Humans vs. Zombies: Data-driven Modeling of Disease Spread

Ognyan Simeonov  
*Bates College*, osimeono@bates.edu

Kari Lemelin Fliss  
*Unity College*, kariafliss@gmail.com

Jennifer Driscoll  
*Unity College*, jmdriscoll0725@gmail.com

Carrie Diaz Eaton  
*Bates College*, cdeaton@bates.edu

Follow this and additional works at: [https://ir.library.illinoisstate.edu/spora](https://ir.library.illinoisstate.edu/spora)

**Recommended Citation**

DOI: [https://doi.org/10.30707/SPORA8.1.1664052111.377374](https://doi.org/10.30707/SPORA8.1.1664052111.377374)  
Available at: [https://ir.library.illinoisstate.edu/spora/vol8/iss1/9](https://ir.library.illinoisstate.edu/spora/vol8/iss1/9)

This Mathematics Research is brought to you for free and open access by ISU ReD: Research and eData. It has been accepted for inclusion in Spora: A Journal of Biomathematics by an authorized editor of ISU ReD: Research and eData. For more information, please contact ISUReD@ilstu.edu.
Humans vs. Zombies: Data-driven Modeling of Disease Spread

Cover Page Footnote
This work would not have been possible without the help of the student moderators who worked hard on organizing and planning each game. We also thank undergraduate students Timothy Beaulieu and Michelle Lewis, who authored the QUBES HvZ site. The authors would also like to thank Dr. Erin Bodine, Dr. Alex Capaldi, Dr. Jean Marie Linhart, Dr. Robert Smith?, and Dr. Meredith Greer, for sharing their modeling work with HvZ on their campuses. We would also like to thank Dr. Meredith Greer for providing suggestions that helped improve and clarify this paper. Lastly, the authors are thankful to the group who created the game and manage the website www.hvzsourse.com, without whom HvZ wouldn't exist.
Humans vs. Zombies: Data-driven Modeling of Disease Spread

Ognyan Simeonov¹⁺, Kari Lemelin Fliss², Jennifer Driscoll², Carrie Diaz Eaton¹

Abstract

Disease spread in close-knit communities depends heavily on the natural immunity of the individuals in the community as well as on the individuals’ interactions within the environment. This study uses data from a game of tag called Humans vs. Zombies, played on a small liberal arts campus, to examine how this “human element” can affect the spread of diseases in such communities. We fit five disease models to our data and find the best-fit parameters for each model. We conclude that an SIR model with multiple susceptibility classes and sleep cycles modifications provides the best fit, showing that human behavior is an essential element in disease modeling and that the diurnal routine of the individuals plays an important role in disease spread. Other college campuses could model similar games to observe how different cultures of people affect the results, which also doubles as a fun student engagement opportunity.

Keywords: disease modeling, human behavior, gamification, zombie, optimization

1 Introduction

Disease modeling is a well-developed area of mathematical biology [8, 9, 19]. However, validation of disease models in human populations is difficult. It is neither moral nor legal to release a disease into a human population to see how it spreads and compare that to the model being tested. Therefore, certain variables that can play an essential role in the disease spread, such as human behavior during an outbreak, need to be estimated.

One solution to the problem of validating a disease model is the use of games with human players to test model parameters and predictions. One example of this is the 2005 outbreak in the game World of Warcraft (WOW). Caused by a programming oversight, a virtual virus started spreading in the game, infecting thousands of characters. Players tried preventing their characters from getting infected because there was a penalty for dying, but lower-level characters who were more susceptible kept reentering the population and becoming infected by higher-level characters who were less susceptible and acted as carriers. Even though the WOW pandemic was not favorable for the players, theudiologists could use the data from the game to draw insight about the dissemination of infectious real-life diseases [2].

This example shows that data from games can be a powerful tool for disease modeling. Game simulations can also be a powerful tool for engaging students in mathematical modeling [6]. In this study, we examine how a game of tag played on college campuses called Humans vs. Zombies (HvZ) can be used for modeling disease spread as part of a modeling class project.

HvZ started at Goucher College in 2005 and has grown to more than 1,200 games played each year [22]. This is a role-playing game of tag that can last anywhere from a few days to several weeks. Players begin as humans and one is randomly chosen to be the original zombie. Humans try to avoid becoming infected as they get around campus, while the zombies’ only goal is to infect as many humans as possible. The full rules are available at humansvszombies.org [22], but these guidelines can be modified by the moderator(s) who plan the game at each location. As the week progresses, susceptible individuals become zombies, and zombies become corpses. This game presents unique opportunities for disease model testing as well as an ability to engage students in mathematics.

Current literature in zombie modeling reveals the following conclusions. A well known article by [14] started a wave of popular science discussion, revealing to the public the uses of mathematical modeling to survive a zombie attack (e.g., Zombies and Calculus [11], and Braaaiiinnnsss [20]). SIR models are popular for studying the zombie virus because their simple construction makes results and dynamics easy to interpret, and they can also be manipulated to accommodate a variety of parameters [5, 7, 19].

Munz et al. [14] theorize that quarantine would not make any difference during a zombie outbreak because any amount of zombies that do not make it into a quarantine zone would restart the epidemic. They also develop a model showing a cure would only work if it is 100% ef-
fective; otherwise, a very small human population would coexist with the much larger zombie population, and thus, impulsive eradication would be the best method of eliminating the zombie disease. McGahan et al. [11] use data from HvZ to compare models of the disease spread in the game. They develop 28 models and conclude that zombies hunt in groups and some zombies are better hunters than others. Based on analysis of our games and student gameplay experience, we conjectured that differential human survival skills, as well as players’ sleep cycles, play an important role in disease dynamics. In this article, we connect zombie disease modeling techniques with data from field collection to compare a series of models including SIR and predator-prey, while accounting for multiple susceptibility classes and the sleep cycles of the players.

Lastly, from a student engagement perspective, the game of Humans vs. Zombies allows students to actively participate in planning as a moderator or in the game execution [10]. Students who might not be interested in mathematical research, now find themselves actively engaged. Students who have the tools to participate in the modeling can also show leadership as a moderator and improve their understanding of how the models connect to the game. Thus, in the last section, we include an educational component of how the game can be used to teach undergraduate students simple but powerful modeling techniques.

2 Data & Methods

Ordinary differential equation (ODE) modeling, such as that in SIR and predator-prey models, is an oversimplification of the ecological dynamics. SIR models assume well-mixed populations with no individual variation, and thus, they are unable to incorporate spatial information or effects. In addition, although well-suited to assess the average or long-term effects of disease outbreaks, SIR models are less suited to model accurately the trajectory before a disease reaches equilibrium in a single, finite population. However, we can introduce modifications in SIR models that account for populations with a variety of characteristics and ones that consists of different susceptibility classes. These models can be implemented using ODE solver and optimization software packages, and the results can be easily interpreted. Thus, from the point of view of engaging students on the undergraduate level, we are choosing a powerful modeling technique that builds upon the SIR model and requires only a year of calculus and differential equations.

We explore data on the transmission of the zombie disease by modeling it under several scenarios: the SIR model, predator-prey with functional response II & III, SIR model with multiple susceptibility classes, and SIR model with incorporated sinusoidal sleep cycles of the players. We then compare these models to see what is the best descriptor of the zombie outbreak. We also develop a protocol to keep data collection comparable between years that can be found in the Appendix.

2.1 Gameplay and data collection

Humans vs. Zombies (HvZ) is a week-long game of tag played on college campuses around the world. When the game is initiated, one to five players are chosen as original zombies and everyone else is a human. Zombies must tag a human once every 48 hours or they starve out of the game and become corpses. Humans can protect themselves by stunning zombies for a period of 10 minutes using Nerf guns or socks. Once a day, a mission occurs and all players are required to attend. The purpose of missions for disease modeling is to allow for even population mixing among humans and zombies. Even population mixing is achieved because the humans and the zombies have to be in the same location during the missions. Therefore, the humans are prompted to come in close proximity with the zombies and interact both with other humans and with the zombies. Rewards and penalties are used as an incentive to participate in the missions. The game also accommodates the use of control strategies to modify the outcomes. This could include quarantine of infected individuals, temporary vaccines, or a limited number of cures. A detailed description of the gameplay, missions, and organization of the game can be found in the Appendix.

In order to compare games across campuses, it is important to note the different cultures of students and how they may affect the game outcome. Data collection for the purpose of this study was completed at Unity College in Maine. The Unity College campus is small; with just over 500 students attending the school, at the time of data collection it is a very close-knit community. The college also owns a sizable woodlot that runs adjacent to campus buildings. The typical student at Unity College feels comfortable going into the woods, which opens up opportunities for humans to hide from zombies or zombies to ambush humans.

The first game used for this study occurred in the fall of 2012. The game lasted for 164 hours (6 days and 20 hours) with a total of 72 players (14% of the student body). This project involved the use of a control strategy 72 hours into the game. The control strategy entailed 12 cures given to zombies, moving these individuals back into the susceptible class. The second data collection period occurred in the fall of 2013 and was used to validate the model selection results from the previous year. This game lasted 114 hours (4 days and 18 hours) and no control strategies were used in the second game period. Our
Figure 1: HZC compartmental model, where Humans are H, Zombies are Z, and Corpses are C. Two rate constants, \(k\) and \(b\) are used to determine the rate of movement among classes.

Data were recorded every hour only, and given that the timelines of the games were only a few days, modeling to the hour was more suitable than modeling daily rates.

### 2.2 Models

We establish five ODE models to show the movement of individuals from one state to another. First, we compare the SIR model with two variations of the predator-prey model. Then, we build upon the SIR model, adding multiple susceptibility classes and sleep cycles variations.

#### 2.2.1 HZC model

The classic SIR disease model was used as a basis for all of the equations used. This model is considered as one of the simplest ways to model disease spread and has been used to model zombie outbreaks previously. We altered the SIR model, changing the susceptible class to humans (H), the infected class to zombies (Z), and the removed class to corpses (C), creating the HZC model in Figure 1. The individuals move from the human to the zombie class at a rate \(k\) and from the zombie to the corpse class at a rate \(b\). In this model the \(k\) parameter represents the number of contacts a susceptible individual has in a time period that would cause a spread of the disease, and the parameter \(b\) represents one over the starvation time.

\[
\frac{dH}{dt} = -kHZ, \quad \frac{dZ}{dt} = kHZ - bZ, \quad \frac{dC}{dt} = bZ \tag{1}
\]

#### 2.2.2 Predator-Prey II & III

We also tested a predator-prey modification in an effort to account for the observation that the spread of the disease is intentional via zombies instead of random as is the case in the SIR model. In this scenario, zombies would be the predators and humans their prey. Type II and type III modifications are both used because they each demonstrate different qualities that could be present in zombie-human interactions. The type II functional response shown in (2) consists of the following characteristics: search rate is constant; predation rate saturates at high prey density; prey mortality declines with decreasing prey density.

\[
\frac{dH}{dt} = -\frac{kHZ}{a + H} \tag{2}
\]

\[
\frac{dZ}{dt} = \frac{kHZ}{a + H} - bZ, \quad \frac{dC}{dt} = bZ
\]

In the type III modification displayed in (3) predators increase search activity with increasing prey density, prey mortality can increase with increasing prey density, and the consumption rate of zombies becomes lower as time progresses. In both the type II and III modifications a new parameter \((a)\) was introduced, accounting for search time, which is an important piece of a predator-prey relationship. This accounts for the time that zombies would spend looking for humans in the game.

\[
\frac{dH}{dt} = -\frac{kH^2Z}{a^2 + H^2}, \quad \frac{dZ}{dt} = \frac{kH^2Z}{a^2 + H^2} - bZ, \quad \frac{dC}{dt} = bZ \tag{3}
\]

The key difference between the type II and type III modifications is that while a type II curve represents predator populations that need a certain amount of time to capture and consume their prey, the type III curve is more indicative of predators that will not capture and consume the prey at a great intensity below a certain threshold prey density level, but above this level, they would consume prey at larger amounts. For more explanation on these models refer to [17], who provides great visualizations and comparisons between the three predator-prey functional responses.

#### 2.2.3 Multiple susceptibility classes (HWZC model)

Multiple susceptibility classes can be used to modify the model assumption that all susceptible individuals are equivalent. Thus, we decided to incorporate a warrior class in our model. This HZC model with an added warrior class (HWZC) was meant to account for humans (warriors) who are particularly hard to infect (Figure 2). The warrior class is similar to a class of people with better immune systems or natural resistance to a disease. The rate constant used for the warrior class is \(h\). The additional class requires an additional differential equation shown in (4).

\[
\frac{dH}{dt} = -hHZ, \quad \frac{dZ}{dt} = hHZ - bZ, \quad \frac{dW}{dt} = hW + \frac{k}{h}HZ - bW, \quad \frac{dC}{dt} = bZ
\]

In the 2012 game, once the control strategy was performed and the cures were given, all but two people given the cure were turned back into a zombie within 18 hours. The thought was that each of these people belonged in the human class to begin with and only warriors were likely to survive the whole week. The percentage of players who survived until the last mission was 19% of the total population. For the second game, we kept track of the players who were wholly committed to the game and went above and beyond what was expected. We estimate
that 22 people met this requirement for the 2013 game which is also 19% of the total population. Therefore, we use 19% of the population as the approximate number of warriors for the model.

\[
\begin{align*}
\frac{dH}{dt} &= -kHZ, \\
\frac{dW}{dt} &= -hWZ, \\
\frac{dC}{dt} &= bZ
\end{align*}
\]

(4)

2.2.4 HWZC + sleep cycles model

In order to keep the models as accurate with the data as possible, we include an adaptation to accommodate for the players’ sleep cycles. By looking at the data, it appears that movement from the human class to the zombie class is most frequent around noon every day and least frequent around midnight each night. For example, there is not a single night in either game when humans were tagged between midnight and 5 A.M. Alternatively, there is always a spike in tags between noon and 5 P.M. Thus, we add a sinusoidal sleep cycle modification to the multiple susceptibility model to determine how the activity level of humans throughout the course of the day changes the rates of infection as shown in (5).

\[
\begin{align*}
\frac{dH}{dt} &= -k\left(0.5\left(\sin\left(\frac{t\pi}{12}\right) + p\right) + 1\right)HZ, \\
\frac{dW}{dt} &= -h\left(0.5\left(\sin\left(\frac{t\pi}{12}\right) + p\right) + 1\right)WZ, \\
\frac{dZ}{dt} &= k\left(0.5\left(\sin\left(\frac{t\pi}{12}\right) + p\right) + 1\right)HZ + h\left(0.5\left(\sin\left(\frac{t\pi}{12}\right) + p\right) + 1\right)WZ - bZ, \\
\frac{dC}{dt} &= bZ
\end{align*}
\]

(5)

2.2.5 Akaike Information Criterion & parameter fitting

We calculate the Akaike Information Criterion (AIC), which allows us to compare the models and identify which one fits the data the best with the least number of parameters used [12]. The following calculation may be used:

\[
AIC = N\ln\left(\frac{\chi^2}{N}\right) + 2K,
\]

(6)

where \(N\) is the sample size, \(\chi^2\) is the sum of squared residuals (SSR), and \(K\) is the number of unknown parameters in the model. By the way we construct the AIC, we get penalized for an increase in the sum of squared residuals and the number of parameters we have in the fitting. This helps us select a best-fit model that is not only as close as possible to the observed data (low SSR) but also uses a minimal number of parameters (prevents overfitting).

We use three parameter optimization methods to find the best-fit parameters for each model: non-linear least squares, differential evolution, and dual annealing. These methods are easy to implement using the scipy.optimize package in Python, and each method uses different techniques to estimate the best-fit parameters [4, 13, 16], so we obtain three independent parameter sets. Parameter fitting was performed with Python and R Studio, and the code is available [18]. All three techniques yield the same parameter values for all time periods. However, the non-linear least squares technique requires approximately 292 function evaluations to find the best-fit parameters for the HWZC + sleep cycles model, while differential evolution and dual annealing require approximately 3,959 and 16,418 function evaluations respectively, so non-linear least squares proved to be the most time-efficient out of the three methods.

3 Results

To model this game, we need to establish values for all the parameters in the models: rate of infection (k) and starvation rate (b) for the HZC model, an additional search time parameter (a) for the predator-prey models, and a warrior rate of infection (h) for the multiple susceptibility model.

3.1 Modeling 2012 data

It is not possible for a zombie to starve during the first 48 hours of the game because zombies starve 48 hours after their last tag. In addition to this, when the game begins, the original zombies are hidden until the first 5 tags. It is not until about 13 hours into the game that humans are being tagged in large enough numbers that
Humans vs. Zombies: Data-driven Modeling of Disease Spread

Simeonov, Lemelin Fliss, Driscoll, Diaz Eaton

Food (human tags) becomes difficult to obtain. Therefore, the zombie that starves first will likely get its last tag 14 hours into the game; this causes the first corpse to appear 62 hours in. This is concurrent with observed data during the 2013 game as well. To accommodate this, \( b_0 \) is used as a constant equal to 0.0 for the first 62 hours of each game and in later phases the result of the least squares optimization helps determine the value of \( b \).

At 75 hours into the game, a control strategy brings 12 zombie players back into the susceptible class via a cure with no immunity to the disease. The purpose of this control strategy is not to rid the game of the zombie disease, but instead to keep the humans around a little longer. Once cures are administered there is a period of 12 hours (9 P.M. to 9 A.M.) with no movement among classes. During the following six hours (9 A.M. to 3 P.M.), 10 humans are tagged and the susceptible population reverts back to approximately the same number of individuals as before. The 10 people who are tagged during this period are all individuals who were cured the night before. Additionally, in 2012 we start with 2 corpses at the beginning of the game that are later moved to the human class because the moderators allowed 2 participants to join the game 21 hours after the game had started. By including the participants that joined late as corpses at the beginning, we preserve the total population of the system constant throughout the whole game.

To accommodate for the changing starvation rate and movement between classes in our models, we split the time span in three periods: from \( t = 0 \) h to \( t = 62 \) h, when the starvation rate is \( b_0 = 0 \); from \( t = 62 \) h to \( t = 75 \) h, when the starvation rate changes; and from \( t = 75 \) h to \( t = 165 \) h, after we have applied the control strategy.

The best fit model reflects a parameter set that yields a minimal sum of squares of residuals (SSR) and has the lowest AIC value. Using the AIC calculation described previously, we determine the best fit model to be the HWZC + sleep cycles. It has the lowest AIC value of 1212.21 compared to the second-best-fit model (type III predator prey) with AIC of 1289.97. The SSR and AIC values for all five models are summarized in Table 1.

Each of the three SIR-based models (HZC, HWZC, HWZC + sleep cycles) show similar best fit values for \( k \) in the first time period. Similarly, the type II and III predator-prey models seem to have similar rate of infection \( h \). Another interesting fact we notice is that both predator-prey models have very high search time parameters (\( a \)) in the first time period, implying that the zombies infect more humans at the beginning.

In the second time period, which continued for about 10 hours, the three SIR-based models have statistically significant results for the starvation rate with the best-fit model’s starvation rate at 0.0054. The results for the third time period are similar—the starvation rates are significant for the three SIR-based models. However, an interesting observation here is that the warrior rate of infection \( h \) equals the human rate of infection \( k \) for the HWZC and the HWZC + sleep cycles models. This suggests that at the end of the gameplay the warriors and the humans work together to fight off the zombies. All parameter values are available in Table 2. Given these parameters, we plot the best-fit model along with the data (Figure 3) to visualize the spread of the disease.

3.2 Modeling 2013 data

In 2013, we didn’t implement any control strategies, and thus, we split the model in just two parts. From the beginning of the game until \( t = 62 \) h when the starvation rate is \( b_0 = 0 \) and from \( t = 62 \) h until the end of the game. We use the Akaike Information Criterion to compare the models (Table 3).

We find that the HWZC + sleep cycles model provides the best fit for the 2013 data as well. However, the AIC values for the best-fit model are only slightly smaller at 1187.05 and 1206.23 respectively compared to the second best fit model (HWZC). We also notice that while the type III predator-prey model provides the second best fit for the 2012 data, it is replaced by the HWZC model in 2013.

After performing the three optimization techniques, we conclude that 14 out of 23 parameters we obtain are statistically significant. The parameters obtained by the
Table 1: The sum of squared residuals, sample size, number of parameters, and AIC values for each 2012 model. Data was collected hourly from hour zero until gameplay concluded at 164 hours, for 492 data points.

<table>
<thead>
<tr>
<th>Model Name</th>
<th>SSR ($\chi^2$)</th>
<th>Sample Size (N)</th>
<th>Number of Parameters (K)</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWZC + sleep cycles</td>
<td>5595.91</td>
<td>492</td>
<td>8</td>
<td>1212.21</td>
</tr>
<tr>
<td>Type III predator-prey</td>
<td>6554.11</td>
<td>492</td>
<td>8</td>
<td>1289.97</td>
</tr>
<tr>
<td>HWZC</td>
<td>6723.68</td>
<td>492</td>
<td>8</td>
<td>1302.54</td>
</tr>
<tr>
<td>HZC</td>
<td>10755.02</td>
<td>492</td>
<td>5</td>
<td>1527.65</td>
</tr>
<tr>
<td>Type II predator-prey</td>
<td>14163.37</td>
<td>492</td>
<td>8</td>
<td>1669.09</td>
</tr>
</tbody>
</table>

Table 2: Parameter values for the five models for all time periods, using the 2012 Humans vs. Zombies data.

### Parameter Values for the First Time Period ($t = 0\text{ h to } t = 62\text{ h}$)

<table>
<thead>
<tr>
<th>Model Name</th>
<th>$k$</th>
<th>$b$</th>
<th>$a$</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWZC + sleep cycles</td>
<td>$0.00526^*$</td>
<td>0</td>
<td>--</td>
<td>$3.7474 \times 10^{-4}$*</td>
</tr>
<tr>
<td>HZC</td>
<td>$0.00178^*$</td>
<td>0</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>HWZC</td>
<td>$0.00245^*$</td>
<td>0</td>
<td>--</td>
<td>$1.8208 \times 10^{-4}$*</td>
</tr>
<tr>
<td>Type II predator-prey</td>
<td>$0.27667^*$</td>
<td>0</td>
<td>$99.99999^*$</td>
<td>--</td>
</tr>
<tr>
<td>Type III predator-prey</td>
<td>$0.30452^*$</td>
<td>0</td>
<td>$75.74294^*$</td>
<td>--</td>
</tr>
</tbody>
</table>

### Parameter Values for the Second Time Period ($t = 62\text{ h to } t = 75\text{ h}$)

<table>
<thead>
<tr>
<th>Model Name</th>
<th>$k$</th>
<th>$b$</th>
<th>$a$</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWZC + sleep cycles</td>
<td>$1.4125 \times 10^{-10}$</td>
<td>$0.00543^*$</td>
<td>--</td>
<td>0</td>
</tr>
<tr>
<td>HZC</td>
<td>$3.8067 \times 10^{-8}$</td>
<td>$0.01301^*$</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>HWZC</td>
<td>$1.0031 \times 10^{-12}$</td>
<td>$0.00573^*$</td>
<td>--</td>
<td>$1.1102 \times 10^{-15}$</td>
</tr>
<tr>
<td>Type II predator-prey</td>
<td>$8.2712 \times 10^{-14}$</td>
<td>$0.01431$</td>
<td>$0.10000$</td>
<td>--</td>
</tr>
<tr>
<td>Type III predator-prey</td>
<td>$2.4016 \times 10^{-9}$</td>
<td>$0.00649$</td>
<td>$75.74294$</td>
<td>--</td>
</tr>
</tbody>
</table>

### Parameter Values for the Third Time Period ($t = 75\text{ h to } t = 164\text{ h}$)

<table>
<thead>
<tr>
<th>Model Name</th>
<th>$k$</th>
<th>$b$</th>
<th>$a$</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWZC + sleep cycles</td>
<td>$0.001214$</td>
<td>$0.04722^*$</td>
<td>--</td>
<td>$0.001214$</td>
</tr>
<tr>
<td>HZC</td>
<td>$3.9048 \times 10^{-5}$</td>
<td>$0.03786^*$</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>HWZC</td>
<td>$4.8610 \times 10^{-4}$</td>
<td>$0.04632^*$</td>
<td>--</td>
<td>$4.8596 \times 10^{-4}$</td>
</tr>
<tr>
<td>Type II predator-prey</td>
<td>$8.2712 \times 10^{-14}$</td>
<td>$0.03579$</td>
<td>$0.10000$</td>
<td>--</td>
</tr>
<tr>
<td>Type III predator-prey</td>
<td>$0.04558$</td>
<td>$0.00815$</td>
<td>$0.05000$</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note: * Significance at 1% level.

Table 3: The sum of squared residuals, sample size, number of parameters, and AIC values for each 2013 model. Data was collected hourly from hour zero until gameplay concluded at 114 hours, for 342 data points.

<table>
<thead>
<tr>
<th>Model Name</th>
<th>SSR ($\chi^2$)</th>
<th>Sample Size (N)</th>
<th>Number of Parameters (K)</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWZC + sleep cycles</td>
<td>10683.79</td>
<td>342</td>
<td>5</td>
<td>1187.05</td>
</tr>
<tr>
<td>HWZC</td>
<td>11196.45</td>
<td>342</td>
<td>5</td>
<td>1203.08</td>
</tr>
<tr>
<td>Type III predator-prey</td>
<td>13789.36</td>
<td>342</td>
<td>5</td>
<td>1274.32</td>
</tr>
<tr>
<td>HZC</td>
<td>15489.60</td>
<td>342</td>
<td>3</td>
<td>1310.08</td>
</tr>
<tr>
<td>Type II predator-prey</td>
<td>28830.55</td>
<td>342</td>
<td>5</td>
<td>1526.56</td>
</tr>
</tbody>
</table>
Humans vs. Zombies: Data-driven Modeling of Disease Spread
Simeonov, Lemelin Fliss, Driscoll, Diaz Eaton

Figure 4: Best-fit model (SIR with a combination of multiple susceptibility classes and sleep cycles) graphed with the data for 2013. Blue is for susceptible individuals (Humans & Warriors), orange is for infected (Zombies), and red is for removed (Corpses).

The sleep cycles of the players and the different susceptibility levels can play a crucial role in determining the rate of the disease spread in the game. This adds to the body of literature about how human behavior may affect disease spread.

Comparing the two predator-prey models leads us to believe that the zombies in the game are not limited in their capacity to infect humans and that increased learning time might enhance the zombies’ abilities to capture humans. Predator prey with functional response type II assumes that the predators need a certain amount of time to capture and consume their prey. However, the predator-prey type II model we examined shows the worst fit of all five models in both years. Thus, we conclude that the zombies can catch large numbers of humans without facing a time constraint in the consumption time, and the zombie disease in our game can spread uncontrollably unless all zombies are eradicated. This is in accordance with the findings of [13] who claim that quarantining the infected would be useless unless all zombies are put in quarantine.

The predator-prey model with functional response type III assumes that the zombies would infect fewer humans towards the end of the game because there are fewer susceptible individuals towards the end of the game. The predator-prey type III model we used proved to be the second best fit for the data in 2012, and thus, we believe that the zombies’ consumption rate decreases as time progresses. However, this model doesn’t explain the large population of humans surviving at the end of the 2012 gameplay. Additionally, predator-prey type II and type III models assume a smaller consumption rate for low density as compared to the HZC model.

Therefore, we take a different approach and examine differences in the susceptible class as opposed to the infected class. We build the HWZC model to account for different infection rates in the susceptible class. The concept of introducing warriors in the game is equivalent to observing humans who have higher natural resistance to a virus and are very hard to infect. The fit of this model is on par with the predator-prey III model, implying that variations in both the susceptible class and the infected class are important for the analysis of the disease spread. These findings add to the HvZ research of [11]. They find that while the classical SIR model provided the worst fit, models that incorporated mechanisms of handling time (time required for the predation interaction) and variation in susceptibility didn’t significantly improve the model. We believe this is due to the different culture of players in the two schools the game was played. Utah State University has much larger student body than Unity College. The two schools also differ in the environment surrounding them and the importance of the game for campus life.

Further development of the model with key behavioral

non-linear least squares technique are summarized in Table 4. Differential evolution and dual annealing provide similar results.

We notice that in the first time period the parameter for infection rate \( (k) \) is smaller for the SIR model compared to the two multiple susceptibility modifications. However, in the second time period the three SIR-based models have similar infection rates. We also notice that in the second time period we obtain almost identical results for the starvation rate in all five models. Also, the two models including multiple susceptibility classes examine similar warrior infection rates \( (h) \), while the two predator prey models have completely different search rates \( (a) \). We use the best-fit parameters to plot the best-fit model against the data in Figure 4.

There are several key differences between the two years that should be noted. In 2012, we estimate that nine humans survived, but we don’t believe any humans survived in 2013. We had 7 zombies on the last day of the first year and 84 the following year. The length of each game was different as well, the first game lasted 164 hours, while the second lasted 115 hours. In general, movement among classes was slower during the second period of data collection in 2013.

4 Discussion

We implement five disease models to accommodate for differences in the classes and conclude that incorporating
Humans vs. Zombies: Data-driven Modeling of Disease Spread

Simeonov, Lemelin Fliss, Driscoll, Diaz Eaton

Table 4: Parameter values for the five models for all time periods, using the 2013 Humans vs. Zombies data.

| Parameter Values for the First Time Period \((t = 0\ h\ to\ t = 62\ h)\) |
|-----------------|-----------------|-----------------|
| Model Name       | \(k\) \(b\) \(a\) \(h\) |
| HWZC + sleep cycles | 0.00206* | 0 | – | \(1.7486 \times 10^{-14}\) |
| HZC              | \(7.6975 \times 10^{-4}\)* | 0 | – | – |
| HWZC             | 0.00104* | 0 | – | \(9.5035 \times 10^{-13}\)* |
| Type II predator-prey | 0.07365 | 0 | 0.24000 | – |
| Type III predator-prey | 0.15980* | 0 | 99.88919*** | – |

| Parameter Values for the Second Time Period \((t = 62\ h\ to\ t = 115\ h)\) |
|-----------------|-----------------|-----------------|
| Model Name       | \(k\) \(b\) \(a\) \(h\) |
| HWZC + sleep cycles | \(4.5846 \times 10^{-4}\) | 0.00549* | – | \(4.5846 \times 10^{-4}\) |
| HZC              | \(2.2957 \times 10^{-4}\)* | 0.00559* | – | – |
| HWZC             | \(2.7660 \times 10^{-4}\) | 0.00547* | – | \(2.7660 \times 10^{-4}\) |
| Type II predator-prey | 0.00429 | 0.00553* | 3.6641 \times 10^{-9} | – |
| Type III predator-prey | 0.00845* | 0.00576* | 11.10767*** | – |

**Note:** *, ** Significance at 1% and 10% levels, respectively.

characteristics continues to improve the model performance. We build upon the HWZC model, adding sinusoidal sleep cycles, to observe if the sleep pattern of the players would affect the results. The sinusoidal model is designed to emulate player sleep cycles and accommodate those peaks and lulls in activity that occur over a 24 hour period of time. Players don’t actually play the game for over 100 hours straight. They keep their regular diurnal routines throughout the game week. In the middle of the night when all the players are sleeping in “no-play zones” for hours at a time, gameplay essentially stops. Incorporating this human behavior characteristic is crucial given the timescale of data collection.

If we re-conceptualize “sleep” as a short-term implementation of quarantine, this results agree with [14]. For diseases where the researcher is more interested in short term transient disease behavior, these hours of non-activity can become important. However, real-life diseases progress much more slowly often, in part because in HvZ, an infected individual is immediately identified as a zombie. Regardless, the inclusion of handling times, a warrior class, and/or sleep as disease dynamic drivers points to the importance of capturing behavior in disease models.

A comparison in the parameter values between the models could also give us valuable insights. For example, we notice that the rate at which the humans move to the zombie class \((k)\) increases as we build upon the HVZ model, adding behavioral characteristics. As we concluded from the AIC analysis, the models that include more behavioral characteristics present a better fit for the data, so we conclude that lacking these modification leads to underestimation of the transition rate from humans to zombies.

We also notice that while in 2012 the transition rate from warriors to zombies \((h)\) stayed consistent between the two time periods, in 2013 the rate increased greatly. This is probably due to the differences between the two years and the lack of cure administration in 2013. Therefore, we conclude that control strategies can have an effect on the overall population, affecting not only the regular susceptible class, but also the stronger individuals, i.e. the warriors in our model.

4.1 Comparing and contrasting 2012 to 2013

The length of the game, number of players, and absence of a control strategy are the only changes made between the two years. However, once data collection for the second year was complete, it became obvious that the parameters were very different from the year before. The biggest difference between the two years is the speed at which the disease spread and endured. The 2012 game has individuals moving very quickly from one class to another (i.e., Human to Zombie). In the 2013 game, individuals stay in each class longer. Humans are less susceptible to the disease.

The differences among years of data collection could have occurred for several reasons, related to HvZ game design. The difficulty level of each mission is important to consider. We did not account for varying mission difficulty when each game was organized. Missions were considered to have one purpose (keep the populations mixing). The 2012 game began with a difficult mission for the humans, whereas, the 2013 game started with an easy
human mission. This could have been what set the stage for a faster acting disease versus a slower moving one. The length of the game is also important to consider. With a slower moving disease and a shorter time frame, the humans in the second year of data collection were overwhelmed by the vast number of zombies who hadn’t starved yet.

Another observation we make is that zombies worked together as a team, helping each other stay alive in 2013, whereas zombies were more solitary the year before. Another factor that likely caused the zombies to move too quickly the first year was a group of three or four humans who sacrificed themselves to the zombie horde and proceeded to systematically wipe out every human they saw. Those players graduated before the second year of data collection occurred. The players who stayed both years took a different approach the second time by sharing tags to keep each other alive as long as possible.

With more years of data, we may be able to more effectively manage disease spread in the HvZ gameplay. In the 2012, game we intervened too late, but by changing the timing and number of individuals affected by a control strategy the model is capable of providing a different outcome. This information can be used to plan future games of HvZ, as it is common for there to be target human zombie ratios at the end of gameplay. With a concerted effort at meta-analysis, we might understand how these games play out differently on different campuses as a way to understand how factors like campus size and culture (e.g., commuter versus residential) may affect the spread of other diseases. This is an opportunity to track disease spread with data from games, as well as, engage students in a fun and educational activity [14].

4.2 Educational component

Although the contribution of this study is in the arena of gathering potential game data upon which to answer disease modeling questions, this project is also one which garners plenty of student attention. It also is an excellent project, ripe for any question at the interface of data and modeling. We encourage readers to consider implementing HvZ at their school as a centerpiece of a problem-based modeling course. In fact, many, if not most, campuses already run games, and so it may be as easy as connecting with the organizers to understand their gameplay methods. Additionally, Students have been able to recreate the database by checking count results hourly.

The structure for adapting this HvZ modeling tool to your own college game would depend on many factors: for example, whether or not this game is already played on your campus, the mathematical preparation of students, and the student and instructor level of involvement in the game planning and execution. The calculus sequence at Unity College at the time of the game already introduced students to differential equations, so co-authors Driscoll and Lemelin, who had taken a biological modeling-oriented Calculus I and II, were prepared to engage in modeling with ODEs (see full curriculum description in a previous paper [6]). The subsequent independent study focused on qualitative analysis of systems of ODEs and the applications to disease in particular using python and gave students the time and space to think about how this applied to game planning, management, and data collection and management. Not all students worked on the HvZ game in this course, but each student came with “favorite disease” in mind, including Lyme disease and chronic wasting disease (found in Cervids).

Our dataset can certainly be modeled in other ways, and can be used to ask questions other than those posed above. For example, Dr. Erin N. Bodine at Rhodes College has advised undergraduate students in the creation of NetLogo models for HvZ which can incorporate individual behavior variation as well as spatial data, and Dr. Meredith Greer at Bates College taught a zombie-inspired disease modeling seminar, using Stella. This approach also allows for variation between individuals and a consideration of game-play history, which is a limiting assumption of our ordinary differential equation approach. If students have a course in partial differential equations or agent-based modeling, they might be able to consider spatial models or utilize time-delay approaches. We used three parameter optimization techniques: non-linear least squares, differential evolution, and dual annealing offered in the scipy.optimize Python package. However, there are other techniques such as Markov-Chain Monte Carlo (MCMC) and machine learning that can be used to find the best-fit parameters. We could imagine a possible integration into an optimization course where it represents an opportunity to engage students in exploring how numerical optimization methods work for data-fitting.

Please note that the game leadership takes a significant portion of time. If you already have a class devoted to ODEs and the game leadership is run separately, for example through a HvZ club as is done on many campuses, this could be adapted simply as a course module. However, we found that many of the insights in model behavior, for example, the observation of warrior players which lead us to pursue multiple susceptibility classes, came as a direct result of the hands-on participation of the students in both model development and gameplay and leadership.

To our knowledge, in 2012, our lab at Unity College was the first to undertake modeling the HvZ gameplay data. However, after sharing results, other colleges became involved, customizing these unique contexts. For example, the Bodine implementation at Rhodes College was in a discrete modeling course, offered post-calculus.
The Greer implementation at Bates College was in a first-year course assuming no prior knowledge of calculus or modeling, making modeling and computation even more accessible. In both cases, these implementations were facilitated by the HvZ management team being a separate team than the modeling team. Co-author S had prior experience playing the HvZ game at Bates, but no prior experience in modeling prior to this project. Co-authors L and D both benefited from prior gameplay experience before running their own versions of the game. We used this project to explore ODE modeling in preparation to engage in thesis research on COVID-19 modeling at small campuses.

We are excited to welcome you into the HvZ research fun! In the Appendix, we have included a suggested protocol for those who would like to get started modeling Humans vs. Zombies. For those colleges that are using similar protocols, we encourage collaboration between HvZ modeling sites to understand the differences in how this “disease” is transmitted across different types of campuses or how effective some models are over others for game prediction and human-zombie management. This may also need additional student survey data to elicit the proportion of warrior class members and the types of student strategies used on campus so that behavioral differences can be accounted for in any meta-analysis. In the spirit of both Open Education and Open Science, we have also freely provided the code used for this study in a GitHub repository [18] and have created a group on QUBES you can join to access additional resources and ask questions [3]. The HvZ source website that we used no longer exists. However, there are open alternatives which guarantee access to your own data [21].

5 Student Researcher Reflections

OS: Humans vs. Zombies provided me with a unique opportunity to explore disease modeling through the lens of a fun and engaging game. Working on this project helped me understand the dynamics of disease spread and build the models and code for my undergraduate math thesis about COVID-19 modeling. However, the most valuable aspect of HvZ was how accessible it made the topic of ODE disease modeling for my friends and peers. People who had not taken any math classes in years would come to my presentation and tell me that they never believed math could be so interesting and easy to understand. It was truly inspiring to use HvZ not only as a research project, but also as a tool for making mathematics more inclusive.

KLF: HvZ and the modeling that went with it was an empowering experience for me. Whether players were interested in or aware of the math involved didn’t matter for their enjoyment of the game. Everyone could choose to go as deep with it as they wanted and as someone who has always enjoyed math I had the pleasure of seeing that interest piqued in my peers who were more math averse.

JD: Through the guise of an unconventional game of tag, HvZ allowed me to realize how insightful ODE modeling can be when predicting real-world disease behavior. During my four years at Unity College, I helped in data collection for the game, as well as analyzing and modeling it into a digestible, educational component for the students involved. This led me to further my exploration of the applications of ODEs in disease modeling in an independent study my senior year. I learned to incorporate different variables into the disease models used and to communicate my findings in a way that showed their practicality. Assisting in the implementation of the game and interpretation of the HvZ data helped me to develop the leadership, communication, and analytical skills that I continue to use throughout my career.

6 Conclusion

We find that data from simulations can be essential in modeling disease spread in close-knit communities. This article provides evidence for some well-established facts in disease modeling: variations exist in both the susceptible and the infected classes, and these variations play a crucial role in the spread of diseases in the community. There are individuals with stronger immune systems that are hard to infect, as demonstrated by the warrior class in our models. We also find that the sleep cycles of the players can significantly affect the disease spread as well.

Besides yielding intriguing insights to disease modeling in close-knit communities, we find that using game simulation to motivate mathematical modeling investigations is a useful pedagogical tool. As students reflect, not only did involvement in the game help them build an important toolset, but even those more math-averse engaged. Building appreciation for mathematics such as disease modeling is important, as we work to improve relationship between science and the public.

Acknowledgments

This work would not have been possible without the help of the student moderators who worked hard on organizing and planning each game. We also thank undergraduate students Timothy Beaulieu and Michelle Lewis, who authored the QUBES HvZ site. The authors would also like to thank Dr. Erin Bodine, Dr. Alex Capaldi, Dr. Jean Marie Linhart, Dr. Robert Smith?, and Dr. Meredith Greer for sharing their HvZ modeling work and experience. We would also like to thank Dr. Meredith Greer for
providing suggestions that helped improve and clarify this paper. Lastly, the authors are thankful to the group who created the game for providing such a fun experience for students and an excellent opportunity for an introduction to mathematical modeling.

Author Contributions

Authors KLF and JD worked on the data collection and the creation of the initial models and programs in MATLAB and R. Author OS extended the models and created the parameter optimization program in Python available at [18]. Author CDE oversaw the whole project from its very beginning, providing academic support and guidance for the undergraduate authors.

Appendix

The game rules can be found at [22]. However, in the following Appendix we take into consideration the data collection and modeling aspects of the game organization.

The Game  Humans vs. Zombies is a game of tag. All players begin as humans, and one is randomly chosen to be the “Original Zombie.” The Original Zombie tags human players and turns them into zombies. Zombies must tag and eat a human every 48 hours or they starve to death and are out of the game. Game rules can be found at humansvszombies.org [22].

Objective  The Zombies win when all human players have been tagged and turned into zombies. The Humans win by surviving long enough for all of the zombies to starve.

Equipment  This gear is required for all players: Bandana, Foam Dart Blaster, marshmallow launcher, and/or socks, and One 3×5 index card (for player ID)

Game Creation  Plan for approximately 5 moderators for every 100 players. Moderators should be told upfront that this is a one to two times a week commitment for the two months leading up to the game, and the work load only gets larger the closer you get to game day.

Game Timeline  To use your data, you will need about a month and a half to work on the math. In the fall semester that means the game will run at the end of October. That leaves two months to prepare for the game. Have an organized schedule for regular moderator meetings and be flexible around your fellow moderators’ schedules.

Moderator Meetings  Meet regularly with your moderators and create a schedule that works for everybody at the first meeting. Assign moderators with different parts of the project: zombies, humans, plot design, mission design, marketing, data collection, photography, and videography.

Plot  Get creative if you decide to follow a plot line. Create a story that missions can work with. Plot ideas can be found on the HvZ wiki page: wiki.humansvszombies.org.

Missions  Missions are critical for population mixing and keeping players engaged. The missions prompt the humans and the zombies to gather at the same locations, and thus, lead to the mixing of the disease carriers and the susceptible individuals. Missions do not need to follow normal game rules. You decide how the rules change and you should be upfront about it to the players. They like to know what to expect. You may introduce non-playing characters (NPC) into missions to enhance the missions or the storyline. There are several mission types: point defense, item retrieval (capture the flag style or search), escort, puzzle, and assassination. Mission types can be combined to enhance game play.

Point Defense missions  Humans must defend a person, place or object for a specified amount of time. Does not work as well when zombies defend. Good for a final mission, humans run out of ammo easily. Ex: Humans must survive in a field for 15 min. until a helicopter arrives. Zombies do not have a stun time, instead when they are stunned they must run to the picnic table in front of a building on campus (respawn point) and touch the center to become unstunned.

Item retrieval missions  Works for both humans and zombies. They may be used to distract zombies during a difficult human mission. They can be set up like capture the flag or the item could be hidden. Ex: Humans must build a radio tower to get help and collect supplies around campus. Zombies need to build a fridge for blood and must collect supplies from around campus.

Escort missions  Humans or zombies escort an important person from location A to location B. Ex: Humans must find the scientist wandering around campus and escort her to a building on campus.

Puzzle missions  Humans or zombies must complete a puzzle to win the mission. Ex: Zombies must complete a half constructed Lego set.
Assassination mission  Humans or zombies find and eliminate (stun) human or zombie, player or NPC. Ex: Humans must find and eliminate a mad scientist who is working with the zombies.

Marketing  Use several different methods to get the word out. Posters are easy to create and videos can start the storyline and get people interested. Have moderators run around outside of sign-up tables dressed as humans and zombies. This can get people really excited for the game. If possible try to have human moderators “protect” students as they approach nearby buildings. Set-up game sign-up at cafeteria, student center or anywhere students congregate. Just make sure you have permission to set up a table there first.

Data Collection  Here is a comprehensive list of the data required to model a game of HvZ: 1. total number of players at each time step; 2. total number of humans at each time step; 3. total number of zombies at each time step; 4. zombie starvation time; 5. length of game; 6. time of control strategy used (if any) and number of players affected. To keep this model and data comparable across multiple campuses some pieces of information must remain the same. We recommend the following considerations: Keep the zombie starvation time at 48 hours and keep missions easy for humans in the beginning and make them harder as the game progresses. Our research led us to use the simple SIR model as explained above, which should be used to model the game on other campuses as well.

Python  We used primarily Python to emulate and graph the models. However, there are also other open source software programs with similar capabilities such as R, with the GUI R Studio. The original coding work done by author KLF used MATLAB, however we emphasize open source software to lower cost barriers for students and for classroom adoption. Author JD, a wildlife biology major, moved this work into R, since R is used heavily in quantitative biology. Author OS built on this prior work, implementing the work in python and creating additional models to consider. All code and data used for this study in a GitHub repository [18], and all readers are welcome to use it or modify it. We would be happy to participate in the future development of this code.

References


