XLIII CILAMCE
21–25, November, 2022
FOZ DO IGUAÇU – PR – BRAZIL

Epidemiologycal SIR model to study 'infodemics´ about child vaccination

J. C. Onofre[1], J. L. Acebal[1]

[1]*Federal Centre for Technological Education of Minas Gerais*
Av. Amazonas, 5253, 30421-169, Belo Horizonte, MG, Brazil
julio.onofre@gmail.com, acebal@cefetmg.br

Abstract.

Technological development has made the internet increasingly accessible in recent decades. However, despite the benefits, overexposure to the dissemination of information has become a social concern. The rapid dissemination of information with fewer criteria plagued by rumours and fake news has compromised its accuracy, clarity and reliability. The process of rapid and massive dissemination of unreliable information has been called an infodemic because it behaves an epidemic. Google Trends is a tool developed to show the relative number of searches for a given term of interest available on the Google platform. In this work, we apply the classic SIR model of mathematical epidemiology to study the case of the time series of the frequency of searches on the controversial term 'childhood vaccination'. The aim of the approach is to assess how accurately the simple form of the epidemiological model can describe the infodemic process. The matter is treated as complex open system and the model parameters like infection and recovery rates from infection are supposed to vary but keeping stable along intervals of engagement and disengagement. We show that the model parameters can provide useful information about the social phenomenon in periods of social media engagement and disengagement in controversial news.

Keywords: SIR Model, infodemics, Google Trends, child vaccination

1 Introduction

Since December 30, 2019, when the Wuhan-China market was closed after an epidemiological alert, the SARS-Cov-2 virus, the etiological agent of Covid-19, has spread rapidly and has gone from isolated cases to a pandemic. On January 24, 2020, the first confirmed case in Europe occurred in France. And, on January 30, 2020, WHO declares pandemic status. On March 20, 2020, the Ministry of Health recognises community transmission in Brazil [1]. The pandemic has caused enormous losses to the world population, both social and economic. Since then, researchers from all over the world have made efforts to mitigate and possibly prevent the spread of the virus. The critical period between the emergence of epidemics, the development of vaccines and the (reasonable) distribution of vaccines among countries was the subject of great engagement on social media.

Amid discussions about possible treatments for Covid-19, the first vaccines emerged in less than a year. Quickly, the effectiveness of treatments and vaccines became the subject of intense debates marked by misinformation and politicisation. In addition to a process of spreading fake news through social networks, news without criteria was given by media vehicles with sensationalism. In this environment, fake news quickly became the topic of the day which confirms observations that fake news spreads significantly faster and more widely than true news in many categories of information [2, 3]. In particular, if this category involves public health, the impact of uncertainties, rumours and fake news on information is a feedback factor in the process of spreading the news and the debate, as scientific sources do not always have the desired reach, are understood or even believed. Engagement and its social effects were enormously intensified when childhood vaccination emerged as a possibility.

According the WHO (World Health Organisation), the social phenomenon characterised by an excess of accurate or inaccurate information which obfuscates access to reliable sources and guidelines is called an infodemic [4]. Infodemia gets its name because of the similarity between the dissemination of information and the spread of a disease in an epidemic [5]. This intense availability of information via news distribution channels and social networks is strongly fuelled by polemics, debates and the legitimate search for reliable sources. Such ebullition results in great social engagement on the topic. As a consequence of this process, a considerable increase in searches for information on the internet is expected. Therefore, the increase in searches for themes through the Google platform can be considered an indicator of the infodemic process associated with these themes. A Recent

works have been studied infodemic processes with use of Sir model [6].

Google Trends is a feature present on the Google services platform for evaluating trends in searches carried out on a topic of interest. It is a new and open-access tool that allows users to interact with research data on the Internet, which can provide deep insights into population behaviour and health-related phenomena [7]. The tool shows the relative frequency over time as a ratio of the maximum frequency reached (fixed at 100 points) in the period of interest for a geographic region that can be selected. Hence, Google Trends serves as a time series recorder of relative search frequencies for a term in this region. Through this time series, it is possible to evidence pulses of engagement and disengagement triggered by social media facts and, following, dumped by information campaigns. Consequently, the tool provides a method for evaluating the infodemic phenomenon.

The first model in mathematical epidemiology of the spread of smallpox was proposed by Daniel Bernoulli in 1760. In 1906, W. H. Hamer suggested that the spread of infection should depend on the number of susceptible and infected individuals. Finally, in 1927, Kermack and McKendrick proposed the SIR model, a model involving differential equations, to describe the spread of infectious diseases among populations of susceptible, infected, and recovered individuals [8, 9]. Since then, the SIR model has been successfully used to predict the effect of epidemics, as well as to provide a tool to study control actions such as vaccination and social distancing.

In the present work, a case study of an infodemic associated with engagement occurred on the term 'child vaccination' [1] is carried out. The time series of searches for the term on Google obtained via Google Trends is partitioned into subintervals associated with pulses or infodemic waves. In addition, each subinterval is also divided into two periods, the period of engagement associated with increasing values of social engagement, followed by the period of disengagement, characterised by a decrease in the social engagement. The rise of engagement is associated with polemics, discussions with the dissemination of non-reliable or non sufficiently verified information. The disengagement period is associated with the progressive action of information media, organisations and health agencies clarifying the matter. The system is considered open since the total number N of individuals can vary, and complex (non autonomous), since the SIR parameters of the *i.e.* model, the contagion and recovery rates can also vary over time. However, they must be constant during periods of engagement or disengagement. Under such assumptions, we eavaluate the description of the system with the SIR model in the study case of the infodemics associated to the term 'childhood vaccination' in Brazil. In section 1, we present basic information about the SIR model. Data and methods are discussed in the Section 3. Results and discussion are provided in Section 4. In Section 5, we draw the conclusions.

2 The SIR model

The SIR model is a mathematical description for the evolution of epidemics from a class called compartmental or population models. The model and its variations have been useful for decades for its simplicity and accuracy in describing the spread of diseases in populations and for assisting in inferences about epidemics and methods of control.

In the SIR model, there are three populations $S(t)$, $I(t)$ and $R(t)$, respectively, the population of susceptible individuals, the population infected by the disease and the population recovered from the disease. When the system is considered closed, the sum of the three populations equals a constant, $S(t) + I(t) + R(t) = N$, the total population number. The evolution of the populations of the SIR-model in time is governed by dynamic rules that can be written in continuous time ($t \in \mathbb{R}$), by using differential equations, or in discrete time, ($t \in \mathbb{Z}$), using difference equations.

Figure 1 shows a discrete-time SIR model. The transitions of individuals between populations in a time interval correspond to fractions of these populations. Thus, at a time t, the nonlinear interaction of infection indicated by the dashed line occurs between individuals from populations $S(t)$ and $I(t)$. The values of these populations make up the rate $\frac{\beta S(t) I(t)}{N}$ of susceptible individuals that become infected in the time interval and change to the infected population. The β parameter is proportional to the rate of social contact times the fraction of contacts that actually produce contagion. Likewise, the population $I(t)$ gives up a fraction of its individuals at a rate $\gamma I(t)$ to the population of recoveries. The γ parameter is the recovery rate, which means it is the inverse of the mean recovery time from the disease.

[1] Search performed for the term in portuguese: 'vacinação infantil'.

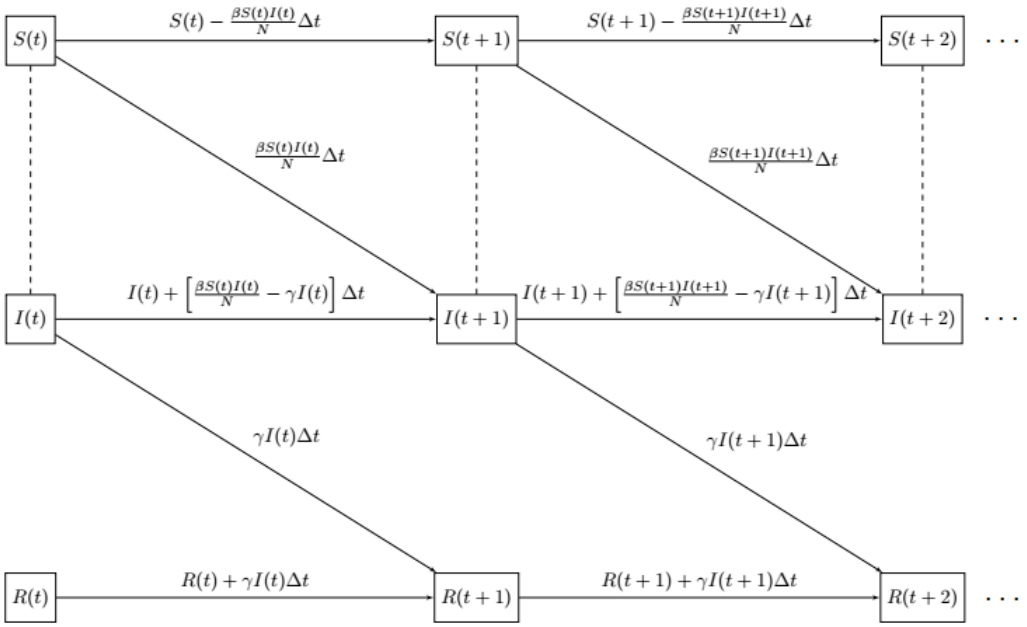

Figure 1. Schematic diagram of the SIR model showing the flow of individuals between populations, according to the parameters of the model and also of their populations.

The dynamics of the discrete-time SIR model can be represented through a system of difference equations:

$$\begin{cases} S(t+1) = S(t) - \left[\dfrac{\beta S(t) I(t)}{N}\right] \Delta t\,, \\ I(t+1) = I(t) + \left[\dfrac{\beta S(t) I(t)}{N} - \gamma I(t)\right] \Delta t\,, \\ R(t+1) = R(t) + [\gamma I(t)]\, \Delta t\,. \end{cases} \tag{1}$$

The time step Δtin the equations is, in principle, arbitrary and can be smaller than the time step of the data to be fitted.

In addition to the model parameters, to determine the intensity of the infectious process it is commonly used the *Basic Reproduction Number*, $R_0 = \frac{I(1)}{I(0)}$, which represents the average number of newly infected in a subsequent period from a single individual infected in an earlier period at the onset of the epidemic process[10]. Thus, the populations at this start are $S(0) \approx N$, $I(0) \approx 1$ and $R(0) = 0$. In this case, it follows from the equation for $I(t)$ in the (1) and the from the definition of R_0 that

$$R_0 = \frac{I(1)}{I(0)} = 1 + \left[\frac{\beta S(0)}{N} - \gamma I(0)\right] \Delta t \approx 1 + \left[\frac{\beta}{\gamma} - 1\right] \gamma \Delta t\,, \tag{2}$$

in which we apply the condition of the beginning of the pandemic. When the time scale is chosen to have $\gamma \Delta t = 1$, we have,

$$R_0 = \frac{\beta}{\gamma}. \tag{3}$$

Hence, this number is expressed in terms of the model parameters. Although there are more general definitions for the reproduction number as $R_t = \frac{I(t+1)}{I(t)}$ which expresses a proportion that varies with the evolution of the process, we will restrict ourselves to R_0, and we will understand that the process that occurs at any time t, this value is compatible with an epidemic that started in the past with R_0 given by the equation (3).

3 Data and Method

To evaluate the infodemic process associated with the theme 'childhood vaccination', a study case was carried out with data collection from the Google Trends service with a geographic filter defined for Brazil, in a period

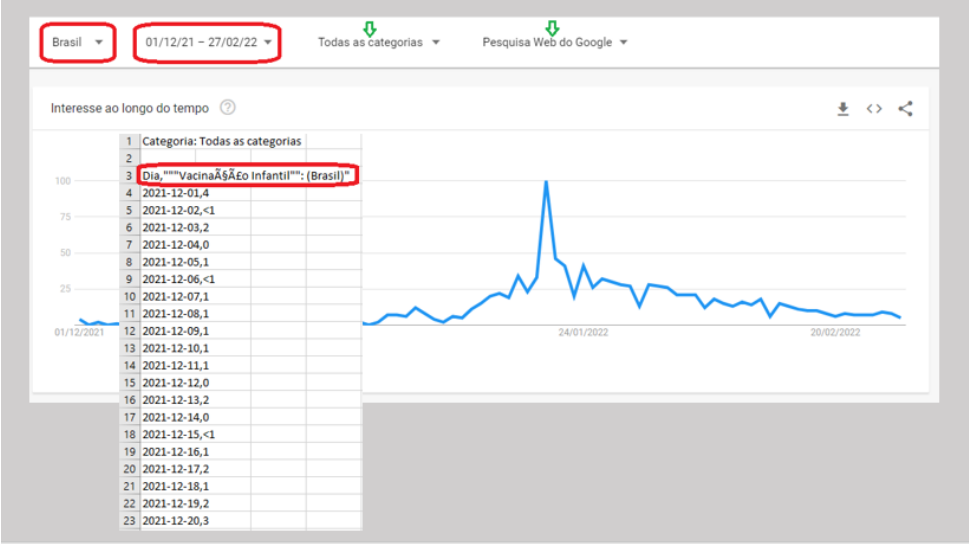

Figure 2. Google Trends Data Sheet and Google Trends Graph - Term "Vacinação Infantil"

between the dates 12/01/2021 to 27/ 02/2022. The time-series chart of the frequency of daily surveys on this subject and the spreadsheet with the corresponding data were obtained (Fig. 2). The initial analysis of the data revealed a certain degree of complexity without, at first, an identification of a typical epidemic process, but perhaps the superposition of several wave pulses, composing a series of events (Fig.3).

The relative frequency of searches in Google was associated with the variable number of infected $I(t)$ of an epidemic in the SIR model (Section 2). The observation of the data raised the fundamental hypothesis that the greater or lesser engagement in the process occurred in periods. In this way, the number N of people involved in the process would vary from a period, according to social media events like the strength of engagement, campaigns of information etc. The rates of the SIR model β and γ which are, respectively, the fraction of susceptible to engage in the subject and the rate of disengagement also can vary at each period. This effect of changing parameters over time is expected in real epidemics as control actions such as social distancing, sanitary measures and vaccines are of common use whenever possible. In this scenario, the time series was partitioned into intervals whose logical sequences are related to processes of engagement (increase) and disengagement (decrease) in each period (see Fig.3).

The partition of data took place into 8 subintervals, each of then was divided into two subsequent periods (see Table 1), corresponding to the engagement period, followed by the disengagement period. The engagement period would correspond to the ascending data sequence that is supposedly the result of a media event that drives an engagement process (infection rising). The disengagement period of these sub-intervals would correspond to the disengagement process that would result from actions to clarify public opinion (epidemic control actions).

For each increasing or decreasing period of each subinterval, a curve of the infected population of the SIR model is fitted to the data. Thus, for each case, a set of model parameters N , β, and γ is obtained via optimisation of the SIR model implemented in Python language. The method used to solve the ODE system was through the *odeint* function of the scipy.integrate sub-package and the optimisation method used was the *Trust Region Reflective* minimising the *least-square* objective function. After determining the parameters, the basal reproduction number R_0 of each case was obtained in order to characterise the intensity with which the infodemic process developed in each analysed period.

Table 1. Description of the sub-intervals of the study period partition into engagement and disengagement pulses in terms of days of each sub-interval, the engagement and disengagement stage and their respective peak.

Period	Days	Peak	Period of
p_1	22º ao 24º	12/24/21 (24º day)	Engagement
p_2	24º ao 26º	12/24/21 (24º day)	Disengagement
p_3	26º ao 28º	12/28/21 (28º day)	Engagement
p_4	28º ao 31º	12/28/21 (28º day)	Disengagement
p_5	31º ao 37º	01/06/22 (37º day)	Engagement
p_6	37º ao 40º	01/06/22 (37º day)	Disengagement
p_7	40º ao 51º	01/20/22 (51º day)	Engagement
p_8	51º ao 89º	01/20/22 (51º day)	Disengagement

4 Results

The model parameters were obtained via optimisation in each sub-interval, distinguishing the periods of engagement and disengagement, according to the initial assumption of having periods, respectively, of misinformation, rumours and fake news and another with educational actions of information through the media and agencies. In each of these periods of engagement and disengagement, the parameters of the SIR model were optimised via Trust Region Reflective optimisation method and had values of R_0 and Root Mean Square Error (NRMSE) normalised to mean of data in the period.

The values obtained by optimisation via the Trust Region Reflective method were described in the 2 table. The curves for the infected population fitted the data, at least, with a stronger correlation (Fig. 4). Finally, Table 2 shows the values of R_0 at each stage of engagement and disengagement of each sub-interval. In each period, the curve behaviour is defined as a result of two tendencies of engagement and disengagement expresses mathematically by the values of, respectively, the rate of infection β and recovering from infection γ, as well as the number N of individuals denotes a proportion of the number of individuals that are able to be interested in the idea. This fact explains the high values of N in some periods. Although official actions on media tend to promote reliable information and disengagement, the action of education may sometimes reach a greater number of individuals. In some intervals, the high values of R_0 are explained by the small number of recovery rates, indicating that in those periods, the trend is markedly pointed to rapid engagement, with few or almost non-existent educational actions.

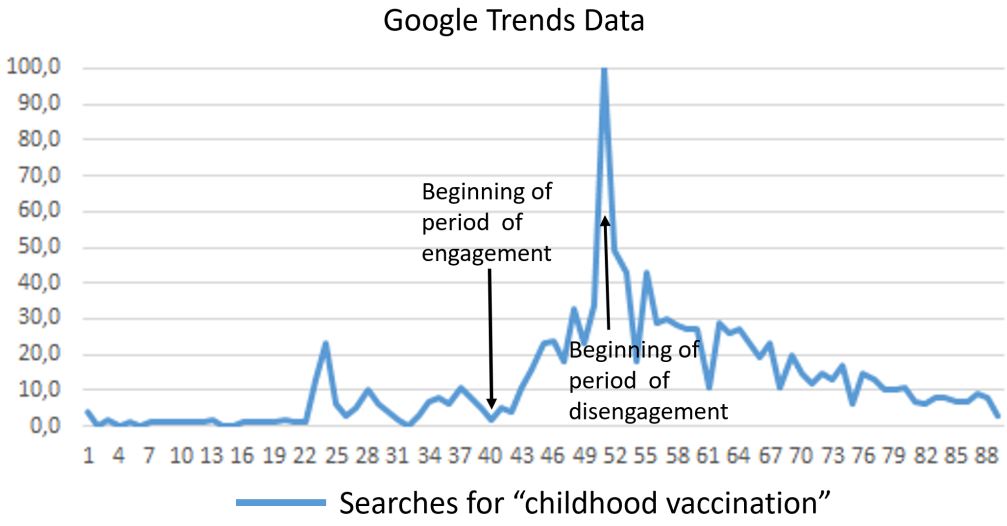

Figure 3. Google Trends search on the term ”Childhood Vaccination”.

Table 2. Parameter values N, β, γ optimised by calculating the approximate value of R_0 in each sub-period and process steps.

	p_1		p_2		p_3		p_4	
N	43	4463	29	19	22	11	202	36705
β	2.311	$9.787 \cdot 10^{-4}$	$7.987 \cdot 10^{-1}$	0.8819	0.6429	0.6537	$3.399 \cdot 10^{-1}$	$1.148 \cdot 10^{-5}$
γ	0.4521	$8.2945 \cdot 10^{-1}$	$3.7566 \cdot 10^{-26}$	0.7359	0.145	0.659	$4.5478 \cdot 10^{-24}$	$7.6987 \cdot 10^{-2}$
R_0	5.113	0.0012	$2.126 \cdot 10^{25}$	1.198	4.435	0.992	$7.474 \cdot 10^{22}$	0.0001
NRMSE	0.4715	0.2314	0.1	0.1599	0.2614	0.2276	0.5702	0.4748
ρ_S	1.0	1.0	1.0	1.0	0.7	1.0	0.9317	0.9028

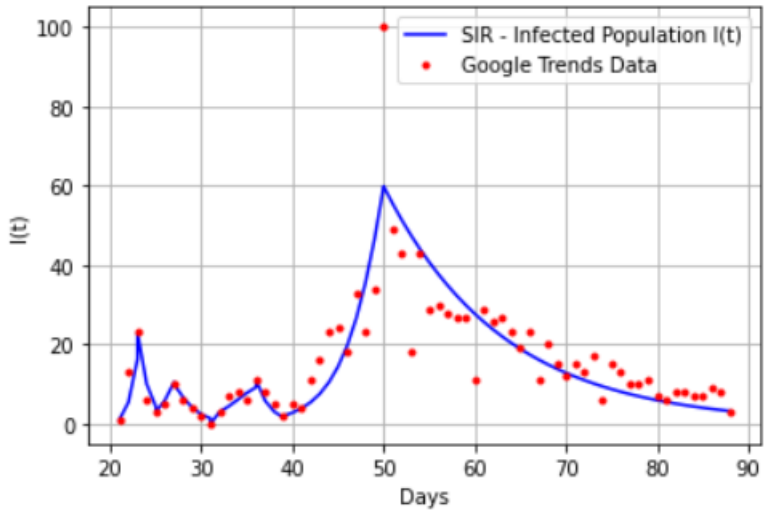

Figure 4. Frequency data of relative searches performed on Google in the analyzed period and curve adjusted by optimizing the parameters of the SIR model.

5 Conclusions

The SIR model, in its simplest form, was able to satisfactorily describe the infodemic process in this case study. This fact can serve to corroborate the epidemic character of the dissemination of information.

The adopted assumptions with the partition of the complex process observation time into sub-intervals, and into periods of engagement and disengagement proved to be effective. On the one hand, it was based on the realistic varying characteristics of the epidemics process over time due to vaccination, control social distancing. On the other hand, fitting the model to the data revealed that the hypothesis was adequate.

Other refinements can be made in the methodology of this work to make the method an instrument of social analysis. One possible refinement is the automatic identification of sub-intervals with computational intelligence. Also, other optimisation methods can be used with more complex data, such as Genetic Algorithms.

Authorship statement. This section is mandatory and should be positioned immediately before the References section. The text should be exactly as follows: The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] N. H. THEY. Uma breve linha do tempo. *UFRGS Litoral*, 2020.

[2] Z. Zhao, J. Zhao, Y. Sano, O. Levy, H. Takayasu, M. Takayasu, D. Li, J. Wu, and S. Havlin. Fake news propagates differently from real news even at early stages of spreading. *EPJ data science*, vol. 9, n. 1, pp. 7, 2020.
[3] S. Vosoughi, D. Roy, and S. Aral. The spread of true and false news online. *science*, vol. 359, n. 6380, pp. 1146–1151, 2018.
[4] W. H. O. (WHO) and others. Understanding the infodemic and misinformation in the fight against covid-19. *Retrieved March*, vol. 10, pp. 2021, 2020.
[5] J. Zarocostas. How to fight an infodemic. *The lancet*, vol. 395, n. 10225, pp. 676, 2020.
[6] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala. The covid-19 social media infodemic. *Scientific reports*, vol. 10, n. 1, pp. 1–10, 2020.
[7] S. V. Nuti, B. Wayda, I. Ranasinghe, S. Wang, R. P. Dreyer, S. I. Chen, and K. Murugiah. The use of google trends in health care research: a systematic review. *PloS one*, vol. 9, n. 10, pp. e109583, 2014.
[8] F. Brauer. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling*, vol. 2, n. 2, pp. 113–127, 2017.
[9] C. Colombo and M. Diamanti. The smallpox vaccine: The dispute between bernoulli and d'alembert and the calculus of probabilities. *Lettera Matematica*, vol. 2, n. 4, pp. 185–192, 2015.
[10] S. Contreras, H. A. Villavicencio, D. Medina-Ortiz, C. P. Saavedra, and Á. Olivera-Nappa. Real-time estimation of r t for supporting public-health policies against covid-19. *Frontiers in public health*, vol. 8, pp. 556689, 2020.