Epidemic reproduction numbers and herd immunity

B V Girdler-Brown,[1] MB ChB, MMed, FCPHM (SA), FFPH (UK); **J Murray,**[2] BA, MB BCh, FFPath (SA)

[1] *School of Health Systems and Public Health, Faculty of Health Sciences, University of Pretoria, South Africa*
[2] *School of Public Health, University of the Witwatersrand, Johannesburg, South Africa*

Corresponding author: B V Girdler-Brown (brendan.girdler-brown@up.ac.za)

In this educational article, we explore the meanings and the applications of the reproductive numbers, namely the basic reproductive number R_0, and the effective reproductive number R. We also give examples of how these numbers may be affected by the values of β, κ and D, and we go on to explore the relationship of the R_0 value to the concept of herd immunity. Finally, we discuss how these concepts may contribute to the control of infectious disease outbreaks.

During the COVID-19 pandemic, there has been much written in the scientific and popular media about the reproduction numbers, herd immunity and modelling of the epidemic at country level. It is not unusual to find lawyers, economists and business people speaking and writing at length about the reproduction numbers and the concept of herd immunity in the popular press, as well as on television. It is important, therefore, that medical professionals, and especially those in the field of public health, are well informed about the meanings of these terms and concepts. This educational article explains the meanings of the reproduction numbers and of the herd immunity concept. It also discusses the relationships between these numbers and the herd immunity level. Finally, there is a brief discussion about how the reproduction numbers are estimated, and why they are important for epidemic management.

The basic reproduction number, $R0$

The definition of $R0$

Dietz[1] has traced the origin of the concept of the basic reproduction number to 1886, when it was developed for use in the field of demography. The earliest reference to the symbol R_0 that Dietz was able to find was by Lotke[2] in 1939.

Giesecke[3] has defined R_0 as 'the average number of persons directly infected by an infectious case during her entire infectious period, when she *enters a totally susceptible population*'. One might add to this, 'and in the absence of any disease-specific control interventions having been implemented'.

Giesecke[3] refers to this as the 'basic reproductive rate'. Others might refer to R_0 as the basic reproductive ratio,[4] and still others as the basic reproduction number. R_0 is dimensionless and is expressed per case, rather than per unit of time, and so it is best not to refer to R_0 as a rate. Although 'ratio' is also perfectly acceptable, we have chosen, in this article, to refer to R_0 as the basic reproduction number.[5]

The determinants of $R0$

The value of R_0 is related to the risk of transmission per contact, β, the number of contacts per unit of time, κ, and the duration of infectiousness, D. These relationships give rise to the following formula:[3]

$$R_0 = β \times κ \times D$$

Hence if the risk of transmission per contact is 0.1, the number of contacts is 10 per day and the duration of infectiousness is 4 days, then the basic reproduction number = $0.1 \times 10 \times 4 = 4$.

With regard to the valuation of β, the concept of a 'contact' varies with the disease and the type of contact. Giesecke[3] cites, as an example, the transmission of HIV. He points out that β for HIV transmission, during a contact between an infected and an uninfected person, is 0 for a single episode of shaking hands, between 0.001 and 0.1 for one episode of unprotected sexual intercourse, and virtually 1 for a blood transfusion with infected blood.

For SARS-CoV-2, the virus that causes the disease COVID-19, the β value may be different for close contact with a sneezing infectious person, on the one hand, and for touching an escalator rail in a busy shopping centre that may carry viable virus in droplets, deposited a few minutes before, on the other hand.

Concerning the value of κ, this parameter value will differ depending on the route of transmission. For a sexually transmitted disease, for example, merely being in a person's presence will not constitute an effective contact for the transmission of the infectious agent. For COVID-19 disease, there are, broadly, two recognised kinds of contact: being within close distance of an infectious person; and touching a surface that is contaminated with viable SARS-CoV-2 virus.

For any infectious disease, the value of D may include a period of time prior to the onset of symptoms, and may terminate before, or even after, the point where an infected person has fully recovered. In addition, there may be asymptomatic infected individuals who are also infectious for a period of time.

The value of R_0 is only meaningful for a fairly homogeneous group of wholly susceptible people, with this group's unique mixing and behaviour patterns. An 'average' value of R_0 is less useful for a heterogeneous community with wide variations in social and behavioural circumstances.

It should be noted that R_0 refers to the reproduction number in a wholly susceptible population, and in the absence of any specific control measures being implemented.

The value of R_0 gives us a clear indication of the relative potential infectiousness of the infectious agent *ceteris paribus*. Therefore, it is customary to estimate R_0 at the beginning of an epidemic.

In addition, we note the following:
- If the value of R_0 is >1 then each case will infect more than one person; there will be an epidemic
- If the value of R_0 = 1 then each case will infect one person; there will be new cases, but the number of prevalent cases will not change much over time; and
- If the value of R_0 is <1 then it is unlikely that the disease will secure a foothold in the population.

The value of R_0 differs from disease to disease, and may vary considerably between different communities and countries for the same disease, even when the same method is used to estimate R_0. Hence it is important to estimate R_0 for individual countries or subpopulations, rather than to rely on an average number, or one that was estimated for a very different country.[6]

This is because the values of the disease-specific parameters are very variable, depending on local customs, social habits and living/commuting arrangements.

As the value of R_0 may vary between communities, owing to differing social circumstances and behavioural patterns (notably the patterns of public transport usage), an average overall value of R_0 is not likely to be very meaningful for a country such as South Africa (SA), where there is much heterogeneity in lifestyles.

Even within a fairly homogeneous community, the assumption is made, when estimating R_0, that mixing is occurring in a fairly random manner. In practice, this is often not the case.

As we shall see, later, R_0 is also useful for estimating required vaccine efficacy, and vaccine coverage levels, in order to stop transmission of an infectious disease.

The estimation of R_0

The precise estimation of R_0 is not straightforward. As R_0 is the reproduction number at a time when the entire population is susceptible, it must be estimated near the start of an epidemic. At the start of an epidemic, case numbers may be low, or highly variable, and the epidemic may even expire due to chance, especially if R_0 is close to unity. The confidence intervals for R_0 are likely to be quite wide.

For novel infectious agents it is sometimes assumed that the entire population is susceptible at the start of an outbreak; however, there may be some cross-immunity prevalent in the community owing to past exposure to related pre-existing strains, as is known to be the case with influenza.[7]

If the values of β (the risk of transmission per contact), κ (the number of effective contacts per unit of time) and D (the duration of infectiousness) have been established, in an intervention-free setting, through empirical research, then R_0 may be estimated using the formula β × κ × D.

Such detailed information will not be available for an epidemic of a novel infectious agent. Estimations of R_0 may, then, be made from knowledge about the epidemic curve and the average generation time, as well as the generation time distribution.[8] The generation time is the time between the date of infection in each primary case and the date of infection in each subsequent, secondary, case.

The dates of infection may be very difficult to establish for most case pairs (primary and secondary). Therefore, the serial interval is often used as a surrogate for the generation time. The serial interval is the length of time between the onset of symptoms of a primary case and the onset of symptoms in its secondary case(s). The computed value of R_0 is very sensitive to assumptions about the shape of the serial interval distribution; great care must be taken, therefore, to determine the distribution in a given setting.[8,9]

Du *et al.*[10] have given a description of how they estimated the serial interval for COVID-19 in China. They estimated this interval to be 3.96 days, on average (standard deviation = 4.75). For 59/468 cases the onset of symptoms occurred in the secondary case before they were evident in the primary case. This suggests there is some presymptomatic transmission taking place. Since they had some negative values for the serial interval, the frequency histogram for their serial interval times resembles a normal distribution.

White and Pagano[11] have described a maximum likelihood estimation method for R_0. The required inputs are the epidemic curve data, the estimated serial interval and the type of distribution for the serial interval data.

There is statistical code available that permits the estimation of R_0, using R statistical software, if the epidemic curve, the serial interval and its frequency distribution are known.[12]

Giesecke[3] has cited a rough estimation, 'approximate' formula that may be used for the estimation of R_0:

$$R_0 = 1 + L/A$$

L is the average life span of the population of interest, and A is the average age of those with the disease at the time of infection. Applying this rough method to the COVID-19 epidemic in SA, the value of L is currently 64.12.[13]

The average age of cases of COVID-19 in SA is reported in 5-year age categories. An estimated average age has been calculated from these official tables on 28 May 2020 to be 39.41 years old.[14]

Applying the 'rough estimate' formula for COVID-19 in SA:

$$R_0 \cong 1+(64.12/39.41) = 2.63$$

This is a very approximate estimation from a rough-and-ready method. In addition, the average ages used in this calculation were from the first 85 days of the epidemic, and included an unknown number of imported cases at the beginning. An earlier report, from the popular media, cited a mean age of 41.8 years for the first

274 cases that included an unknown number of imported cases.[15] If this earlier value is used, the calculated 'rough' value of R_0 is 2.53.

The National Institute for Communicable Diseases (NICD) has published, on its website, the results of estimations of COVID-19 values for R_0.[16] The analysts constructed the epidemic curve using the dates of symptoms onset, where known, and imputed values for these where the dates of symptoms onset were not recorded. They assumed a gamma distribution for the serial interval data, and used White and Pagano's[11] maximum likelihood estimation method.

They obtained a value for R_0 for the whole of SA of 2.07 (95% confidence interval 1.69 - 2.50). At a subnational level, there was considerable variation in the estimates of R_0. These ranged from 1.7 - 2.5, measured approximately 2 - 3 weeks after introduction of the initial case(s).

Subsequently, the reproductive number would be expected to fall, following the possible spread of acquired immunity and/or the implementation of various control interventions that would be expected to affect the values of β, κ and D.

This changing value of R is sometimes depicted by R, sometimes by R_t (the 'time-varying reproduction number') and sometimes by R_e (the 'effective reproduction number'). In this article, we refer to it as the effective reproduction number and depict it by R with no subscript.

The effective reproduction number, *R*
The definition of *R*
R, the effective reproduction number, may be defined as the average number of people infected by each new case of the disease. Unlike R_0, which refers to the number of people infected by each case in a wholly susceptible population, and in the absence of any control interventions, R is the number infected when there is some immunity and/or some intervention present.

From this definition it is clear that, as immunity is acquired in the population during the course of the epidemic, or as control interventions are implemented, the effective reproduction number can be expected to fall.

Classically, for diseases that result in immunity, the value of R declines as more and more people become immune and are therefore no longer susceptible. This spread of immunity affects the number of susceptible contacts available to contract the disease.

There are, in addition, disease control interventions that, if effectively implemented, will also reduce the value of R. A falling value for R may, therefore, be used to evaluate the impact of these interventions. The aim of the control measures is to reduce the value of R to <1. When this happens, the number of new cases will decline over time. If the value of R is maintained <1 for a sufficient length of time, there will eventually be no new cases.

The value of R is affected by the values of β, κ and D. Therefore, any interventions that will affect these parameter values will also affect the value of R.

Changing values of β
The value of β may be reduced for infectious diseases by, for example, frequent washing of hands (in the reduction of seasonal flu transmission[17]), or the use of tenofovir-containing vaginal gel (for reducing the transmission risk for HIV).[18] For the COVID-19 epidemic, the following interventions are some that have been advocated to reduce the value of β:
- the universal wearing of face masks in public places
- the frequent washing or sanitising of hands
- avoidance of touching one's face.

Changing values of κ
The number of contacts per unit of time, κ, depends on the definition of what would constitute an effective contact (i.e. a contact that has the potential to result in transmission). For SARS-CoV-2 there are at least two recognised types of contact that would be regarded as effective, namely being within sneezing distance of an infectious person who is coughing, sneezing, speaking etc., and touching a recently deposited droplet containing viable virus. The value of κ may be reduced by:
- reducing human interactions by discouraging unnecessary social mixing at work, on business, at worship, at leisure, on shopping sprees or in travel (i.e. the 'lockdown')
- suspending unnecessary travel between centres of population;
- frequent sanitising of public spaces, especially bathrooms, lifts, escalator railings, tables, taxis, buses, trains etc.
- enforcing 'social distancing rules' in queues, transport, at places of worship, etc.

Changing values of D
Finally, the value of D, the duration of infectiousness, may be reduced effectively by:
- isolating known or suspected individuals until they are no longer infectious
- placing known contacts of infected individuals in quarantine;
- testing of symptomatic individuals in an effort to isolate those infected as soon as possible
- testing of asymptomatic close contacts of known cases in an effort to detect cases early, while still asymptomatic
- in some high-risk situations such as mine work, care homes and among healthcare workers and other institutional staff, universal testing in order to pre-empt institutional outbreaks by isolating those found to test positive.

If the prevalence of true infections is low, say, <1 per thousand, then the majority of cases that are detected by universal testing, if the number being tested exceeds 2 000, will be false positive results, even if the specificity of the test is 99.9%. The lower the incidence and specificity, and the higher the number being tested, the greater the problem of false positives will become. Universal testing of the general public will not offer significant benefits, and will deplete the number of test kits available for more targeted testing.

In the future, it is possible that new treatments for the COVID-19 illness may be found that will reduce the duration of infectiousness among those infected.

The serial estimation of *R* during an epidemic
As the number of recovered, and hence immune, members of the community increases, the probability that an infectious

person will come into contact with a susceptible person, just by chance, will decline. As a result, transmissions will decline and the effective reproduction number will also decline. Eventually, the value of R will fall below unity, and then the number of new cases will start to fall (although the total number of historical cases will continue to rise).

However, the value of R can also be reduced by control measures and interventions, such as those already discussed for the COVID-19 epidemic, including restricting people to their homes except for essential outside visits for food purchases and to access medical care. Tracking the value of R will then permit public health workers and policy-makers to assess the impact of their control efforts. It will also help policy-makers to decide when, and how quickly, to ease restrictions on movement. This needs to be done cautiously for COVID-19, as this pandemic is a novel experience for the world. There should be constant monitoring of the value of R so that, if the relaxation of restrictions results in a rise in R to values >1, then some restrictions may need to be reintroduced to bring the value of R <1 again. This tracking of R is thus a vital component in the management of the COVID-19 pandemic.

If the value of R is maintained at 1, then the number of new cases reported each day will remain unchanged. For example, if there were 1 000 new cases a day at the time when R reached unity, there will continue to be 1 000 new cases a day from then onwards (unless the value of R dips to <1 owing to accumulating numbers of immunes in the population). This means, of course, that the cumulative number of cases will continue to rise, in spite of the fact that R has been reduced to 1.

Once R falls to <1, there will still be additional new cases each day for some time. A value of R <1, for example, a value of 0.8, means that 100 cases will transmit to 80 new cases. These 80 new cases will transmit to 64 new cases, and so on. As a result, we would expect the numbers of new cases to dwindle, but it will not be as though a tap has been turned off.

A further complication is that the above comments are made on the assumption of uniform mixing of people who are infected and people who are still susceptible ('homogeneity'). In reality, this is not usually the case. Transmission tends to occur more often within clusters of people than between clusters. This is because there is more contact between individuals within a cluster than there is between people who do not share a cluster. A cluster might be people living together, working together, worshipping together, etc. In other words, there is considerable heterogeneity of the mixing and socialising patterns within society. Therefore, clustered outbreaks (for example, in a community of mineworkers) have the potential to push up the number of new cases in a sporadic way, even if the average, community-wide, value of R is <1.

The estimation of R

The estimation of R is usually performed in a similar way to the estimation of R_0, using an accepted method such as that already described by White and Pagano.[11] However, as the values of β, κ and D are now changing as the epidemic proceeds, the value of R will be expected to change over time. If the level of immunity is rising significantly, or if control measures are succeeding,

then the value of R is expected to decline. On the other hand, if the control measures are not succeeding, then R would not be expected to fall.

There are additional Bayesian methods that are more mathematical, but popular.

The code[12] already referred to that is used for the estimation of R_0 may also be used for the estimation of R, using knowledge of the epidemic curve, the serial interval and its distribution.

In SA, the tracking of R has been carried out by a team of mathematicians in collaboration with infectious disease epidemiologists and other experts. The NICD has made the results available, along with details about the methodology, on their website.[16]

The relationship of R_0 to the herd immunity level

The term 'herd immunity' was first coined by Topley and Wilson[19] in 1923, although not explicitly defined by them. Fine et al.[20] have explained the concept very clearly, and also pointed out that the term has been defined in a number of different ways depending on the context in which the term is used.

For the purposes of this article, we define herd immunity as the threshold level of immunity required in a population that will result in failure of an infectious disease to spread within that population. This immunity may be acquired through vaccination or by natural immunity following experience of the illness in question.

This argument depends on two important assumptions about the population in question:
(i) Immunity is randomly spread throughout the population
(ii) mixing of individuals within the population occurs at random with regard to immune status of individuals in the population.

Under these assumptions, if the immune proportion exceeds:

$$\frac{(R_0 - 1)}{R_0}$$ then transmission will be unlikely to occur if an infectious person enters the population. The above relationship for the herd immunity reduces, algebraically, to: $$1 - \frac{1}{R_0}$$

As an example, if R_0 = 2.5, then the herd immunity threshold level is: 1 − (1/2.5) = 0.6.

This herd immunity level may be expressed as a percentage of the population: 60%.

The implication is that, if 60% of the population is immune (and if the two assumptions are realistic), then, if an infectious individual enters the population, there will be no secondary cases expected due to that person's presence in the population. As pointed out earlier, the value of R_0 may vary from community to community depending on the lifestyle in each community. Hence there is probably no 'national' herd immunity threshold for a country such as SA, where there are widely differing lifestyles and incomes among the population.

Concerning the two assumptions

The two assumptions that underlie the preceding statements about herd immunity are not always met. For example, with

regard to the first assumption, the uniform distribution of immunity throughout the population, one may find that, in general, immunity levels may be very high. Immunity levels may even be above the threshold required for herd immunity against a specific disease. However, there may be pockets of people within the general population where the level of immunity is below this protective threshold.

This may, for example, be due to:
• poor access to vaccination services
• cultural resistance to vaccination
• mistrust of the vaccines on offer.

If an infected, and infectious, person enters one of these communities with low immunity levels, then a local epidemic may arise. An example of such an outbreak (of measles, in this case), in the presence of high general levels of immunisation, has been described for a measles outbreak in a religious community in the USA. The community had only 14% of children vaccinated, despite 88% vaccine coverage in the state overall. This 88% overall coverage would have been sufficient to expect herd immunity if the R_0 value for measles in this state was assumed to be just <10.

Vaccination was not prohibited by the religion practised in the affected community, but their traditional lifestyle had limited their engagement with public health initiatives.[21] In this outbreak, there were 383 cases of measles, 380 of them from the religious community. The other 3 cases in the affected state were from outside the community, but they were epidemiologically linked to the same outbreak strain.

With regard to the assumption of random mixing, again referring to the measles outbreak example, the source case was someone who had visited co-religionists in a foreign country, and then returned to his community. Had he returned to the USA but, instead of going home, mixed 'at random' with his fellow countrymen, it would have been unlikely that he would have come into contact with a susceptible person during his infectious period. Transmission would, therefore, not have occurred.

This outbreak took place because there was a cluster of non-immune people, and because the source case had not mixed 'at random'. In this example, neither of the assumptions was met.

A further observation about herd immunity, and relevant to the current epidemic of COVID-19, is that if an effective vaccine is found against SARS-CoV-2, then as long as 60% of the population is rendered immune through vaccination, the disease COVID-19 will not be expected to spread should it be reintroduced to the population after the end of the current outbreak.

We should note this does not mean that only 60% of people will need to be vaccinated. Sixty percent will have to be immunised. If the vaccine efficacy is, say, only 80%, then the percentage that must be vaccinated, in order to obtain 60% immunity, will be (60/0.8) = 75%. There may, too, be some residual immunity, acquired naturally, from previous exposure to SARS-CoV-2. This may need to be taken into consideration.

As a final comment, the vaccination coverage should be sufficiently spread out to achieve 60% immunity and above in each and every community, particularly at-risk communities, as found in care homes. It will not be good enough to achieve an overall 60% immunity level while there exist pockets of people with below-threshold levels of immunity. This is particularly the case if these community members are at high risk of poor outcomes following COVID-19 disease. Because there is unlikely to be homogeneity of the distribution of the immunes, and random mixing of people, even this vaccine coverage may not be sufficient to prevent outbreaks of COVID-19 subsequently. Furthermore, as the value of R_0 may be higher in high-risk communities, such as hospitals and institutions that care for the elderly and vulnerable, and it might be impossible to estimate, it might be wiser to aim for universal vaccination coverage in such subpopulations, irrespective of the national, or even regional, R_0 value.

Conclusion

We have defined R_0 as the number of people infected, on average, by each infectious person during the infectious period of an illness, given a totally susceptible population with no control programmes implemented. This number is reasonably constant for a given community of people, but may vary between communities with different lifestyles. As a result, R_0 is most meaningful when it has been estimated at the subpopulation level.

R, in contrast, is defined as the number of people, on average, infected by each infectious person in a population that may contain some immune people, and may also have experienced implemented control measures. R, too, is most useful when estimated at a subpopulation level.

The values of R_0, and R, are both affected by the values of β, κ and D. These values of β, κ and D may be impacted by the spread of immunity in the community, as well as the implementation of control measures. As a result, the value of R is expected to change as the epidemic evolves. Tracking this value of R helps authorities to monitor the success or otherwise of their disease control interventions. If this is done at the subpopulation level, then it helps to identify areas where control measures might have been less well implemented, or where the community may need some additional control measures or support in order to bring the epidemic under control locally.

The concept of herd immunity is related to the value of R_0. It is the level of immunity required to prevent further outbreaks, and is calculated from the value of R_0. As the value of R_0 may vary for different communities depending on their social habits or working and living circumstances, relevant herd immunity thresholds may vary from region to region or from community to community.

Herd immunity may be achieved through vaccination, considering the vaccine efficacy. There are two important conditions that must apply for herd immunity to be effective at preventing future outbreaks: uniform distribution of immunity; and random mixing of people. Care needs to be taken, therefore, to ensure that adequate immunity levels are maintained in highly vulnerable groups such as homes for the elderly, or people living with poor access to health services. It is recommended that universal vaccination is aimed for in such smaller groups of people, once a safe and effective vaccine becomes available.

BIOSTATISTICS NOTEBOOK

1. Dietz K. The estimation of the basic reproduction number for infectious diseases. Stat Methods Med Res 1993;2(1):23-41. https://doi.org/10.1177/096228029300200103
2. Lotke AJ. Théorie Analytique des Associations Biologique. Deuxième Partie. Analyse Démographique avec Application Particuliére à l'Espèce Humaine. Paris: Hermann, 1939.
3. Giesecke J. Modern Infectious Disease Epidemiology. 2nd edition. Boca Raton: CRC Press, 2017.
4. Diekmann O, Heesterbeek JAP. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Chichester: John Wiley and Son, 2001.
5. Jones JH. Notes on R0. Stanford University. 2007. https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf (accessed 28 May 2020).
6. Guerra FM, Bolotin S, Lim G, et al. The basic reproduction number (R0) of measles: A systematic review. Lancet Infect Dis 2017;17(12):e420-e428. https://doi.org/10.1016/s1473-3099(17)30307-9
7. Nuño M, Chowell G, Wang X, Castillo Chavez C. On the role of cross-immunity and vaccination in the survival of less-fit flu strains. Theor Pop Biol 2007;71(1):20-29. https://doi.org/10.1016/j.tpb.2006.07.002
8. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Royal Sci B 2006;247(1609):599-604. https://doi.org/10.1098/rspb.2006.3754
9. Wearing HJ, Rohani P, Keeling MJ. Appropriate models for the management of infectious diseases. PLoS Med 2005;2(7):e174. https://doi.org/10.1371/journal.pmed.0020174
10. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Ancel Meyers L. Serial interval of COVID-19 among publicly reported confirmed cases. Emerg Infect Dis 2020;26(6):(epub ahead of print). https://doi.org/10.3201/eid2606.200357
11. White LF, Pagano M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. Stat Med 2008;27(16):2999-3016. https://doi.org/10.1002/sim.3136
12. Obadia T, Haneef R, Boëlle P. The R0 package: A toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med Inform Decis Mak 2012;12:147. https://doi.org/10.1186/1472-6947-12-147
13. Macrotrends. https://www.macrotrends.net/countries/ZAF/south-africa/life-expectancy (accessed 28 May 2020).
14. National Department of Health, South Africa. Update on COVID-19 (28 May 2020). Pretoria: NDoH, 2020. https://sacoronavirus.co.za/2020/05/29/update-on-covid-19-28th-may-2020/ (accessed 28 May 2020).
15. Keeton C. COVID-19 in numbers: SA and the world. TimesLIVE, 23 March 2020. https://www.timeslive.co.za/news/south-africa/2020-03-23-covid-19-in-numbers-sa-and-the-world/ (accessed 28 May 2020).
16. National Institute for Communicable Diseases. The initial and daily COVID-19 effective reproductive number (R) in South Africa. Pretoria: NICD, 2020. https://www.nicd.ac.za/wp-content/uploads/2020/05/The-Initial-and-Daily-COVID-19-Effective-Reproductive-Number-R-in-South-Africa-002.pdf (accessed 1 June 2020).
17. Liu M, Ou J, Zhang L, et al. Protective effect of hand-washing and good hygienic habits against seasonal influenza: A case-control study. Medicine 2016;95(11):e3046. https://doi.org/10.1097/md.0000000000003046
18. McConville C, Boyd P, Major I. Efficacy of tenofovir 1% vaginal gel in reducing the risk of HIV-1 and HSV-2 infection. Clin Med Insights Womens Health 2014;7:1-8. https://doi.org/10.4137%2FCMWH.S10353
19. Topley WWC, Wilson GS. The spread of bacterial infection: The problem of herd immunity. J Hygiene 1923;21(3):243-249. https://doi.org/10.1017%2Fs0022172400031478
20. Fine P, Eames K, Heyman DL. 'Herd immunity': A rough guide. Clin Infect Dis 2011;52(7):911-916. https://doi.org/10.1093/cid/cir007
21. Gastañaduy PA, Budd J, Fisher N, et al. A measles outbreak in an underimmunized Amish community in Ohio. N Engl J Med 2016;375(14):1343-1354. https://doi.org/10.1056/nejmoa1602295