4 - 10^6 \text{ cells}}.$$  

How does this estimate stack up against the number of potential host cells for the virus? The best-characterized route of infection for SARS-CoV-2 is through cells of the respiratory system, specifically the pneumocytes (a total of about $\approx 10^{11}$ cells in the body), alveolar macrophages ($\approx 10^{10}$ cells) and the mucus cells in the nasal cavity ($\approx 10^9$ cells).
cells) (Stone et al. 1992; Crapo et al. 1982). Other cell types, like enterocytes (gut epithelial cells) can also be infected (Lamers et al. 2020). Our best estimate for the size of the total potential pool of host cells is thus $\sim 10^{11}$ cells, and the fraction of cells infected during peak infection therefore represents a small fraction of this potential pool (1 in $10^5$-$10^7$).

![Figure 2: Estimation of the number of infected cells and their fraction out of the potential relevant host cells.](image)

**Virion production rate and the total number of virions produced throughout an infection**

In addition to analyzing the state of an infected individual during peak infection, we can also ask what is the total number of virions produced over the course of an infection, as well as what is the rate of virion production inside a human host. First, in order to estimate the total number of virions produced during an infection, we rely on the average number of viruses an infected cell yields, which is about 10-100 (based on measurements from different kinds of betacoronavirus (Robb and Bond 1979; Hirano et al. 1976), and our estimate of the number of virions within a single cell at any given moment). The total number of virions produced over the course of an infection can be
estimated by multiplying the burst size by the total cumulative number of infected cells as
\[
N_{\text{cells}} \times Y_{\text{virions per cell}} = \hat{N}_{\text{total virions throughout disease}}
\]
\[
10^4 - 10^6[\text{cells}] \times 10 - 100[\text{virions per cell}] = 10^5 - 10^8 [\text{infectious virions, TCID50}]
\] (3)

This value exceeds the peak number of virions we estimated by about an order of magnitude. In terms of RNA copies, this translates into $10^9$-$10^{12}$ RNA copies.

In order to estimate the rate of production of virions we divide the burst size by the duration it takes the infected cells to produce that burst size, known to be a few days (Bar-On et al. 2020). We thus estimate that each infected cell produces ~10 infectious virions per day on average. To estimate the maximal production rate for all infected cells, we multiply this production rate per single infected cell by the $10^4$-$10^6$ infected cells. Thus the maximal production rate of virions is given by:
\[
N_{\text{cells}} \times Q_{\text{virions per cell per day}} = P_{\text{virions per day}}
\]
\[
10^4 - 10^6[\text{cells}] \times 10 [\text{virions per cell per day}] = 10^5 - 10^7 [\text{infectious virions (TCID50) per day}].
\] (4)

While the estimates were performed using a reference value for the lung mass taken from adult men, they can be generalized for women and children. We rely on the multiplication of the viral concentration in the lungs and the total mass of the lungs. Reference values for the lung mass show a value smaller by 20% for women, and 75%-25% for children aged 5-15 years (ICRP 2002). Although COVID-19 is known to affect adult men more than women and children (Salje et al. 2020; WHO 2020), there is scarce information regarding difference in viral concentrations across gender and age. One preprint (Jones et al. 2020) suggests that viral concentration in children is lower by up to an order of magnitude, but the change they measured is not consistent across the entire age range. Assuming the change in measured viral load represents a similar change in viral concentration in the lung tissue, and combining the concentrations with the reduced lung mass, we get that the number of virions in an infected woman is similar to that estimated for men, and that an infected child is probably carrying an order of magnitude less virions.
Discussion

To provide some context for our estimate of the total number of virions per infected individual, we can compare it with the total of $3 \times 10^{13}$ human cells in the body (Sender et al. 2016b). We find that even for the higher end of the range of estimates for viral particles there is still a 10-100 factor in favor of the human cells. A more functional comparison is with the number of antibodies the body produces to combat the viral infection. Total IgG antibody levels specific for SARS-CoV-2 spike protein were measured 3 weeks following symptom onset showing a concentration in the serum on the order of $\sim 10\ \mu g/mL$ (Iyer et al. 2020). Only a fraction of $\approx 5\%$ of the total anti-spike IgG antibodies has the capacity to neutralize the virus ($f_{\text{neutralizing}}$) (Rogers et al. 2020).

Combining the concentration of neutralizing IgG antibodies with a mean IgG molecular weight ($MW_{\text{IgG}}$) of 150 kDa (Janeway et al. 2001) we can derive the number of neutralizing antibody molecules in an mL of serum:

$$C_{\text{IgG}} \times f_{\text{neutralizing}} \times \frac{1}{MW_{\text{IgG}}} \times N_{\text{Avogadro}} = \hat{N}_{\text{neutralizing}}$$

$$10^{-5}\left[\frac{\text{g IgG}}{\text{mL}}\right] \times 5\% \left[\frac{\text{neutralizing}}{\text{IgG}}\right] \times \frac{1}{150,000}\left[\frac{\text{mol}}{\text{g}}\right] \times 6 \times 10^{23}\left[\frac{\text{molecules}}{\text{mol}}\right] = 3 \times 10^{12}\left[\frac{\text{neutralizing molecules}}{\text{mL}}\right].$$

Combining this estimate with the measurement of viral concentration within the lung tissue (Munster et al. 2020) and accounting for 30-40 spike trimers on each SARS-CoV-2 virion (Yao et al. 2020, Turoňová et al. 2020) we can estimate the ratio of neutralizing antibodies to viral spike proteins:

$$\frac{\hat{N}_{\text{neutralizing}}}{C_{\text{virions}} \times N_{\text{spike proteins}}} = R_{\text{neutralizing antibodies/spike proteins}}$$

$$= 10^{3} - 10^{5}\left[\frac{\text{neutralizing antibodies}}{\text{spike protein}}\right].$$

Previous work on morphologically similar viruses such as influenza and flavivirus found that a ratio of 1 neutralizing antibody to 2-4 receptor-binding proteins was sufficient to neutralize binding of a virion to its cellular receptor (Taylor et al. 1987; Pierson and Diamond 2015). Taken at face value, our estimate seems to suggest an excess of neutralizing antibody molecules, even if we consider that the antibodies concentrations 30-in the lung tissue is lower than that of the blood, and the extensive glycosylation patterns found on the spike protein that shield many of its epitopes (Turoňová et al. 2020) from neutralizing antibody binding and thus decrease the efficiency of neutralization (Schön et al. 2020). However, it is important to remember that in order to estimate the effectiveness of antibody neutralization, we need to estimate the fractional occupancy of the viral epitopes by antibodies. This fractional occupancy is determined by the strength of the binding of the neutralizing antibodies to the viral particles, given...
by the dissociation constant $K_d$ (Pierson and Diamond 2015). Following the first order relation:

$$\text{fraction bound by nAb} = \frac{[nAb]}{[nAb] + K_d}$$  \hspace{1cm} (7)

As the dissociation constants are mostly in the range of 1-10 nM (Chi et al. 2020; Seydoux et al. 2020) we get:

$$[nAb] = 3 \times 10^{12} \left[ \frac{\text{neutralizing molecules}}{\text{mL}} \right] \times \frac{1}{6 \times 10^{29}} \left[ \frac{\text{mol}}{\text{molecules}} \right] \times 1000 \left[ \frac{\text{mol}}{\text{L}} \right] = 3 \times 10^{-9} \left[ \frac{\text{mol}}{\text{L}} \right] = 3 \left[ nM \right]$$

$$\text{fraction receptors bound by nAb} = \frac{3 \left[ nM \right]}{3 \left[ nM \right] + 1 - 10 \left[ nM \right]} = 25\% - 75\%$$

Thus, the concentration of antibodies is needed to ensure that enough of the epitopes are bound, regardless of the high ratio between the number of neutralizing antibodies and viral particles.

The total number of virions and their tissue concentration can also shed some light on the mechanism of action of antiviral drugs, such as Remdesivir. In their study with rhesus macaques (Williamson et al. 2020) measured a median concentration of $\approx 3$ nmol of the activated drug within 1 gram of homogenized lung tissue, after six days of treatment. This is equal to $2 \times 10^{15}$ molecules per 1 gram of tissue, about 7-8 orders of magnitude more than the concentration of virions within the tissue. It appears that this ratio of antiviral drug to viral concentration is needed as the reported concentrations are also close to the measured EC50 (50% effective concentration) of $\approx 10 \mu$M (Choy et al. 2020).

One can contextualize these estimates using an absolute mass perspective. Each virion has a mass of $\approx 1$ fg (Bar-On et al. 2020). Therefore even when the body carries $10^9$-$10^{11}$ viral particles, these have a mass of only about 1-100 μg, i.e. 1-100 times less than the mass of a poppy seed. Taking a global view across all of humanity and assuming a total of 10-100 million infected people at a given time (including the undetected), we arrive at a total mass of all the virions residing in humanity to be on the order of 0.1-1 kg. Furthermore, using the total number of viral particles produced throughout an infection ($10^9$-$10^{12}$ per person) we can derive the total mass of all the SARS-CoV-2 viral particles ever produced throughout this current pandemic. Assuming the total number of infected people actually will be 1 billion people we find the total mass of viral particles produced for such widespread infection to be on the order of

$$N_{\text{infected people}} \times M_{\text{viral particle}} \times N_{\text{viral particles produced per person}} = M_{\text{all viral particles produced}}$$  \hspace{1cm} (9)
We have shown how our estimate for the total number of virions could be used to estimate currently unknown quantities such as the fraction of potential host cells that are infected during peak infection. Single-cell RNA-sequencing studies (Sungnak et al., 2020; Zeigler et al., 2020; Lukassen et al., 2020) indicate that a few percent of the cells in the lungs and airways express ACE2 (angiotensin-converting enzyme 2) and TMPRSS2 (Transmembrane protease, serine 2), the receptor and main protease SARS-CoV-2 relies on for infecting cells. Most of the cells that have been found to express both are type 2 pneumocytes. While these results might be biased due to drop-out effects in measurements of only a few molecules (Valyaeva et al., 2020; Zeigler et al., 2020), it is still reasonable that 1%-10% of the lung and airway cells contain the necessary receptor to be infected by SARS-CoV-2. This number is several orders of magnitude higher than our estimate for the total number of infected cells during peak infection - a fraction of 1 in $10^5$-$10^7$ of the potential lung cells. This suggests that out of the cells expressing both ACE2 and TMPRSS2, only a small fraction, e.g. $10^{-3}$-$10^{-5}$, get infected by the virus.

There is very high variability in viral loads, exceeding 5 orders of magnitudes, as can be seen from samples taken from the upper respiratory system (Jacot et al., 2020). This wide variation reflects the difference between people as well as differences in viral load through the progression of infection within an infected individual (He et al., 2020). In our estimates we consider only a smaller range of viral concentration, corresponding to the center of the distribution of the measured values (interquartile range - between the quantiles 25%-75%). Thus, extreme cases could exceed the range provided by additional two orders of magnitudes, reaching values of $10^{13}$ viral particles in a single person at the peak of infection, while up to 10% of the cells expressing both ACE2 and TMPRSS2 are infected.

Major knowledge gaps that we identified are the virion yield per infected cell which is known only from a few studies on different kinds of betacoronavirus from over 40 years ago (Robb and Bond 1979; Hirano et al. 1976).

The global impact of SARS-CoV-2 is clearly evident from its influence on public health and the global economy. Contrasting these global effects with the very moderate cumulative mass of these viruses highlights the limits of our day-to-day intuition in understanding and combating this pandemic and the need to rely on sound quantitative information rather than gut feelings. Having better quantitative information on the process of infection at the cellular level, the intra-host level and the inter-host level will hopefully empower researchers with better tools for thinking about ways to combat COVID-19 and curb its spread and that of future pandemics.
Acknowledgments

Gidon Eshel, Shai Fuchs, Thierry Mora, Eran Segal, Maya Shamir, Ziv Shulman, Harinder Singh, Itai Benhar, Aleksandra Walczak. Figure created using Biorender.

Funding: This research was supported by the European Research Council (Project NOVCARBFIX 646827), Israel Science Foundation (Grant 740/16), Beck-Canadian Center for Alternative Energy Research, Dana and Yossie Hollander, Ullmann Family Foundation, Helmsley Charitable Foundation, Larson Charitable Foundation, Wolfson Family Charitable Trust, Charles Rothschild, Selmo Nussenbaum, Miel de Botton (R.M.), the National Institutes of Health (1R35 GM118043-01 (Maximizing Investigators Research Award)) (R.P.) the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center and by a research grant from Madame Olga Klein – Astrachan (R.S.). R.M. is the Charles and Louise Gartner Professional Chair. Y.M.B. is an Azrieli Fellow

References


Lukassen, S., Chua, R.L., Trefzer, T., Kahn, N.C., Schneider, M.A., Muley, T., Winter,


Schön, M.P., Berking, C., Biedermann, T., Buhl, T., Erpenbeck, L., Eyerich, K., Eyerich,


