COMPUTER MODEL FOR PREDICTING EFFECTS OF STRAIN-SPECIFIC IMMUNITY

Manuel Castañón–Puga[1], Camilo Khatchikian[2], E. Dante Suarez[3], Dustin Brisson[4] and Alfredo Tirado–Ramos[5]

[1]*Chemistry and Engineering School, Universidad Autónoma de Baja California*
Tijuana, BC 22424, Mexico
[2]*Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth*
Lebanon, NH 03756, USA
[3]*Trinity University*
San Antonio, TX 78212, USA
[4]*Department of Biology, University of Pennsylvania*
Philadelphia, PA 19104
[5]*Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth*
Lebanon, NH 03756, USA

ABSTRACT

Different outcomes may be expected when a particular host is exposed to an infectious disease, depending on its immunity status. For example, because of a previous encounter with the disease, the host may not become infected as consequence of the presence of antibodies. While sometimes it is difficult to establish when this disease avoidance occurred, the recognition of the existence and duration of such immunity is critical for the development of vaccines and accurate epidemiological models. In this paper, we propose a generalized computer model that can examine a wide range of epidemiological data and generate expected patterns of strain-specific infections and reinfection. These expected patterns can be compared and tested with empirical patterns existing in available databases.

KEYWORDS

Immunity, Reinfection, Model, Mining, eHealth

1. INTRODUCTION

Different outcomes can be expected when a particular host is exposed to a compatible infectious disease: a) the host become infected; b) the host does not become infected as, it is naturally immune to the disease; or c) the host does not become infected because it has developed resistance from a previous encounter with the disease. In the latter case—when immunity is acquired after a previous infection—the immunity could last for a finite timeframe or be permanent. The existence and duration of such immunity is a critical feature for the development of vaccines (for example, see Wormser 2021), as well as for the modeling of disease transmission, and the ultimate effects on entire populations (*i.e.*, Lee *et al.* 2021).

Identifying (and quantifying) the existence of acquired immunity from direct data is challenging, as the process involves having certainty of the existence of an "infectious event" (i.e., an exposure to the infectious agent) that did not result in an infection, where failure to produce a disease in the host could be due to innate or acquired immunity. It is feasible to distinguish between these two scenarios if there is a previous record of a successful infection, as such record would invalidate the innate immunity case.

The existence of records that confirm previous infections and exposure to an "infectious event" are difficult to find. Such difficulty is centered in the lack of evidence that a particular individual was exposed to an infectious agent, and that the reason of avoidance of the disease was the existence of acquired immunity. Previously, Khatchikian *et al.*, 2014, proposed that by examining and modeling the reinfection patterns of different Lyme disease strains (etiological agent *Borrelia burgdorferi*), it is possible to infer the presence and duration of strain-specific immunity. The model developed by Khatchikian and collaborators was limited to the outcomes of 17 clinical patients in the observed data set, which contained documented Lyme disease

reinfections where the specific strain was identified, and correspondingly characterized according to a surface protein.

Current developments in electronic health records allow for the aggregation of systems that are publicly available, leveraging accessible research data. Furthermore, there has been recent and rapid growth of molecular sequences deposited in public repositories (such as Genbank, Benson *et al.* 2013), which allow the identification of the agents responsible for infections, and thus facilitate the accessibility and availability to such records (*i.e.*, Alex *et al.* 2017). Access to large amounts of epidemiologically relevant data highlights the potential utility of models that can take advantage of such data and produce pertinent inferences.

We propose a generalized computer model that can be used to examine a wide range of existing data. Such information may currently exist in clinical records, offering flexibility regarding the number of patients considered, infectious opportunities, number of strains involved, number of infections, and prevented infections. To this end, we provide examples and discuss the potential advantages and limitations of the proposed methodology.

2. METHODS

2.1 Model

The model allows the simulation of strain-specific immunity by generating expected patterns of infections and reinfections of different infectious agents' strains (*i.e.*, immunity-recognizable types), based on set model parameters. The model is intended to explore the parameter space, allowing the researcher to generate patterns that match patient-observed datasets. The model proposed here is the first attempt to expand and generalize the initial, infectious agent-specific (Lyme disease) model previously developed in Khatchikian *et al.*, 2014. The present model has greater flexibility and is intended to use data from multiple sources, including clinical records, online repositories, case reports, and simulated datasets.

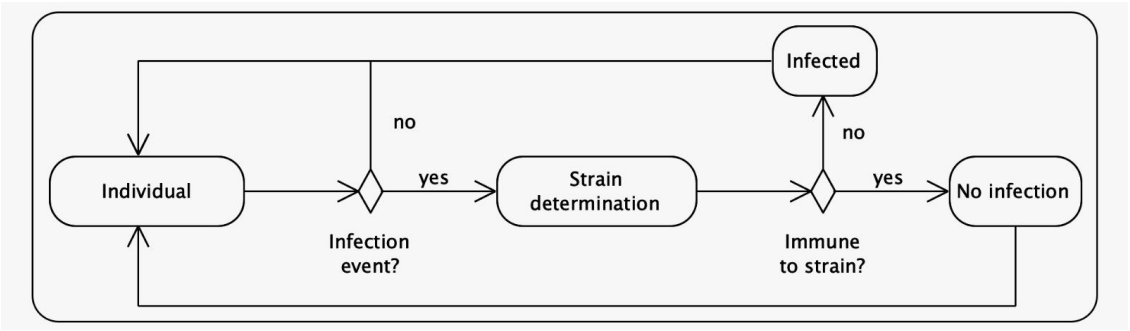

Figure 1. Schematic representation of the steps in the simulation model. Each individual's probability of exposure to an infection event are determined by the *probability of infection*, if it is exposed the model determines to which strain it is exposed according to the *strain distribution*. If the individual has been exposed to that same strain, the model considered that it is immune, and no new infection is recorded; otherwise infection occurs. Each individual goes thought this cycle n times according to the *time in simulation*. The total number of individuals is determined by the *number of individuals*

2.2 Script Coding

The model is coded in R (R Core Team 2017). We use the R implementation in RStudio (Rstudio Team 2015). The overall logic and data flow of the model is presented in Figure 1. To illustrate the model, we run it with simple settings (Table 1).

Table 1. General parameters of the model and specific values used in test simulation (right column)

Model parameter	Notes	Value in test run
Number of individuals	Number of potentially susceptible hosts	2,000
Time in simulation	Number of cycles in the simulation (*i.e.*, number of times that each host could be exposed to an infection)	20
Probability of infection	Chance of each host to be challenged by an infectious event	0.1
Strain distribution	Probability of infection with each specific strain	A=0.5 B=0.25 C=0.125 D=0.125

3. RESULTS

The model presented here allows to generate expected patterns of strain-specific infections and reinfection, which can in turn be compared and tested with observed or predicted patterns existing in databases.

Our test run of the script (see Table 1) allows for the comparison of the pattern (*i.e.*, distribution) of first strains vs. overall strain infections. It is therefore possible to infer the existence of strain-specific immunity by observing the changes in the strain prevalence (Figure 2).

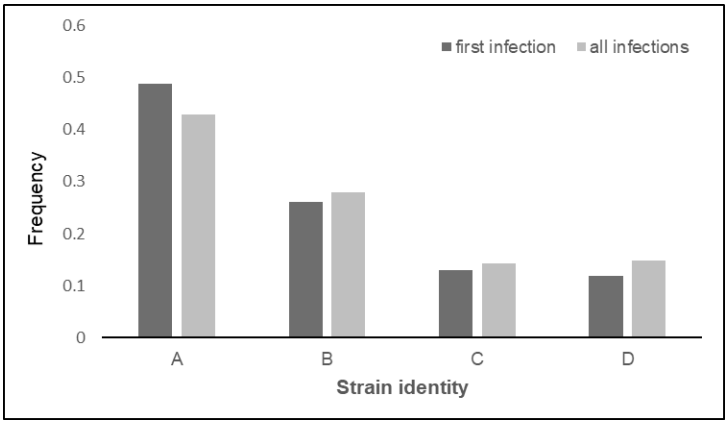

Figure 2. Frequency of strains (A, B, C, or D) in first strains vs. all infections. The strain present with high frequency (strain A) is less present in all infections, while less common strains are more frequent

4. CONCLUSIONS

Our proposed model opens the door to the exploration and simulation of different infections and reinfection patterns for diseases that have differentiable immune strains. The proposed methodology produces numerical expectations that can be statistically compared with observed datasets, generates inferences of an immune-specific event (as previously demonstrated by Khatchikian et al., 2014.), and thus allows the researcher to make inferences and predictions of the real epidemiological effect. The approach proposed here is limited to the existence of empirical records that can be compared with simulated values. Finally, this model has potential teaching applications, as it allows the easy manipulation of its parameters to generate a wide range of outcomes that demonstrate basic epidemiological concepts.

REFERENCES

Alex FB *et al.*, 2017. Exhaustively Characterizing a Patient Cohort by Prevalence of EMR Facts: a Generalized, Vendor-Agnostic Method for Quality Control and Research. *AMIA Annual Symposium Proceeding*, 458-464.

Benson DA *et al.*, 2013. GenBank. *Nucleic Acids Research*, 41, D36-D42.

Khatchikian CE *et al.*, 2014. Evidence for strain-specific immunity in patients treated for early Lyme disease. *Infection and immunity*, 82(4), 1408–1413. https://doi.org/10.1128/IAI.01451-13.

Lee TJ *et al.*, 2021. Network models in epidemiology. In: *Handbook of Statistics* (Eds. Srinivasa Rao ARS, Rao CR). Elsevier Vol 44. https://doi.org/10.1016/bs.host.2020.09.002.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

RStudio Team, 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com/.

Wormser GP, 2021. A brief history of OspA vaccines including their impact on diagnostic testing for Lyme disease. *Diagnostic Microbiology and Infectious Disease*, 102(1), 115572. doi: 10.1016/j.diagmicrobio.2021.115572.