The Impact of Preferred Mixing on the Sexual Transmission of ZIKV at the 2016 Summer Olympics

Hilary Marrero[1], Whitney Mgbara[2],
Pradyuta Padmanabhan[3], Hailey Stirneman[4],
Baltazar Espinoza[2], Baojun Song[5]

[1] University of Puerto Rico, Mayaquez, PR

[2] Arizona State University, Tempe, AZ

[3] Foxcroft School, Middleburg, VA

[4] Texas State University, San Marcos, TX

[5] Montclair State University, Montclair, NJ

Abstract

Zika virus (ZIKV) has started to plague South and Central America. Following the outbreak in Brazil, cases of microcephaly and other neurological disorders arose steadily. Health officials linked these cases to ZIKV which led the International Health Regulations (IHR) committee to deem ZIKV as a Public Health Emergency of International Concern (PHEIC) in 2016. Despite the control strategies, ZIKV continues to spread in Brazil where the 2016 Summer Olympic games are to be held. Mass gatherings such as sporting events typically increase sexual activity and the spread of sexually transmitted diseases. Thus holding the Olympics in Rio de Janeiro subjects people to a range of health risks. This was confirmed in February 2015 when researchers found that males can transmit the disease via sexual intercourse. Using SIR models, we investigate three male-seeking-male (MSM) subpopulations: male natives in Rio, male visitors from outside of Rio, and male sex workers in Rio. We formulated the basic reproduction number and also calculated implicitly the final size of the epidemic. For specific values of preference, we calculated a basic reproductive number of 4.7886 which is similar to 4.4, the basic reproductive number calculated by Dr. Towers and colleagues for the Zika outbreak in Barranquilla, Colombia [25]. The final PRCC charts show that the most sensitive parameters are c_1, c_2, c_3. This indicates that the most important parameters to consider when trying to decrease the infected individuals of each subpopulation is the number of sexual contacts that each individual is involved in. Overall, the preference parameter, f_i, does not reduce the final size for a population the way the c_i parameter can for final size distribution.

Summary

1 Introduction

1.1 ZIKV Epidemiology and Global Spread

ZIKV is considered part of the *flavivirus* genus along with other viruses such as dengue and West Nile virus. Zika virus is classified as an arbovirus due to its dominant transmission mechanism coming from an infected arthropod: the female mosquito of the *Aedes aegypti* genus [20]. In 1947, the virus was isolated from a *rhesus* monkey in the Zika forest of Uganda, where ZIKV received its name [20]. By 1954, ZIKV spread to West Africa with recurrent outbreaks reported not only in tropical Africa, but also in Southeast Asia [12]. Between 2007 to 2016, a total of 44 countries and territories documented Zika viral transmission; 33 of which reported transmission between 2015 and 2016 [21]. This recent rise in the spread of Zika virus (ZIKV) has given health officials cause for concern.

The Yap Islands of the Federated States of Micronesia reported in 2007, the first large outbreak of ZIKV that was outside of Asia and Africa [20]. After 2015, the disease spread to areas outside of the traditional regions, eventually reaching South and Central America. ZIKV eventually progressed to in Brazil in May 2015 [9]. At the time, researchers believed that ZIKV was originally introduced during the 2014 FIFA World Cup event in where no Pacific countries with documented ZIKV had competed. It was also suggested that ZIKV was introduced in Rio de Janeiro (Rio) during the August 2014 Va'a World Sprints Canoe Championship, an event where various Pacific countries participated. Mass gatherings such as the previously mentioned events, attract global travelers, which exposes many people to a large range of health risks. Now, there is data available and the situation does not seem safe: Rio's suspected Zika cases are the highest of any state in Brazil (26,000), and its Zika incidence rate is the fourth worst (157 per 100,000) [1]. According to Brazil's official data, Rio is not on the fringes of the outbreak, but inside its heart [1]. This is not comforting data, considering that the city of Rio de Janeiro is the location of the 2016 Summer Olympics.

1.2 Transmission of ZIKV in Brazil and Olympics

The dominant ZIKV transmission mechanism in Brazil is via mosquitoes which are prominent in tropical regions. Typically, the virus is transmitted from human to human by the bite of an infected female mosquitoes [20]. Infected humans are the main carriers and spreaders of ZIKV. Humans serve as a source for uninfected mosquitoes because when a mosquito feeds on an infected human, it becomes infected as well [17]. An infected mosquito can transmit ZIKV for the rest of its life. After an infected mosquito bite, disease symptoms usually appear following an incubation period of three to twelve days [17].

In addition to incidences of mosquito-borne ZIKV transmission, ZIKV has been detected in serum, saliva, urine, and semen [14]. These incidences led to the realization that ZIKV can be transmitted sexually through vaginal, anal, and oral intercourse [20]. Since 2015, countries such as Argentina, Canada, Chile, France, Italy, New Zealand, Peru, Portugal, and more recently, the USA have had cases of sexually transmitted ZIKV [9]. This is a concern because within the span of a month, six cases of sexually transmitted ZIKV arose throughout the United States, starting with the first incident in Dallas, TX on February 2, 2016 [9] [18]. It was also found that Europe had its first case of sexually transmitted ZIKV in France on February 2016 [12].

ZIKV infections are often asymptomatic or have mild symptoms that can last around two to seven days and have an insignificant impact on sexual activity [25]. Since ZIKV is sexually transmissible, it is necessary to avoid sexual contact or to use preventive measures such as condoms. This is particularly important to pregnant women in areas where ZIKV is circulating. Based on the reported dengue data from 2015, WHO estimated that up to four million people in the Americas could be infected by ZIKV in 2016 [21]. Without effective intervention, the situation has the potential to worsen, due in part to the upcoming 2016 Summer Olympics in Rio de Janeiro (Rio).

1.3 Model Overview

When ZIKV emerged in South America, the spread of ZIKV was rapid throughout South and Central America, reaching Mexico in November 2015 [9]. As the number of incidences of microcephaly, Guillain-Barré syndrome, and other neurological disorders linked to ZIKV increased, the

World Health Organization (WHO) declared ZIKV a Public Health Emergency of International Concern (PHEIC) on February of 2016 [19]. ZIKV is spreading throughout the South and Central Americas and is a potential threat to the visitors and participants of the 2016 Rio Olympics [9]. Another concern is that in Brazil, being a sex worker is legal [7]. Sex is not frowned upon and there is expected to be an increase in sexual activity during the Olympics.

Males have been known to sexually transmit ZIKV due a large viral load of ZIKV found in their semen that can last up to six months [14]. Since male individuals represent a long term reservoir for ZIKV, they are considered super spreaders. There is a large amount of men who have sex with men (MSM) in Brazil and Rio in particular, contains 14.30 % of Brazil's LGB population [22]. This, as well as the fact that sex working is legal in Brazil, led us to contruct a male-to-male model where we look at three groups: homosexual males that are visitors outside of Rio, natives in Rio who are not sex workers, and male sex workers in Rio.

The overall goal of this paper is to find out how the behavior of MSM affect the transmission dynamics of ZIKV at the 2016 Olympic Games in Rio. More specifically, to show that heterogeneous mixing enhances the transmission of ZIKV. Therefore, we propose a deterministic ODE system of equations.

2 Methods

2.1 Model Framework

For our model, we consider the sexual transmission of ZIKV and take into consideration male-to-male sexual contact. We focus on the total population of men seeking men (MSM) at Rio de Janeiro during the Olympics. We then split that population into three subpopulations: MSM visitors outside of Rio, natives in Rio, and male sex workers in Rio. We then describe each of those populations as a separate class. The dynamics of these three separate populations incorporate a short time scale, 30 days, and the analysis is over the duration of a single outbreak.

2.2 Assumptions

Since the incubation period of the disease is relatively small, it is neglected in this research. Thus a typical individual becomes infectious instantaneously after being infected. It is also assumed that initially all visitors are susceptible [20].

Disease free equilibrium was computed in order to get an analytical expression of R_0. However, the endemic equilibrium is not computed due to the short time period of the Olympics. The Olympics are 17 days long, but we are looking at a 30 day time scale due to taking traveling into consideration.

We let f_i notate an individual's preference of another individual in the same group. We can assume that $f_3 << f_1 < f_2$. This is because we have conclude that male sex workers are the least likely to have sexual contact within their group due to the lack of gain from the experiences versus sexual contacts with the other groups. We then assume visitors will be next to have contact within their own group due to the fact that they will most likely be mixing with the male sex worker or the male native groups. We assume that the native population will have the most mixing within their own group because the Olympics will affect them the least. We are assuming that they will behave as normal for the most part, and will mostly keep to themselves despite the inflow of visitors from the event.

2.3 ZIKV Epidemic Model

We denote N_i as the total subpopulation size of each class, with i establishing which subpopulation we are looking at. We let $i = 1, 2, 3$ to denote visitors, natives, and male sex workers, respectively. The transmission process is modeled through the interactions of the following epidemiological variables. We let S_i, I_i, and R_i denote the susceptible, infectious, and recovered individuals respectively,

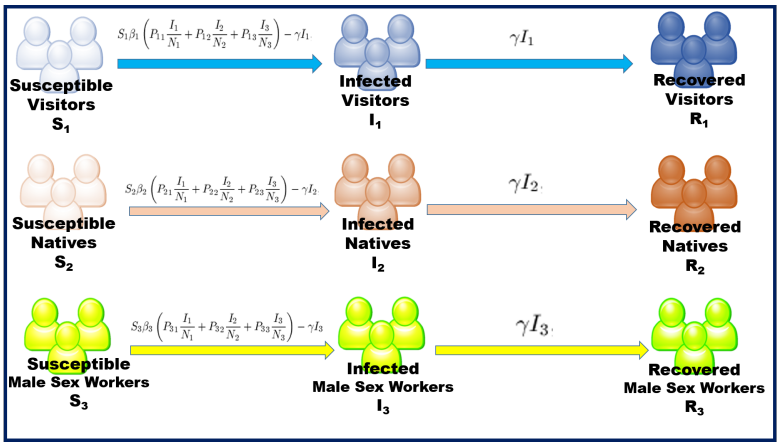


Figure 1: Flowchart for the SIR model of each subpopulation.

Our model describes the impact of heterogeneous mixing in Rio during the Olympics and the deterministic ODE system of equations is given by:

$$\begin{aligned}
\frac{dS_1}{dt} &= -S_1\beta_1\left(P_{11}\frac{I_1}{N_1}+P_{12}\frac{I_2}{N_2}+P_{13}\frac{I_3}{N_3}\right),\\
\frac{dI_1}{dt} &= S_1\beta_1\left(P_{11}\frac{I_1}{N_1}+P_{12}\frac{I_2}{N_2}+P_{13}\frac{I_3}{N_3}\right)-\gamma I_1,\\
\frac{dR_1}{dt} &= \gamma I_1,
\end{aligned}$$

$$\begin{aligned}
\frac{dS_2}{dt} &= -S_2\beta_2\left(P_{21}\frac{I_1}{N_1}+P_{22}\frac{I_2}{N_2}+P_{23}\frac{I_3}{N_3}\right),\\
\frac{dI_2}{dt} &= S_2\beta_2\left(P_{21}\frac{I_1}{N_1}+P_{22}\frac{I_2}{N_2}+P_{23}\frac{I_3}{N_3}\right)-\gamma I_2,\\
\frac{dR_2}{dt} &= \gamma I_2,
\end{aligned}$$

$$\begin{aligned}
\frac{dS_3}{dt} &= -S_3\beta_3\left(P_{31}\frac{I_1}{N_1}+P_{32}\frac{I_2}{N_2}+P_{33}\frac{I_3}{N_3}\right),\\
\frac{dI_3}{dt} &= S_3\beta_3\left(P_{31}\frac{I_1}{N_1}+P_{32}\frac{I_2}{N_2}+P_{33}\frac{I_3}{N_3}\right)-\gamma I_3,\\
\frac{dR_3}{dt} &= \gamma I_3,
\end{aligned}$$

where the total number of each subpopulation is, $N_i = S_i + I_i + R_i$.

Table 1: Parameter Definitions and Units

Parameters	Description	Units
β_i	rate of infectivity	time per contact
c_i	contacts per unit time	$\frac{contacts}{time}$
f_i	preference of contacting an individual in the same group	unitless
γ	recovery rate	$\frac{1}{time}$

Table 2: Parameter Values

Parameters	Description	Range	Values	References
β_i	Rate of Infectivity	-	$c_i * .043$	[27]
c_1	Visitor's Contacts per Week	$(\frac{3}{28}, \frac{6}{28})$	$\frac{6.8}{28}$	[2]
c_2	Native's Contacts per Week	$(\frac{3}{28}, \frac{6}{28})$	$\frac{3.4}{28}$	[2]
c_3	Sex Worker's Contacts per Week	$(\frac{2}{7}, \frac{20}{7})$	$\frac{11}{7}$	[23]
f_i	Preference of Contacting an Individual in the Same Group	$(0, 1)$	Estimation	-
γ	Recovery Rate	-	$\frac{1}{180}$	[12]

We set $\beta_i = pc_i$, where p is the probability that a contact is effective for ZIKV transmission given that a susceptible has contact with an infected individual and c_i is the average number of contacts of group i per person per unit time. Therefore, β_i is interpreted as the average number of effective contacts a susceptible has per unit of time. Furthermore, when $f_i = f_j = 0$ the preference term is removed from the equation and there is simply proportional mixing between subpopulations.

We assumed the preferred mixing of the three subpopulations and used the formula,

$$P_{ij} = f_i \delta_{ij} + (1 - f_i) \frac{(1 - f_j) c_j N_j}{\sum_{n=1}^{3} (1 - f_n) c_n N_n} \tag{1}$$

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & \text{otherwise} \end{cases}$$

where P_{ij} = probability that an individual in group i has a contact with an individual on group j given that the i group has engaged in a sexual contact. Where f_i is an individual preference for their own group and (1-f_i) is an individual's preference for a different group [3] [11] [6].

Table 3: Subpopulation Values

Parameters	**Description**	**Values**	**References**
N_1	Population of MSM Visitors	$40,777$	[8] [24]
N_2	Population of MSM Natives	$130,718$	[16] [22]
N_3	Population of Male Sex Workers	$5,000$	[15] [22]
S_1	Susceptible Visitors	$40,777$	Estimation
S_2	Susceptible Natives	$129,760$	Estimation
S_3	Susceptible Sex Workers	$4,963$	Estimation
I_1	Initially Infected Visitors	0	[13]
I_2	Initially Infected Natives	958	[13]
I_3	Initially Infected Sex Workers	37	[13]
R_1	Initially Recovered Visitors	0	Estimation
R_2	Initially Recovered Natives	0	Estimation
R_3	Initially Recovered Sex Workers	0	Estimation

There are 500,000 expected visitors coming to the Olympics this year, and 242,718 will be male [8]. Since 21% of males in the world are men who have sex with men, there should be approximately 50,971 MSM males coming from outside of Rio to the Olympics this year [24]. However, we assume that only 80% of the MSM visitor population will be participating in sexual activity, which leads us to consider 40,776 visitor individuals in our SIR model.

There are approximately 3.082 million males in the city of Rio, and 14.30% of individuals in Brazil are MSM males, therefore our number of MSM male natives is 440,726 [22]. We then subtracted the male sex worker population of 5,000 individuals and that gave us 435,726 native individuals. However, we assumed that 3% of the MSM natives will be participating in the sexual activity during the Olympics, which leads us to consider 130,726 native individuals in our SIR model.

It is estimated that there is approximately at least 450 sex workers working on every given night at the Olympics [15]. We multiplied that number by 7 for each day of the week and received a number of 3,150 individuals. However, since that is the worse case scenerio, we decided to round that number to 5,000 individuals. Therefore, 5,000 male sex workers is the amount of individuals we consider in our model.

3 Mathematical Analysis

3.1 The Derivation of R_0

We are considering three subpopulations in Rio de Janeiro: male visitors, male natives, and male sex workers. In order to derive the basic reproductive number R_0, we use the Next Generation Matrix [4]. The computation of the basic reproduction number follows below. Two matrices X and Y are defined as: matrix X (disease compartments) is the infected population and matrix Y includes the susceptible and recovered populations. Matrix F includes the newly infected individuals and cannot be negative while matrix V includes the output from the infected compartments:

$$X^T = \begin{bmatrix} I_1 & I_2 & I_3 \end{bmatrix},$$

$$\mathrm{Y}^T = \begin{bmatrix} S_1 & S_2 & S_3 & R_1 & R_2 & R_3 \end{bmatrix},$$

$$\mathcal{F} = \begin{bmatrix} S_1\beta_1 P_{11}\frac{I_1}{N_1} + S_1\beta_1 P_{12}\frac{I_2}{N_2} + S_1\beta_1 P_{13}\frac{I_3}{N_3} \\ S_2\beta_2 P_{21}\frac{I_1}{N_1} + S_2\beta_2 P_{22}\frac{I_2}{N_2} + S_2\beta_2 P_{23}\frac{I_3}{N_3} \\ S_3\beta_3 P_{31}\frac{I_1}{N_1} + S_3\beta_3 P_{32}\frac{I_2}{N_2} + S_3\beta_3 P_{33}\frac{I_3}{N_3} \end{bmatrix}, \qquad \mathcal{V} = \begin{bmatrix} \gamma I_1 \\ \gamma I_2 \\ \gamma I_3 \end{bmatrix},$$

Considering the disease free equilibrium point (N_1,0,0,N_2,0,0,N_3,0,0), the matrix is as seen below:

$$F = \begin{bmatrix} \beta_1 P_{11} & \frac{N_1\beta_1 P_{12}}{N_2} & \frac{N_1\beta_1 P_{13}}{N_3} \\ \frac{N_2\beta_2 P_{21}}{N_1} & \beta_2 P_{22} & \frac{N_2\beta_2 P_{23}}{N_3} \\ \frac{N_3\beta_3 P_{31}}{N_1} & \frac{N_3\beta_3 P_{32}}{N_2} & \beta_3 P_{33} \end{bmatrix}, \qquad V = \begin{bmatrix} \gamma & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma \end{bmatrix}, zs$$

$$V^{-1} = \begin{bmatrix} \frac{1}{\gamma} & 0 & 0 \\ 0 & \frac{1}{\gamma} & 0 \\ 0 & 0 & \frac{1}{\gamma} \end{bmatrix}, \qquad FV^{-1} = \begin{bmatrix} \frac{\beta_1 P_{11}}{\gamma} & \frac{N_1\beta_1 P_{12}}{N_2\gamma} & \frac{N_1\beta_1 P_{13}}{N_3\gamma} \\ \frac{N_2\beta_2 P_{21}}{N_1\gamma} & \frac{\beta_2 P_{22}}{\gamma} & \frac{N_2\beta_2 P_{23}}{N_3\gamma} \\ \frac{N_3\beta_3 P_{31}}{N_1\gamma} & \frac{N_3\beta_3 P_{32}}{N_2\gamma} & \frac{\beta_3 P_{33}}{\gamma} \end{bmatrix},$$

Let: $a = \frac{N_1}{N_2}, b = \frac{N_1}{N_3}, c = \frac{N_2}{N_3} = \frac{b}{a}, f = \frac{\beta_1}{\gamma}, g = \frac{\beta_2}{\gamma}$ *and* $h = \frac{\beta_3}{\gamma}$

By these substitutions we are able to simplify the previous matrix to the following:

$$FV^{-1} = \begin{bmatrix} fP_{11} & afP_{12} & bfP_{13} \\ \frac{gP_{21}}{a} & gP_{22} & cgP_{23} \\ \frac{hP_{31}}{b} & \frac{hP_{32}}{c} & hP_{33} \end{bmatrix}$$

Therefore, the characteristic polynomial of FV^{-1} is given by:

$$P(\lambda) = |A - \lambda I| = -\lambda^3 + tr_1(A)\lambda^2 - tr_2(A)\lambda + tr_3(A)$$

For $A = FV^{-1}$ we obtain:

$$\begin{aligned}
tr_1 =& fP_{11} + gP_{22} + hP_{33},\\
tr_2 =& fgP_{11}P_{22} - fgP_{21}P_{12} + ghP_{22}P_{33} - ghP_{32}P_{23} + fhP_{11}P_{33} - fhP_{31}P_{13},\\
tr_3 =& fghP_{11}P_{22}P_{33} - fghP_{11}P_{32}P_{23} - fghP_{12}P_{21}P_{33}\\
&+ fghP_{12}P_{31}P_{23} + fghP_{13}P_{21}P_{32} - fghP_{13}P_{31}P_{22}.
\end{aligned}$$

The characteristic polynomial is now:

$$\begin{aligned}
P(\lambda) =& -\lambda^3 + (fP_{11} + gP_{22} + hP_{33})\lambda^2 - (fgP_{11}P_{22} - fgP_{21}P_{12}\\
&+ ghP_{22}P_{33} - ghP_{32}P_{23} + fhP_{11}P_{33} - fhP_{31}P_{13})\lambda + fghP_{11}P_{22}P_{33}\\
&- fghP_{11}P_{32}P_{23} - fghP_{12}P_{21}P_{33} + fghP_{12}P_{31}P_{23} + fghP_{13}P_{21}P_{32}\\
&- fghP_{13}P_{31}P_{22}.
\end{aligned}$$

In order to use the basic reproductive number formula [5], the polynomial is multiplied by -1 and is obtained $\lambda^3 - r\lambda^2 + s\lambda - t = 0$ and we need to get $\lambda^3 - r\lambda^2 + s\lambda + t = 0$. Then:

$$\begin{aligned}
r =& fP_{11} + gP_{22} + hP_{33}\\
s =& fgP_{11}P_{22} - fgP_{21}P_{12} + ghP_{22}P_{33} - ghP_{32}P_{23} + fhP_{11}P_{33} - fhP_{31}P_{13}\\
t =& -(fghP_{11}P_{22}P_{33} - fghP_{11}P_{32}P_{23} - fghP_{12}P_{21}P_{33} + fghP_{12}P_{31}P_{23}\\
&+ fghP_{13}P_{21}P_{32} - fghP_{13}P_{31}P_{22}).
\end{aligned}$$

Each one of the, R_0^{ij}, represents the basic reproductive number of an infected individual from

a sub-population i with a susceptible individual from a sub-population j. Those are the basic reproductive numbers for the interaction between two sub-populations. Since in this model we are considering three sub-populations, we have an R_0 that represents the interaction between three sub-populations. If the R_0 is bigger than 1, there is an epidemic, but if is less than 1 there is no epidemic.

$$
\begin{aligned}
R_0^{ij} =& \frac{\beta_i P_{ij} N_i}{\gamma N_j}, \text{ where } i,j \in \{1,2,3\} \\
R_0^{11} =& \frac{\beta_1 P_{11}}{\gamma}, R_0^{12} = \frac{\beta_1 P_{12} N_1}{\gamma N_2}, R_0^{13} = \frac{\beta_1 P_{13} N_1}{\gamma N_3}, R_0^{21} = \frac{\beta_2 P_{21} N_2}{\gamma N_1}, \\
R_0^{22} =& \frac{\beta_2 P_{22}}{\gamma}, R_0^{23} = \frac{\beta_2 P_{23} N_2}{\gamma N_3}, R_0^{31} = \frac{\beta_3 P_{31} N_3}{\gamma N_1}, R_0^{32} = \frac{\beta_3 P_{32} N_3}{\gamma N_2}, R_0^{33} = \frac{\beta_3 P_{33}}{\gamma}
\end{aligned}
$$

Then, the dominant root of the polynomial is the basic reproductive number [5]

$$
\begin{aligned}
R_0 =& \frac{r}{3} + \sqrt[3]{\left(\frac{2r^3 - 9rs - 27t}{54}\right) + \sqrt{\left(\frac{3s - r^2}{9}\right)^3 + \left(\frac{2r^3 - 9rs - 27t}{54}\right)^2}} \\
&+ \sqrt[3]{\left(\frac{2r^3 - 9rs - 27t}{54}\right) - \sqrt{\left(\frac{3s - r^2}{9}\right)^3 + \left(\frac{2r^3 - 9rs - 27t}{54}\right)^2}}
\end{aligned}
$$

Then, r, s and t are the coefficients of λ in the characteristic polynomial and are written in terms of the R_0^{ij}:

$$
\begin{aligned}
r =& R_0^{11} + R_0^{22} + R_0^{33}, \\
s =& R_0^{11} R_0^{22} - \frac{R_0^{11}}{P_{11}} \frac{R_0^{22}}{P_{22}} P_{21} P_{12} + R_0^{22} R_0^{33} - \frac{R_0^{33}}{P_{33}} \frac{R_0^{22}}{P_{22}} P_{32} P_{23} + R_0^{11} R_0^{33} \\
&- \frac{R_0^{33}}{P_{33}} \frac{R_0^{11}}{P_{11}} P_{31} P_{13} \\
t =& - R_0^{22} R_0^{33} R_0^{11} + \frac{R_0^{22}}{P_{22}} \frac{R_0^{33}}{P_{33}} R_0^{11} P_{32} P_{23} + \frac{R_0^{22}}{P_{22}} \frac{R_0^{11}}{P_{11}} R_0^{33} P_{12} P_{21} \\
&- \frac{R_0^{22}}{P_{22}} \frac{R_0^{33}}{P_{33}} \frac{R_0^{11}}{P_{11}} P_{12} P_{31} P_{23} - \frac{R_0^{22}}{P_{22}} \frac{R_0^{33}}{P_{33}} \frac{R_0^{11}}{P_{11}} P_{13} P_{21} P_{32} + \frac{R_0^{33}}{P_{33}} \frac{R_0^{11}}{P_{11}} R_0^{22} P_{13} P_{31}
\end{aligned}
$$

Special Case for the R_0:

If we want to consider just one population, we can set $f_i = 0$, $c_1 = c_2 = c_3$ and $N_1 = N_2 = N_3$.

If the contacts of the three subpopulations are the same, then the infectious rate is the same ($\beta_1 = \beta_2 = \beta_3$). Then $P_{ij} = \frac{1}{3}$ and $R_0^{ij} = \frac{\beta}{3\gamma}$.

Therefore, the $R_0 = \frac{\beta}{\gamma}$ like the simple SIR model for one population.

3.2 Final Size

Final Size is the final number of susceptible individuals in any population after the disease has ceased to exist ($I = 0$). Let $i = 1, 2, 3$ we can consider the original six "Susceptible" and "Infected" differential equations where they will be written as:

$$\begin{aligned}\frac{dS_i}{dt} &= -S_i\beta_i\left(P_{i1}\frac{I_1}{N_1} + P_{i2}\frac{I_2}{N_2} + P_{i3}\frac{I_3}{N_3}\right)\\ \frac{dI_i}{dt} &= \ S_i\beta_i\left(P_{i1}\frac{I_1}{N_1} + P_{i2}\frac{I_2}{N_2} + P_{i3}\frac{I_3}{N_3}\right) - \gamma I_1.\end{aligned}$$

Adding these equations, we get:

$$(S_i(t) + I_i(t))' = -\gamma I_i(t)$$

Since $(S_i(t) + I_i(t))' = -\gamma I_i(t)$, we can make the statement that $S_i(t) + I_i(t)$ is a positive decreasing function, therefore the limit exists. The derivative of a positive decreasing function tends to zero, and this yields that $\gamma I_i(t) \to 0$, and since $\gamma > 0$ this implies $I_i(t) \to 0$. The goal of this process is to get to a point where we can understand what $S_i(\infty)$, and in order to do this, we can integrate both sides of the equation from 0 to ∞. Integrating this equation we get

$$\begin{aligned}(S_i(t) + I_i(t))\,|_0^\infty &= -\gamma\int_0^\infty I_i(t)\ dt\\ (S_i(\infty) + I_i(\infty)) - (S_i(0) + I_i(0)) &= -\gamma\int_0^\infty I_i(t)dt\\ S_i(\infty) - N_i &= -\gamma\int_0^\infty I_i(t)\ dt\end{aligned}$$

The derivative of a positive decreasing function tends to zero, therefore $\gamma I_i(t) \to 0$, and since $\gamma > 0$ this implies $I_i(t) \to 0$. Hence, $S_i(0) + I_i(0) = N_i$. Then,

$$\gamma \int_0^\infty I_i(t)\, dt = N_i - S_i(\infty) \tag{2}$$

considering the original equation in a simplified form, we use the method of separation of variables [4]:

$$\begin{aligned}
\int_0^\infty \frac{dS_i}{S_i} &= -\beta_i \sum_{j=1}^{3} \frac{P_{ij}}{N_j} \int_0^\infty I_j(t), \\
\log(S_i)|_0^\infty &= -\sum_{j=1}^{3} \frac{\beta_i P_{ij}}{N_j} \int_0^\infty I_j(t), \\
\log\left(S_i(\infty)\right) - \log\left(S_i(0)\right) &= -\beta_i \sum_{j=1}^{3} \frac{P_{ij}}{N_j} \int_0^\infty I_j(t), \\
\log \frac{S_i(0)}{S_i(\infty)} &= \beta_i \sum_{j=1}^{3} \frac{P_{ij}}{N_j} \int_0^\infty I_j(t).
\end{aligned}$$

Expanding this equation we have

$$\log \frac{S_i(0)}{S_i(\infty)} = \beta_i \left[\frac{P_{i1}}{N_1} \int_0^\infty I_1(t) + \frac{P_{i2}}{N_2} \int_0^\infty I_2(t) + \frac{P_{i3}}{N_3} \int_0^\infty I_3(t) \right], \tag{3}$$

substituting relation from equation (2) into the right hand side of (3) we get,

$$\begin{aligned}
\log \frac{S_i(0)}{S_i(\infty)} &= \beta_i \left[\frac{P_{i1}}{N_1} \left(\frac{N_1 - S_1(\infty)}{\gamma} \right) + \frac{P_{i2}}{N_2} \left(\frac{N_2 - S_2(\infty)}{\gamma} \right) + \frac{P_{i3}}{N_3} \left(\frac{N_3 - S_3(\infty)}{\gamma} \right) \right], \\
&= \frac{\beta_i}{\gamma} \left[P_{i1} \left(1 - \frac{S_1(\infty)}{N_1} \right) + P_{i2} \left(1 - \frac{S_2(\infty)}{N_2} \right) + P_{i3} \left(1 - \frac{S_3(\infty)}{N_3} \right) \right],
\end{aligned}$$

hence the final size relation is a system of three equations in three unknowns $S_1(\infty), S_2(\infty)$ and $S_3(\infty)$ given by:

$$\begin{aligned}
\log \frac{S_1(0)}{S_1(\infty)} &= \frac{\beta_1}{\gamma} \left[P_{11} \left(1 - \frac{S_1(\infty)}{N_1} \right) + P_{12} \left(1 - \frac{S_2(\infty)}{N_2} \right) + P_{13} \left(1 - \frac{S_3(\infty)}{N_3} \right) \right], \\
\log \frac{S_2(0)}{S_2(\infty)} &= \frac{\beta_2}{\gamma} \left[P_{21} \left(1 - \frac{S_1(\infty)}{N_1} \right) + P_{22} \left(1 - \frac{S_2(\infty)}{N_2} \right) + P_{23} \left(1 - \frac{S_3(\infty)}{N_3} \right) \right], \\
\log \frac{S_3(0)}{S_3(\infty)} &= \frac{\beta_3}{\gamma} \left[P_{31} \left(1 - \frac{S_1(\infty)}{N_1} \right) + P_{32} \left(1 - \frac{S_2(\infty)}{N_2} \right) + P_{33} \left(1 - \frac{S_3(\infty)}{N_3} \right) \right].
\end{aligned}$$

Note however that this is a system of non-linear transcendental equations and one must use non-linear equation solvers to obtain the solution. Once we obtain $S_1(\infty), S_2(\infty)$ and $S_3(\infty)$, we can then find the number of disease cases from,

$$\sum_{i=1}^{3} [N_i - S_i(\infty)]$$

When we consider the situation in which the visitors are not accounted for, and there are only the natives and male sex workers, we will be able to look specifically at $S_2(\infty)$ and $S_3(\infty)$.

4 Results

4.1 R_0 Value

The basic reproductive number can be calculated by selecting specific values of individual preference for each subpopulation. We assumed that in a real life scenario, the preference of male sex workers for male sex workers is very low due to the lack of gain from the experience. Also, the preference of natives for their own group is bigger than the preference of visitors for their own group. We let $f_1 = 0.4$, $f_2 = 0.6$ and $f_3 = 0.01$ and use our estimated sexual contact values for each subpopulation to obtain the basic reproductive number of our system, 4.7886. This result is similar to 4.4, the basic reproductive number calculated for the Zika outbreak in Barranquilla, Colombia by Dr. Towers and colleagues [25].

4.2 Special Cases with the SIR model

We ran a simulation where we evaluated the final epidemic size for our model with various f_i (preferences) and c_i (contacts). Table 4 shows the different cases for which the number of infected individuals was calculated at 30 days, that is the approximate time that people could spend in Rio de Janeiro for the Olympic Games. We also calculated the final epidemic size (FES) for the entire length of the outbreak. The first three cases are the special ones when $f_i = 0$, $f_i = 0.5$ and $f_i = 1$. For cases 4, 5 and 6, the preference of individuals within their own group was fixed and the

number of contacts was varied and for cases 7, 8 and 9 the contacts were fixed and the preferences varied.

Table 4: Number of infected individuals and final epidemic size (FES) for different cases of preferences and contacts

Case	Preferences (f_1, f_2, f_3)	Contacts (c_1, c_2, c_3)	Number of infected individuals at 30 days	FES
Case 1	0, 0, 0	$\frac{6.8}{28}, \frac{3.4}{28}, \frac{11}{7}$	279.8	94,330.87
Case 2	0.5, 0.5, 0.5	$\frac{6.8}{28}, \frac{3.4}{28}, \frac{11}{7}$	294.7	83,038.1456
Case 3	1, 1, 1	$\frac{6.8}{28}, \frac{3.4}{28}, \frac{11}{7}$	331.6	13,083.374
Case 4	0.4, 0.7, 0.01	1, 0.5, 10	15,633.2676	172,591.921
Case 5	0.4, 0.7, 0.01	0.5, 0.33, 10	14,273.6013	162,493.0755
Case 6	0.4, 0.7, 0.01	0.5, 0.33, 6	4,241.2	162,394.5997
Case 7	0.3, 0.6, 0.01	$\frac{6.8}{28}, \frac{3.4}{28}, \frac{11}{7}$	271.5	86,860.7888
Case 8	0.5, 0.8, 0.01	$\frac{6.8}{28}, \frac{3.4}{28}, \frac{11}{7}$	271.1	79,372.5979
Case 9	0.5, 0.8, 0.05	$\frac{6.8}{28}, \frac{3.4}{28}, \frac{11}{7}$	272.6	79,163.911

4.3 Peak Time with the SIR Model

Table 5 shows the peak times for different scenarios of contacts and preferences between individuals. The result of the simulation produced a peak time value for each subpopulation in each different case. The peak time is the day at which the number of infected individuals in each sub-population reaches the maximum value. Each one of the nine cases of Table 5 has the same preferences and contacts values as Table 4.

Table 5: Peak Time for different cases of preferences and contacts

Case	Peak time for I_1 (days)	Peak time for I_2 (days)	Peak time for I_3 (days)
Case 1	487	523	331
Case 2	447	462	181
Case 3	0	1	122
Case 4	109	187	37
Case 5	147	229	32
Case 6	176	261	57
Case 7	414	492	258
Case 8	385	487	216
Case 9	385	487	213

The goal was to predict the number of male visitors, natives and sex workers that are going to be infected at the end of the Olympics. Since the Final Epidemic Size (FES) is obtained after the Olympic Games, we calculated the number of infected individuals at 30 days and the actual FES. The FES was found to be at approximately at 4,000 days, when the disease dies out. Table 4 shows the three special cases that were considered when fixing the number of contacts per day and varying the preference of individuals within their own group from 0, 0.5 and 1. The results showed that when f_i is the same for each subpopulation and increases, that the number of infected individuals increased and the FES decreased. Notice that when the $f_i = 1$, there are no infected male visitors which means that the FES is zero. Table 5 shows the peak time for each infected subpopulation in each case. For the first three cases, when f_i is increased the peak time for each infected subpopulation decreased; since individuals are mixing more within their own group, the dynamic of the epidemic is going to be faster and the peak time is obtained faster. When $f_i = 1$ visitors are having contacts only within their own group they have no peak time due to no visitors being initially infected.

For the cases 5 and 6, when the preference of each group is fixed ($f_1 = 0.4$, $f_2 = 0.7$ and $f_3 = 0.01$) and the contacts of male visitors and natives are the same, but the contacts of male sex workers

decrease from 10 to 6. Also, the number of infected individuals at 30 days decreased and the FES does not changed considerably. This is expected since the male sex workers are having less sexual contacts and one assumption was that a percentage of that subpopulation was initially infected. For case 7 and 8 the number of contacts were fixed and the preference of male natives and visitors within their own group was increased so the number of infected individuals at 30 days does not changed considerably, but the FES decreased.

The following figures show the comparison between the random mixing and the preferred mixing for infected individuals from our three subpopulations. In the random mixing, individuals do not make contact within their own group and in the preferred mixing individuals make contact within their own and other group. Figure 2 A) shows the infected male visitors and natives for $f_i = 0$, 2 B) shows the analogous for $f_i = 0.5$ and 2 C) shows for $f_i = 1$. From the graphics we are able to determine that when the preference value increases, the maximum number of infected individuals decreases. Therefore preferred mixing is better for male natives and visitors. When the preference value is 1, visitors are only engaging in sexual contact with themselves and therefore are not infected. Figure 3 A) shows the infected male sex workers for $f_i = 0$, 3 B) shows the analogous for $f_i = 0.5$ and 3 C) shows for $f_i = 1$. The result is that when the preference value increases, the maximum number of infected individuals increases. Therefore, random mixing is favored for the male sex workers.

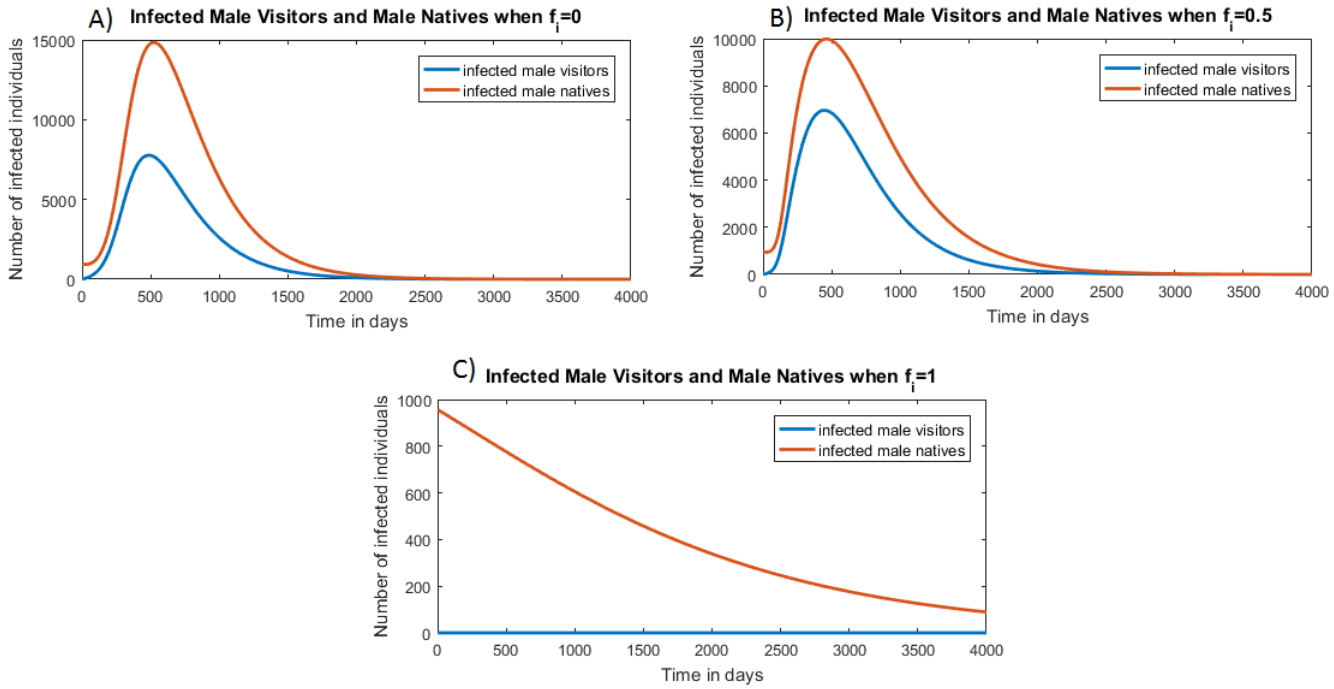


Figure 2: Simulation for Infected Male Visitors and Natives

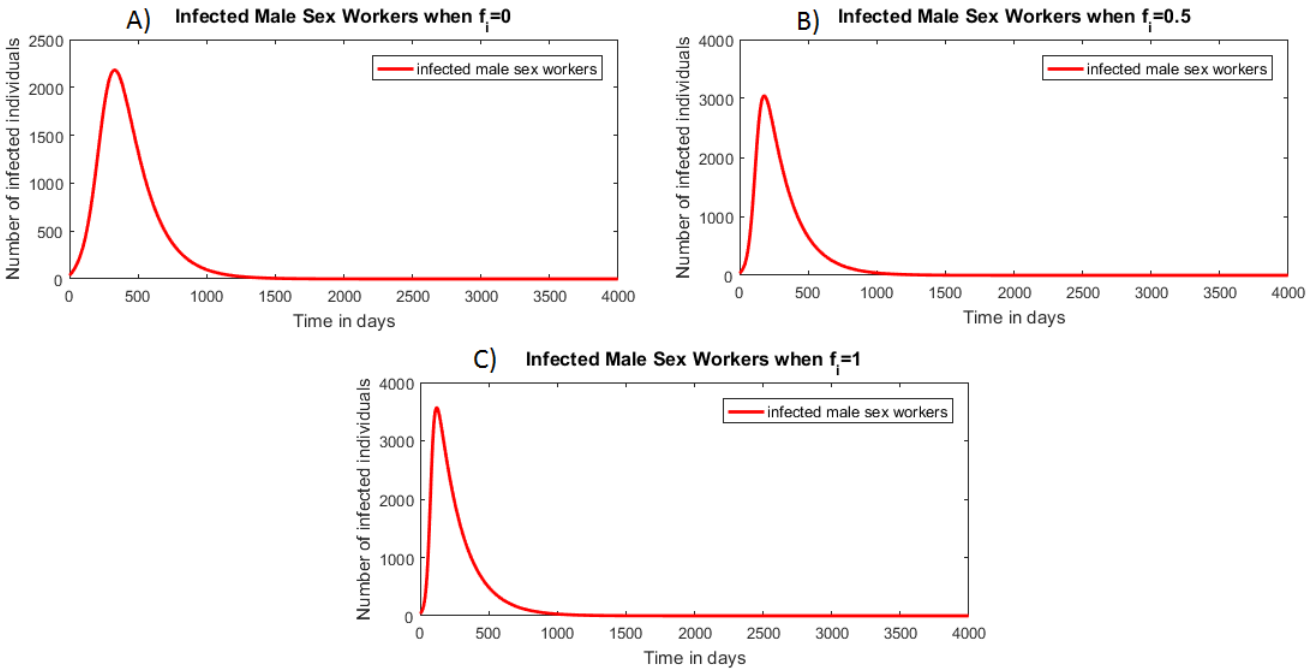


Figure 3: Simulation for Infected Male Sex Workers

The following graphics show the relationship between the peak time and the preference value f_i. Figure 4 shows the peak time for the infected male visitors at different values of f_i. When f_i is the same for all subpopulations and is increasing, the peak time for the infected male visitors decreases, but at a certain value it starts increasing again. Then, for that particular value the system is not monotonic. Figure 5 shows the peak time for the infected male natives at different values of f_i. When f_i is the same for the three subpopulations, it seems that the peak time decreases while f_i increases. Figure 6 shows the peak time for the infected male sex workers and the result is that the peak time decreases while f_i increases.

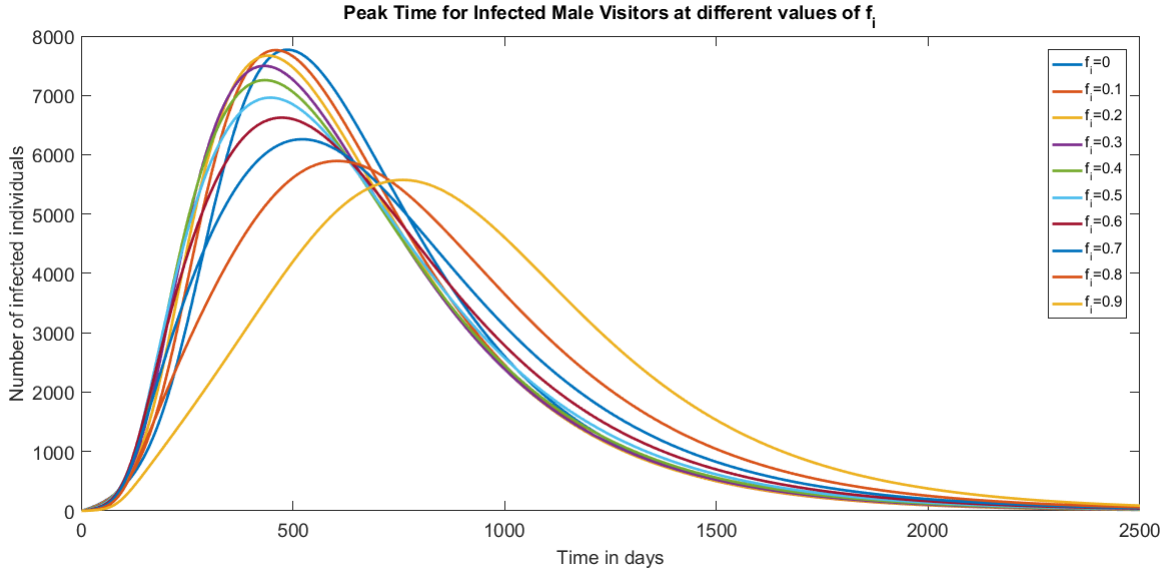


Figure 4: Peak Time for Infected Male Visitors at different values of f_i

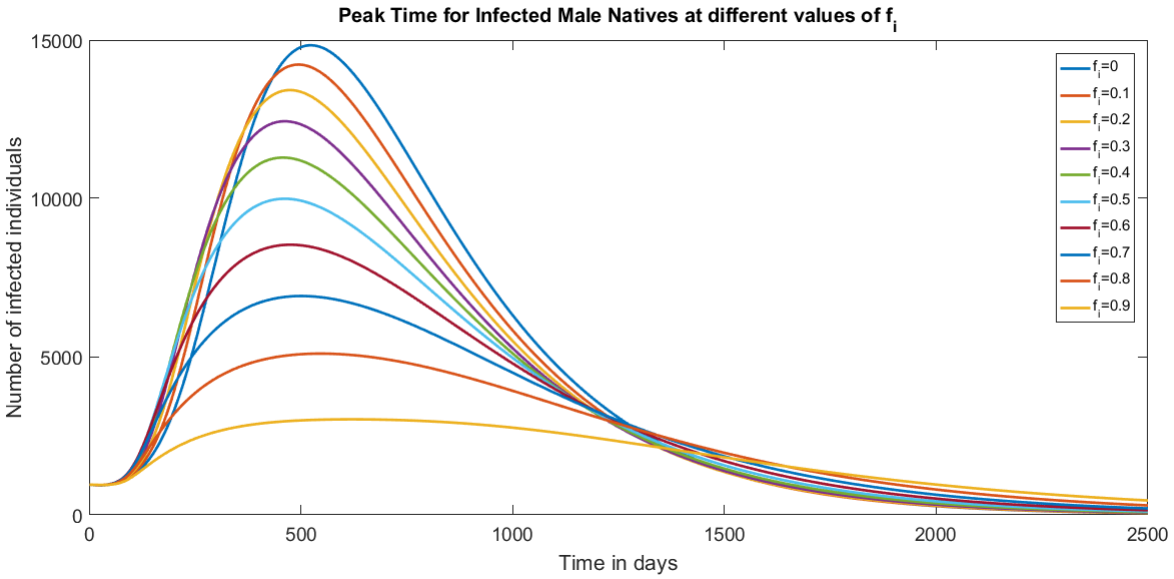


Figure 5: Peak Time for Infected Male Natives at different values of f_i

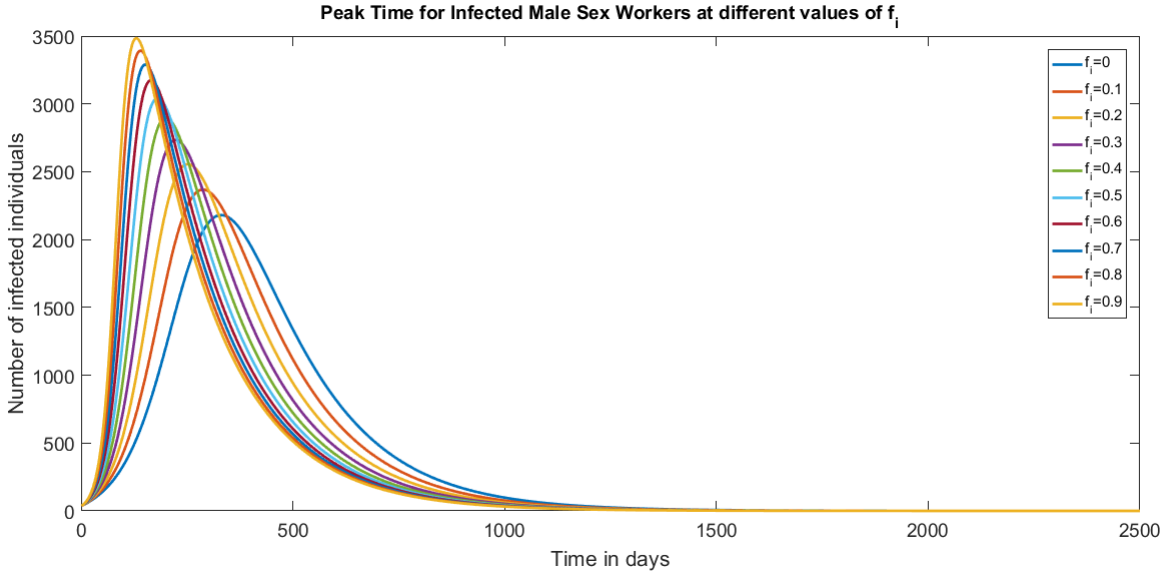


Figure 6: Peak Time for Infected Male Sex Workers at different values of f_i

4.4 Uncertainty Analysis

Uncertainty in the chosen parameter values introduces variability to the model's results. Since there is parameter uncertainty, there is variability in the results. In particular, the more uncertain parameters are, the more significant the variability introduced [26]. Being doubtful of the values of parameters associated with these processes creates model parameter uncertainty [26]. Thus, a sensitivity analysis is often performed to assess this variability in the model prediction [26].

Uncertainties can affect the model output in each time step of the simulation, so it is important to analyze how the parameters affect the model [26]. However, adjusting the model to find solutions for different data sets involves the use of different parameter sets [26]. Since uncertainty involves the idea of randomness, the alteration of parameter values over space and time cannot be predicted with certainty [26]. Thus, we use random variables throughout the uncertainty analysis to determine how the uncertainty in our parameters and variables affects the outcomes of the model [26].

In order do this, it is important to determine the probability that a parameter will be within some specified range of values which is defined by a probability distribution [10]. For this project Latin Hyper Cube Sampling (LHS) will be used to generate values for our parameters corresponding

to preference (f_i) and contacts per day (c_i). The LHS procedure divides the range of values for a given parameter into equally probable intervals, thus the LHS is a stratified form of random sampling. Probability distributions are assigned to parameters, the intervals in the distribution are divided into equi-probable regions, and these intervals are then each sampled without replacement [10]. LHS compared to simple random sampling schemes requires fewer samples to achieve the same level of accuracy [10].

A complete uncertainty analysis involves a comprehensive identification of all parameters of uncertainty that contribute to the joint probability distributions of each input or output variable [26]. R_0 and final size were considered to see how the parameters affect the possible outcomes. Using the LHS method under uniform distribution, three experiments were conducted to analyze the uncertainty in our ODE systems. For both R_0 and final size, the experiments correspond to the same sampling. For the value ranges of f_i and c_i refer to Table **??**. Experiment 1 samples from varying f_i in the specified range in with constant c_i's from LHS. Experiment 2 takes varying values of c_i from the specified range with constant f_i values. Lastly, Experiment 3 uses the same ranges for c_i and f_i in Experiments 1 and 2, meaning that six parameters were varied for each simulation (f_1, f_2, f_3, c_1, c_2, and c_3).

Experiment 1 with Varying f_i	
Constants	Values
N_1	40776.8
N_2	130717.8
N_3	5000.0
c_1	$\frac{6.8}{28}$
c_2	$\frac{3.4}{28}$
c_3	$\frac{11}{7}$

Table 6: R_0 Experiment 1 Constants for Uncertainty Analysis on f_i.

Experiment 2 with Varying f_i	
Constants	Values
N_1	40776.8
N_2	130717.8
N_3	5000.0
f_1	0.0412
f_2	0.8967
f_3	0.0263

Table 7: R_0 Experiment 2 Constants for Uncertainty Analysis on c_i.

Experiment 3 with Varying f_i	
Constants	Values
N_1	40776.8
N_2	130717.8
N_3	5000.0

Table 8: R_0 Experiment 3 Constants for Uncertainty Analysis on c_i and f_i.

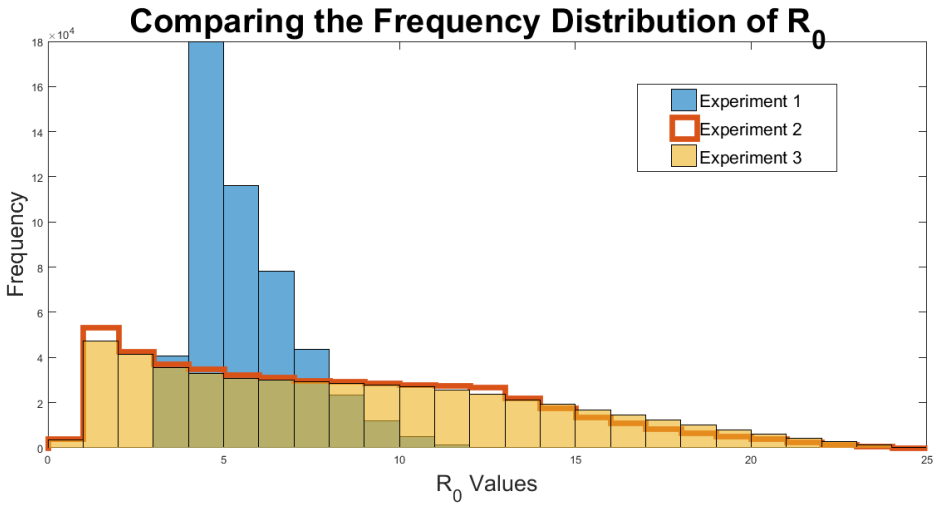


Figure 7: Compares the overall distribution of values for R_0 produced by the experiments.

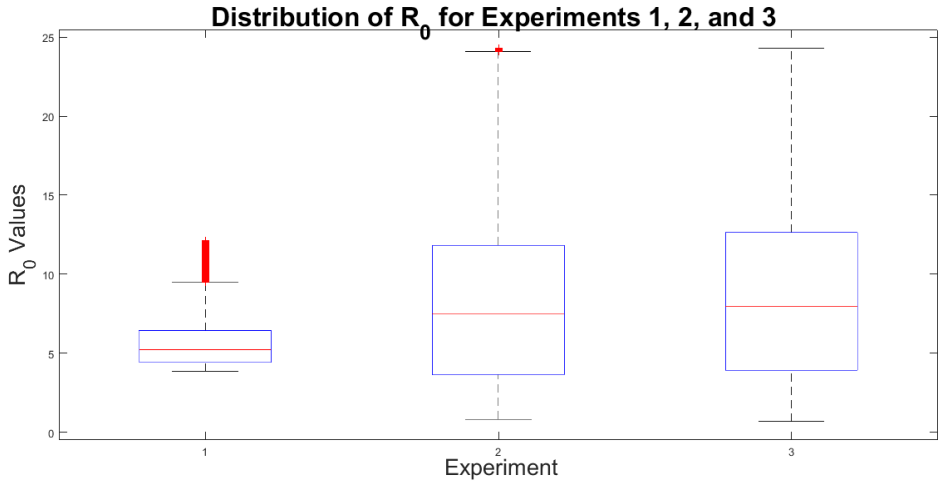


Figure 8: Compares the frequency and distribution of values for for R_0 produced by the experiments.

Measures of central tendency for R_0				
Experiment	**Mode**	**Median**	**Mean**	**Variance**
Varying f_i (1)	3.8384	5.2210	5.6053	2.2348
Varying c_i (2)	0.8147	7.4661	8.1204	26.2594
Varying c_i & f_i (3)	0.6804	7.9714	8.6862	29.8724

Table 9: Values for Data Measurement and Statistics of the Distribution of R_0.

For the experiments on R_0 with the initial assumptions on population size and values for f_i and c_i, the R_0 values have a smaller width of distribution than the R_0 values in Experiment 1 where the uncertainty of f_i is analyzed. Also, Figure 7 on page 25 shows that in Experiment 1, R_0 is less sensitive to changes in f_i. However, Figure 8 on page 26 also shows that changes in f_i can result in R_0 greater than 1, implying that the 2016 Olympics can undergo a widespread occurrence of ZIKV under these initial assumptions.

In Figure 7, the distribution of R_0 is still smaller than the distribution of values for R_0 in the other experiments. The R_0 range goes from 3.5 to 12.5 for Experiment 1 in Figure 8 which reasonable according to Towers et al. the estimated R_0 value is 4.4 [25].

For Experiments 2 and 3, Figure 8 shows that R_0 has a wider distribution of values as c_i varies

uniformly. With f_i constant, this shows that controlling the number of contacts per day may reduce R_0 more efficiently than controls on the preference parameters relating to f_i.

The same experiments were applied to the uncertainty analysis of $S_i(30)$ and $S(30)$ (total Olympics population), which are calculated as the total number of infections during the duration of the Olympics between $t = 0$ and $t = 30$ ($S(0)$ - $S(30)$)

Experiment 1 with Varying f_i	
Constants	Values
N_1	40776.8
N_2	130717.8
N_3	5000.0
c_1	$\frac{6.8}{28}$
c_2	$\frac{3.4}{28}$
c_3	$\frac{11}{7}$

Table 10: $S_i(30)$ Experiment 1 Constants for Uncertainty Analysis on f_i.

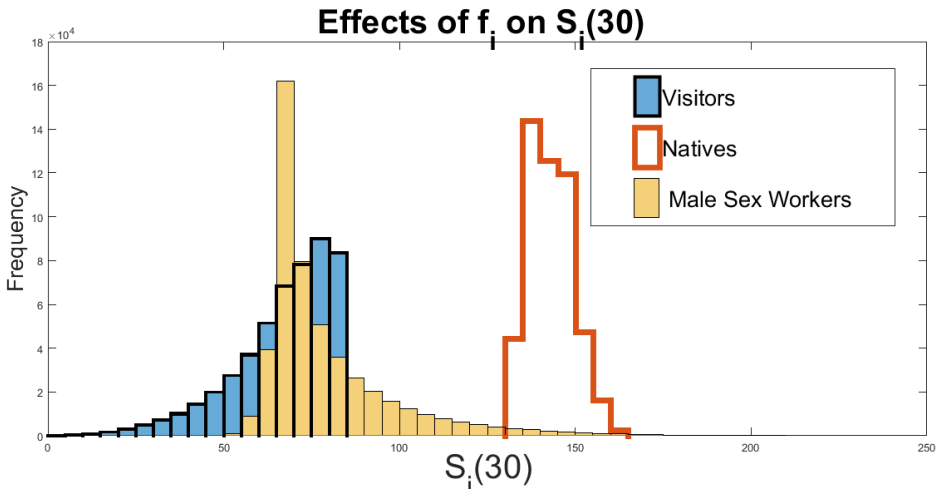


Figure 9: Histogram comparing the frequency distribution of $S(30)$ for visitors, natives, and male sex workers with varying f_i and constant values from table .

Changes in f_i creates skewed data points for the visitors and male sex workers supopulation. Figure 9 on page 27 shows how the distributions peak is off center toward the limit with a tail that

stretches away from it for visitors and male sex workers. The visitor subpopulation is left skewed while the male sex worker subpopulation is right skewed. The natives have the smallest range for its distribution of $S_2(30)$ values, hence f_i does not affect $S_2(30)$ as much as $S_1(30)$ and $S_3(30)$.

For Experiment 2, we used the following values as constants while c_i varied uniformly.

Experiment 2 with Varying f_i	
Constants	Values
N_1	40776.8
N_2	130717.8
N_3	5000.0
f_1	0.1608
f_2	0.3779
f_3	0.0145

Table 11: $S_i(30)$ Experiment 2 Constants for Uncertainty Analysis on c_i.

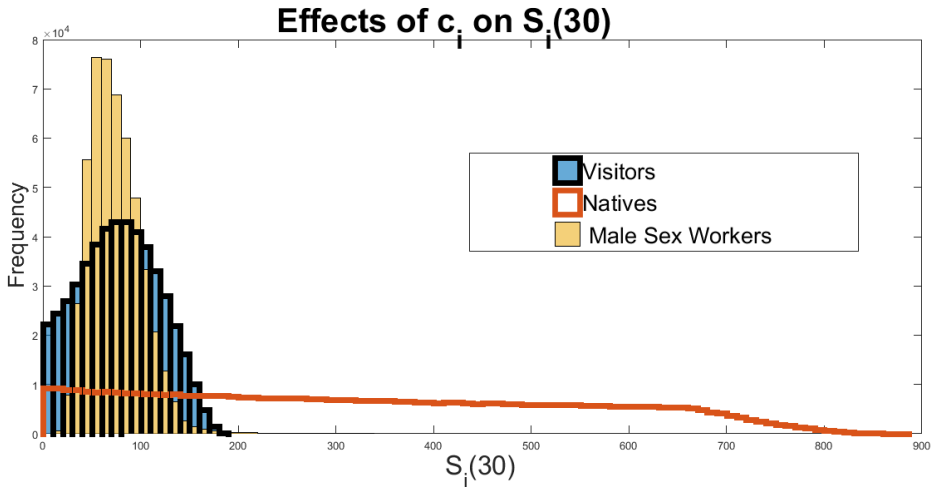


Figure 10: Histogram comparing the frequency distribution of $S(30)$ for visitors, natives, and male sex workers with varying c_i.

Figure 10 on page 28 shows that variations in c_i has the greatest affect on the native's subpopulation at the 2016 Olympics because the range of the distribution for $S_2(30)$ is the greatest for Experiment 2. The male sex worker subpopulation is least affected by the c_i parameter (Figure 10).

Also, for Experiment 2, the range for the distribution of values in all subpopulations increased compared the the range for the distribution of values in Experiment 1 (see Figure 9 and Figure 10)

Experiment 3 with Varying f_i	
Constants	Values
N_1	40776.8
N_2	130717.8
N_3	5000.0

Table 12: $S_i(30)$ Experiment 3 Constants for Uncertainty Analysis on c_i and f_i.

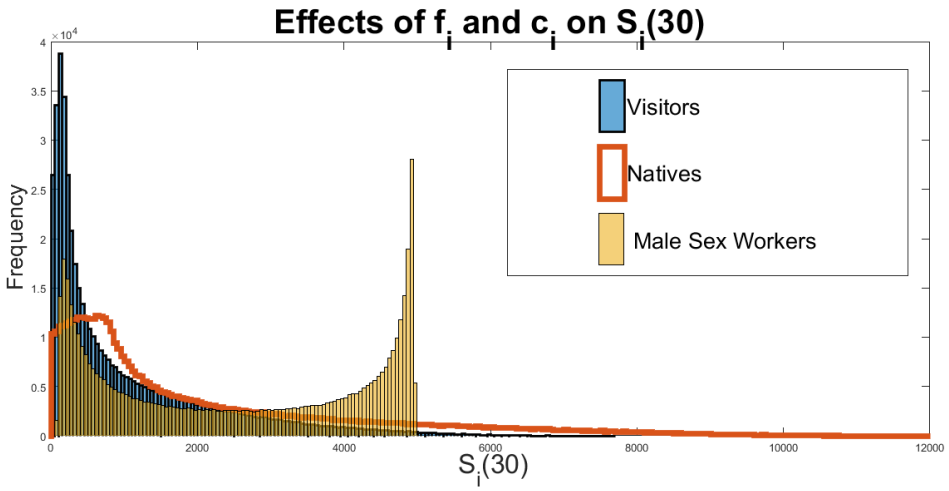


Figure 11: Histogram comparing the frequency distribution of $S(30)$ for visitors, natives, and male sex workers with varying c_i and f_i with constant values from table.

As c_i and f_i both vary, $S_3(30)$ becomes bimodal while $S_1(30)$ and $S_2(30)$ are right skewed (see Figure 11 on page 29. This suggests that the male sex worker subpopulation has two local maximums where the data points stop increasing and start decreasing. These affects can be contributed to the c_i parameter because of how much changes in c_i affected each subpopulation compared to changes in f_i.

Relative Maxima and Minima $S_i(30)$		
Experiment 1	**Range**	**Variance**
$S_1(30)$	83.6731	197.4159
$S_2(30)$	30.4108	37.9532
$S_3(30)$	152.5820	360.7195
Experiment 2	**Min.**	**Variance**
$S_1(30)$	186.2558	1.6353e+03
$S_2(30)$	885.7896	4.6594e+04
$S_3(30)$	322.2019	715.6933
Experiment 3	**Min.**	**Variance**
$S_1(30)$	7.6877e+03	1.4267e+06
$S_2(30)$	1.1994e+04	4.4181e+06
$S_3(30)$	4.9457e+03	3.2140e+06

Table 13: Compares the ranges for $S_i(30)$ between the experiments and the subpopulations. This summarizes how the uncertainty in parameters affects the ranges of values for $S_i(30)$.

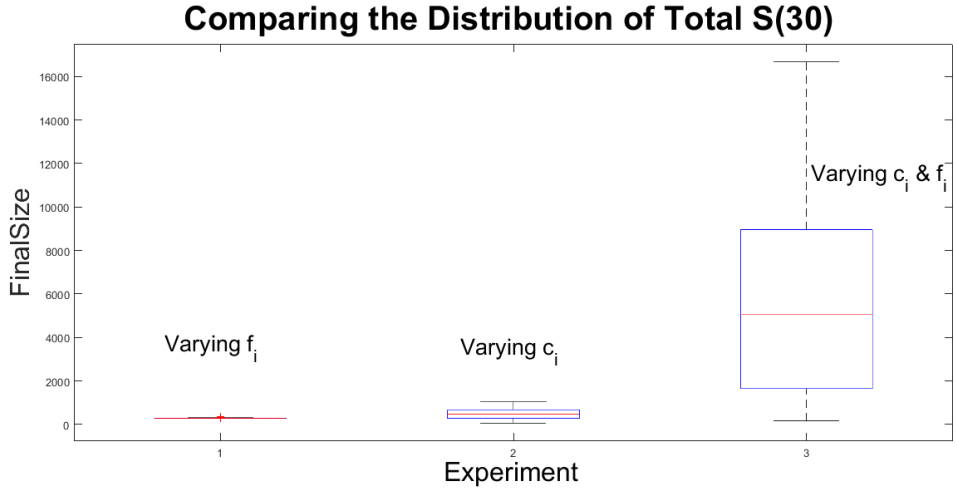


Figure 12: Boxplot comparing the overall distribution of values for $S(30)$, total final size after the Olympics, with the different experiments.

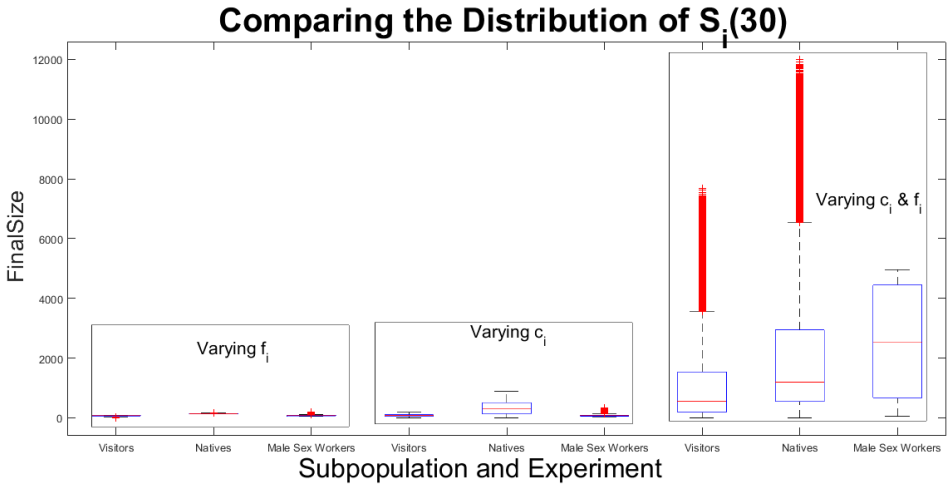


Figure 13: Boxplot comparing the distribution of values for $S_(30)$ between the different subpopulations produced by experiment 1 (left), Experiment 2 (center), and Experiment 3 (right).

Figure 13 on page 31 shows how the range of distributions changes between experiments and subpopulations as the parameter values vary. Recall that Experiment 1 corresponds to affects of variations in the f_i parameter; Experiment 2 corresponds to affects of variations in the c_i parameter; and Experiment 3 corresponds to affects of variations in the c_i and f_i parameters.

Measures of central tendency for $S(30)$				
Experiment	**Mode**	**Median**	**Mean**	**Variance**
1	267.0955	285.5336	288.9192	111.5656
3	62.1114	469.3810	487.4168	5.3551e+04
3	174.4761	5.0424e+03	5.6545e+03	1.7585e+07

Table 14: Values for Data Measurement and Statistics of the Distribution of $S(30)$ under varying f_i, c_i, and c_i and f_i

Figure 13 show that changes in the initial assumptions produce different outcomes for the ranges of values in the final size. The widest distribution of values when varying parameters is produced in Experiment 3 (see Figure 11). However, since the subpopulations generally Figure 9 have a smaller range than Figure 10, the resulting ranges in Figure 11 could be due to the variations in c_i as it is in Figure 10. Also, the total final size, using the estimated values from literature, can

get as large as 160,000, a significantly larger number than all of the possible total final size values produced by the initial assumptions.

Overall, the preference parameter, f_i, does not reduce the final size for a population, as the way the c_i parameter can for final size distribution. Also, the number of contacts per day, c_i has a greater range of uncertainty. Hence, there exists a combination of c_i's that contributes to reduce the total final size nearing 0. For Experiment 3 exhibits the largest range of uncertainty which is a reasonable result due to the fact that 6 parameters (f_1, f_2, f_3, c_1, c_2, and c_3) are being varied uniformly to produce the final sizes.

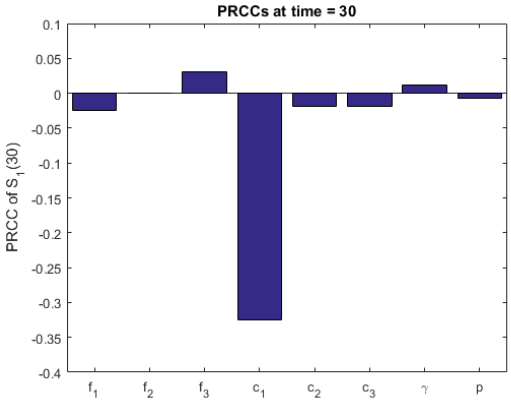
4.5 Sensitivity Analysis

We conducted LHS/PRCC (Latin Hypercube Sampling/Partial Rank Correlation Coeffiecient) Sensitivity Analysis. The purpose of LHS/PRCC analysis is to understand how the uncertainties of the parameters used in a model effect predictions and distributions, and to rank the parameters by impact onto those predictions [10]. As was aforementioned, the LHS was used to generate values for the parameters of preference (f_1, f_2, f_3) and contact (c_1,c_2,c_3). Here, we are considering the PRCC in order to create a single histogram in which the parameters that affect the final sizes of each of our three populations: visitors, natives, and male sex workers.

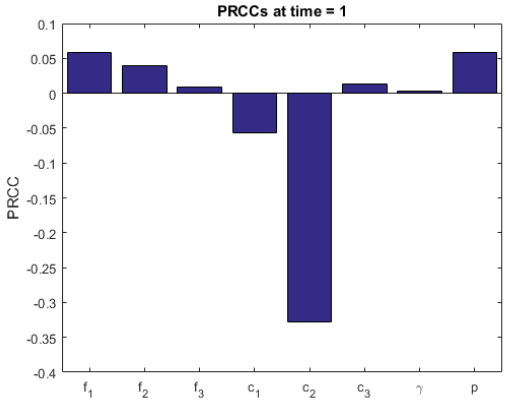
The figures above each show which parameters are the most sensitive to the final size of each population. In Figure **??**, it is shown that the distributions for each of the populations are uniform. This result comes out of the codes that were used to create the charts of PRCC outputs for each final size. In Figure 14a, we can see that the final size of the visitor population is highly impacted by the number of contacts that happen between the visitors and the visitors, natives, and male sex workers, c_1.

Similarly, we can see the correlation between c_2, the number of sexual contacts of the natives to the visitors, natives, and male sex workers, and the final size of the native population. Here we can understand why in Figure 14b these parameters are the most sensitive. We can also notice that these scatter plot correlations are random, since there is a uniform distribution.

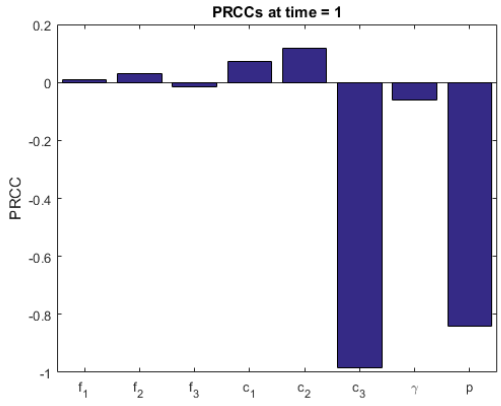
Lastly, we can see the correlation between c_3, which is the number of sexual contacts from the



(a) This shows the PRCC output in bar chart form of the parameters against $S_1(\infty)$.



(b) This shows the PRCC output in bar chart form of the parameters against $S_2(\infty)$.



(c) This shows the PRCC output in bar chart form of the parameters against $S_3(\infty)$.

male sex workers to the visitors, natives, and male sex workers and $S_3(30)$.It can also be noticed that f_3 is negligible since the chance that male sex workers would be interacting with prostitutes in this event period is very small. Here we can see in Figure 14c that only one parameter c_3 is sensitive to the final size of the male sex worker population.

To conclude the description of this analysis, we can see, from Figure 14a,Figure 14b,Figure 14c that the most sensitive parameters are c_1, c_2, and c_3. Therefore, we can say that the number of contacts that each individual makes in these sub populations will have the highest impact upon the number of infected individuals at the end of the 30-day Olympic event period. We can also notice that the histograms showed negative PRCC values, and this is because we would want to

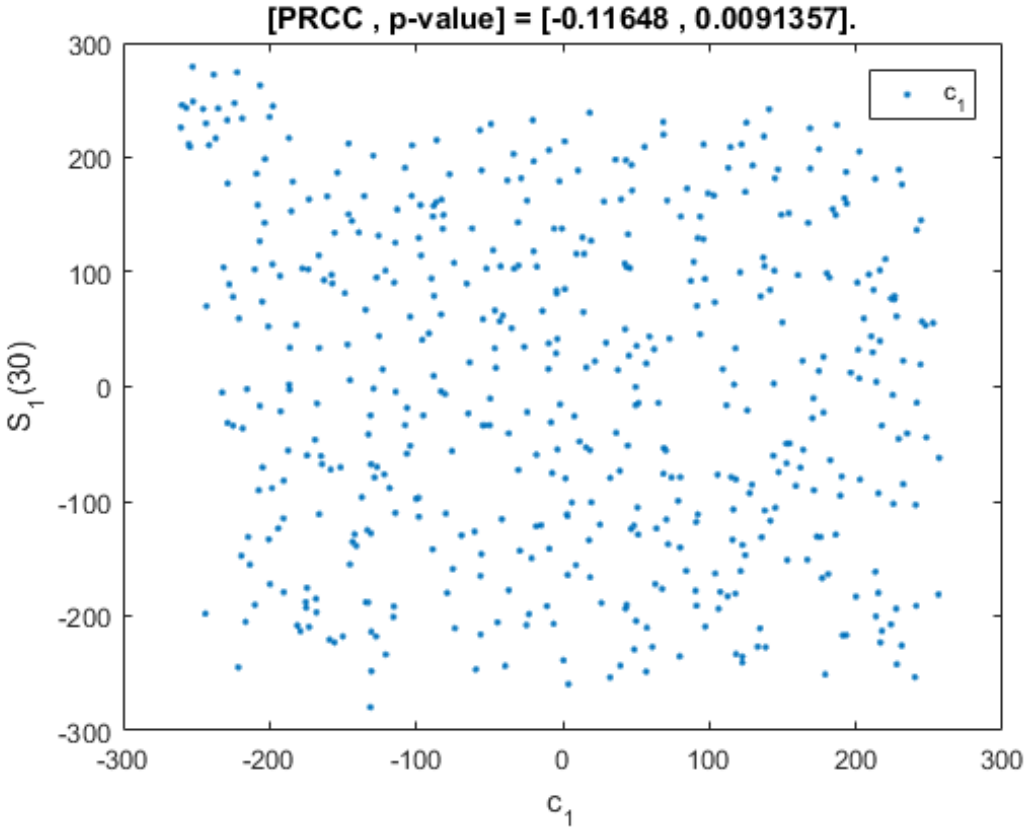


Figure 15: This shows the scatterplot correlation between c_1 and $S_1(30)$.

decrease the number of contacts in each sub population in decrease the number of infected visitors, natives, and male sex workers.

5 Conclusion and Discussion

We ran simulations where we observed special cases with $f_i = 0, 0.5$ and 1. It was determined that the final epidemic size and peak time decrease as time goes to infinity. For 30 days, it was determined that when the preference of individuals within their own group increases, that the number of infected individuals increase as well. We concluded that preferred mixing is better than random mixing because the number of infected male visitors and natives is less overall. Male sex workers have the lowest subpopulation, therefore we did not consider the number of infected

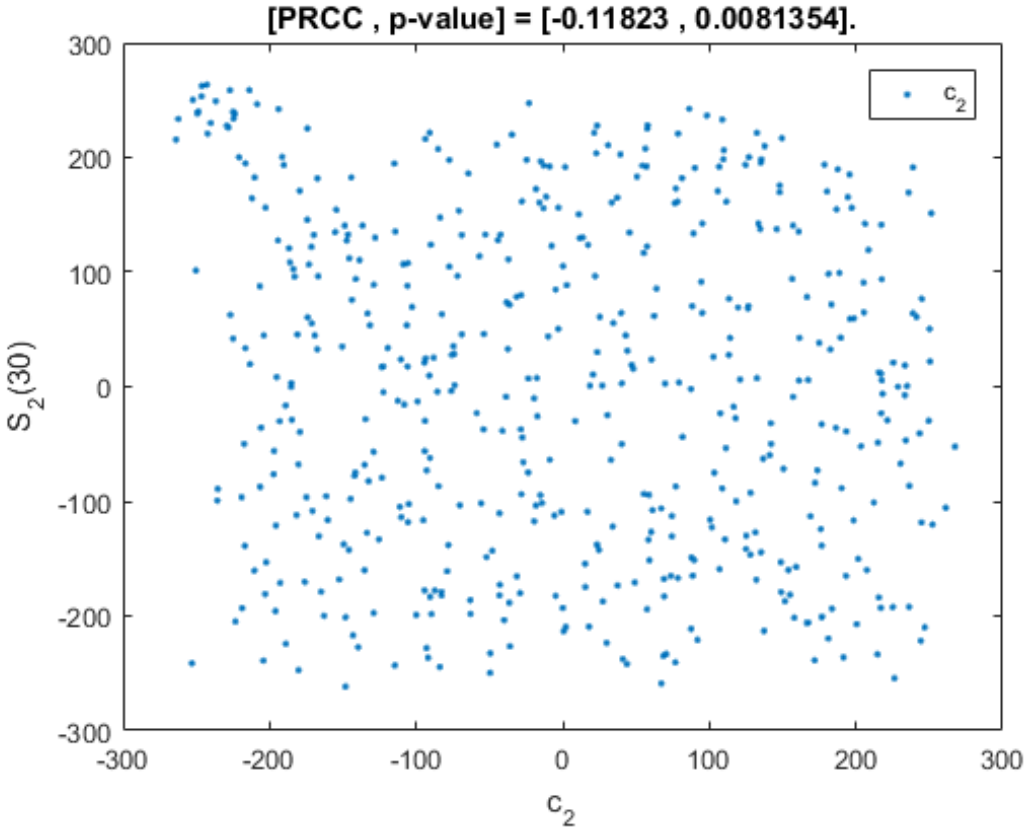


Figure 16: This shows the scatter plot correlation between f_2 and $S_2(\infty)$.

male sex workers since they are a small amount of the overall population. For specific values of preference, we calculated a basic reproductive number of 4.7886 which is similar to 4.4, the basic reproductive number calculated by Dr. Towers and colleagues for the Zika outbreak in Barranquilla, Colombia [25].

The final PRCC charts show that the most sensitive parameters are c_1, c_2, c_3. This indicates that the most important parameters to consider when trying to decrease the infected individuals of each subpopulation is the number of sexual contacts that each individual is involved in. Overall, the preference parameter, f_i, does not reduce the final size for a population the way the c_i parameter can for final size distribution. Also, the number of contacts per day parameter, c_i has a greater range of uncertainty. So, there exists a combination of c_i's that contributes to total a final size

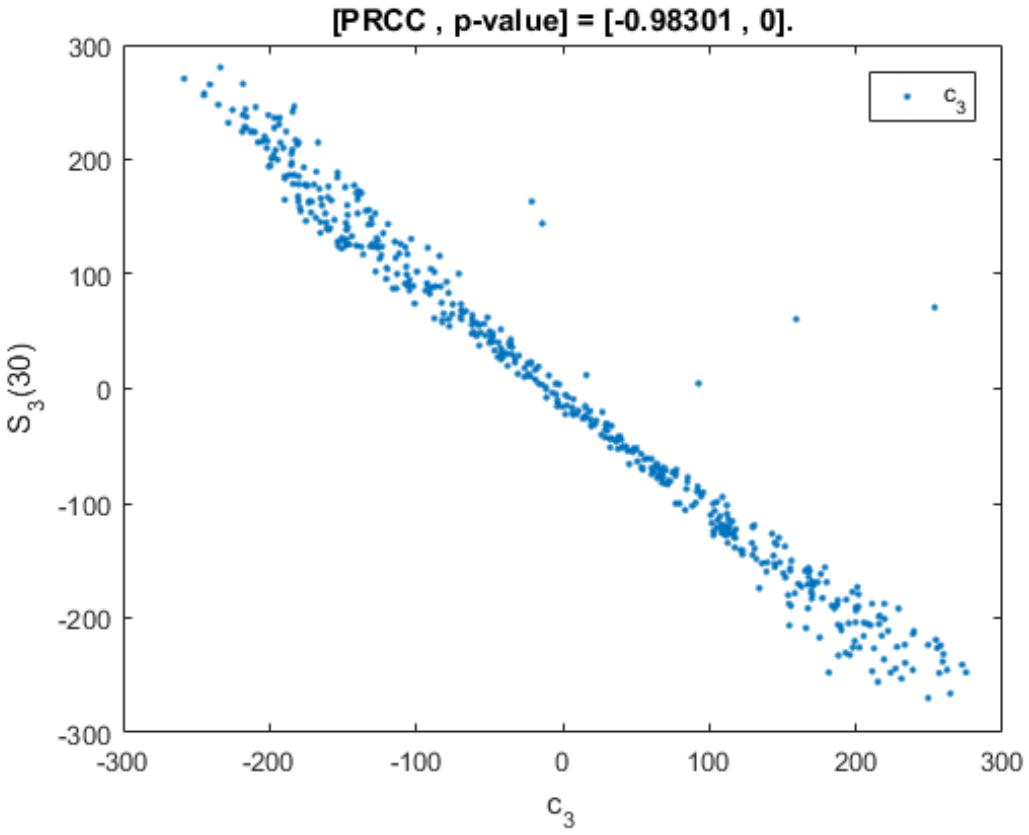


Figure 17: This shoes the scatterplot correlation between f_3 and $S_3(\infty)$.

nearing 0 for the entire population at the 2016 Olympics. For Experiment 3 exhibits the largest range of uncertainty which is a reasonable result due to the fact that 6 parameters (f_1, f_2, f_3, c_1, c_2, and c_3) are being varied uniformly to produce the final sizes.

6 Acknowledgments

We would like to thank the Mathematical and Theoretical Biology Institute (MTBI) co-Directors Dr. Carlos Castillo-Chavez, Dr. Anuj Mubayi, and Dr. Marlio Paredes for giving us the opportunity to participate in this research program. We would also like to thank Associate Director Sherry Woodley and Coordinator Ciera Duran for their efforts in planning and executing the day to day activities of MTBI. We also want to give special thanks to Baojun Song, Baltazar Espinoza,

and Leon Arriola. The research has been carried out at MTBI which is a Research Experience for Undergraduate (REU) summer program at the Simon A. Levin Mathematical, Computational and Modeling Sciences Center (SAL MCMSC) at Arizona State University (ASU). This project has been partially supported by grants from the National Science Foundation (DMS1263374), the Office of the President of ASU, and the Office of the Provost at ASU.

References

[1] A. Attaran, *Off the podium: Why public health concerns for global spread of zika virus means that rio de janeiro's 2016 olympic games must not proceed.* http://harvardpublichealthreview.org/off-the-podium-why-rios-2016-olympic-games-must-not-proceed/, July 10, 2016.

[2] J. Billy, *The sexual behavior of men in the united states*, Family planning perspectives, (1993), pp. 52–60.

[3] F. Brauer, *Heterogeneity in disease transmission style*, Bulletin of Mathematical Biology, (2008), pp. 70:1869–1885.

[4] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology.*, vol. Second Edition, Springer, 2010.

[5] C. Castillo-Chavez, B. Song, and J. Zhangi, *An epidemic model with virtual mass transportation: The case of smallpox*, Bioterrorism: Mathematical Modeling Applications in Homeland Security, 28 (2003), p. 173.

[6] L. Catron, A. L. Forgia, D. Padilla, R. Castro, K. Rios-Soto, and B. Song, *Immigration laws and immigrant health: Modeling the spread of tuberculosis in arizona.*, (2010).

[7] A. S. Chacham, S. G. Diniz, M. B. Maia, A. F. Galati, and L. A. Mirim, *Sexual and reproductive health needs of sex workers: two feminist projects in brazil*, Reproductive Health Matters, 15 (2007), pp. 108–118.

[8] C. Codeço, D. Villela, M. Gomes, L. Bastos, O. Cruz, and F. C. Claudio Struchiner, Luis Max Carvalho, *Zika is not a reason for missing the olympic games in rio de janeiro*, Mem. Inst. Oswaldo Cruz, 111 (2016), pp. 414–415.

[9] D. Gao, Y. Kuang, Y. Lou, G. Chowell, D. He, T. C. Porco, and S. Ruan", *Prevention and control of zika fever as a mosquito-borne and sexually transmitted disease*, Scientific Reports, (2016).

[10] B. Gomero, *Latin hypercube sampling and partial rank correlation coefficient analysis applied to an optimal control problem.* http://trace.tennessee.edu/cgi/viewcontent.cgi?article=2443context=utk$_g radthes, July 20, 2016.$

[11] J. John, C. Simon, J. Koopman, L. Sattenspiel, and T. Perry, *Modeling and analyzing hiv transmission: The effect of contact patterns.*, Mathematical Biosciences, (1988), pp. 92: 119–199.

[12] D. G. M. Jr., C. S. Louis, and N. S. Fleur, *Short answers to hard questions about zika virus.* http://www.nytimes.com/interactive/2016/health/what-is-zika-virus.html$_r$ = $0module$ = $ArrowsNavcontentCollection$ = $Healthaction$ = $keypressregion$ = $FixedLeftpgtype$ = $Multimedia, July19, 2016$.

[13] ——, *Short answers to hard questions about zika virus*, Mathematical Biosciences, (June 24th, 2016).

[14] J. M. Mansuy, M. Dutertre, C. Mengelle, C. Fourcade, B. Marchou, P. Delobel, J. Izopet, and G. Martin-Blondel, *Zika virus: high infectious viral load in semen, a new sexually transmitted pathogen*, Lancet Infect Dis, 16 (2016), pp. 00138–9.

[15] G. Mitchell. private communication, 2016.

[16] I. Mundi, *Brazil demographics profile 2014.* http://www.indexmundi.com/brazil/demographics$_p$$rofile.html$.

[17] Z. V. Net, *Zika virus tranmssion.* http://zikavirusnet.com/transmission.html, July 10, 2016.

[18] C. of Disease Control, *Zika virus in brazil.* http://wwwnc.cdc.gov/travel/notices/alert/zika-brazi, July 10, 2016.

[19] W. H. Organization, *Who statement on the first meeting of the international health regulations (2005) (ihr 2005) emergency committee on zika virus and observed increase in neurological disorders and neonatal malformations.* http://www.who.int/mediacentre/news/statements/2016/1st-emergency-committee-zika/en/, July 10, 2016.

[20] ——, *Zika virus.* http://www.who.int/mediacentre/news/statements/2016/1st-emergency-committee-zika/en/, July 11, 2016.

[21] ——, *Zika situation report.* http://www.who.int/emergencies/zika-virus/situation-report/7-july-2016/en/, July 16, 2016.

[22] S. P. T. Resources, *Lgbt rights in brazil.* https://www.susans.org/wiki/LGBT$_r$$ights_in_Brazil, July12, 2016$.

[23] M. Richter, *Female sex work and international sport events-no major changes in demand or supply of paid sex during the 2010 soccer world cup: a cross-sectional study*, BMC public health, 12.1 (2012).

[24] J. Robinson, *What percentage of the population is gay?* http://www.gallup.com/poll/6961/what-percentage-population-gay.aspx, July 12,2016.

[25] S. Towers, F. Brauer, C. Castillo-Chavez, A. Falconar, A. Mubayi, and C. Romero-Vivas, *Estimation of the reproduction number of the 2015 zika virus outbreak in barranquilla, colombia, and a first estimate of the relative role of sexual transmission*, arXiv preprint arXiv:1606.01422, (2016).

[26] UNESO, *Model sensitivity and uncertainty analysis*, Water Resources Systems Planning and Management, (2005), pp. 255–290.

[27] D. Wilson, M. Law, A. Grulich, D. Cooper, and J. Kaldor", *Relation between hiv viral load and infectiousness: a model-based analysis.*, The Lancet, (2008), pp. 314–320.