

Modelando la Propagación de la Fiebre Tifoidea

Román Zúñiga Macías
Modelación Epidemiológica
Centro de Investigación en Matemáticas (CIMAT)

5 de diciembre de 2022

1. Motivación

Hoy en día, la fiebre tifoidea sigue representando un gran reto a nivel mundial, sobretodo en países en desarrollo. Pese a que se han desarrollado diversos medicamentos, tratamientos e incluso vacunas para combatir la fiebre tifoidea, cada año se siguen presentando en el mundo alrededor de 15 millones de casos y 600,000 muertes debido a esta enfermedad [6].

La fiebre tifoidea es una infección causada por la bacteria *Salmonella typhi* que es transmitida a través de la vía fecal-oral, es decir, un individuo es contagiado cuando ingiere algo que fue contaminado con heces de una persona infectada.

A pesar de que actualmente se reportan pocos casos en países desarrollados, en países en desarrollo esta enfermedad sigue representando un gran problema, siendo África y Asia los continentes más afectados. Además, el reto de combatir esta enfermedad se ha hecho más difícil debido a la aparición de variantes de la *Salmonella typhi* que son resistentes a los antibióticos [1].

La fiebre tifoidea está lejos de ser erradicada y nuestro país no es ajeno a ella. Tan sólo en el año 2020 se reportaron 16,858 casos de fiebre tifoidea en México [11], por lo que resulta necesario conocer y describir su forma de propagación para aplicar medidas de control epidemiológico efectivas y así poder afrontar de forma adecuada cada brote infeccioso que se presente.

Es aquí donde los modelos epidemiológicos juegan un papel importante, ya que permiten describir las principales características de la dinámica de la epidemia y tomar mejores decisiones al prever posibles escenarios de la propagación de la enfermedad. Es por esto, que en el presente trabajo se propone un modelo matemático que captura las principales características de la dinámica de la fiebre tifoidea, el cual podría ser una herramienta para el estudio de los brotes epidémicos causados por esta enfermedad.

2. Introducción

En este proyecto se propone un modelo compartimental para la fiebre tifoidea que considera las principales características de transmisión y evolución de la enfermedad. Cabe mencionar que existen dos formas de transmisión de esta enfermedad [10][13]: por contacto directo entre individuos y por contacto indirecto al ingerir agua o alimentos contaminados por la bacteria *Salmonella typhi*. El modelo que se presentará se

enfocará principalmente en la primera, basándose en la ley de acción de masas y suponiendo que entre más grande sea el número de interacciones entre susceptibles e infectados mayor será el número de nuevas infecciones. Y se supondrá que tales interacciones directas no son sólo a través de contacto físico o cercanía entre los individuos, sino que también se presentan cuando un infectado prepara la comida para uno o más susceptibles.

Para introducir el modelo, primero se contemplará una división de la población total en cuatro categorías: S , que corresponde a las personas susceptibles, es decir, a aquellos que se pueden contagiar; E , que representa a los individuos expuestos, que son aquellos que ya están contagiados pero aún no manifiestan síntomas y todavía no son capaces de contagiar a otros; I , que corresponde a los individuos que ya están contagiados, tienen los síntomas propios de la enfermedad y son capaces de infectar a los demás; y por último, C (por *Carriers*), que representa a la fracción de individuos que una vez que se recuperaron de la enfermedad se convierten en portadores asintomáticos de la bacteria y pueden contagiar a otros.

Estas categorías fueron elegidas porque son las que capturan los estados más importantes de los individuos respecto a la fiebre tifoidea. El compartimento de expuestos E se incluyó porque la *Salmonella typhi* posee un periodo de incubación de 7 a 14 días, después del cual inician los síntomas con fiebre y malestar general [7]. De esta manera, cada individuo infectado tiene un periodo de latencia y demora algunos días en volverse capaz de transmitir la enfermedad.

Por otra parte, se creó un compartimento para los portadores asintomáticos C puesto que se ha estimado que entre el 3 y 5 % de quienes se recuperan de la enfermedad se convierten en portadores asintomáticos [12], que resultan ser una fuente importante de nuevos contagios y que pueden durar años en ese estado. De hecho, en un muestreo reportado en [12] se registró que el 8.8 % de los recuperados de fiebre tifoidea pasaron a ser portadores asintomáticos.

Como se ha mostrado que después de recuperarse de la enfermedad no se adquiere un nivel de protección apreciable contra una futura reinfección de fiebre tifoidea [5], en el modelo se supondrá que después de pasar por el estado infeccioso o de portador asintomático se vuelve a ser susceptible inmediatamente.

En la Figura 1 se representan dichos compartimentos y las transiciones consideradas entre ellos para proponer el modelo que describe la propagación de la fiebre tifoidea.

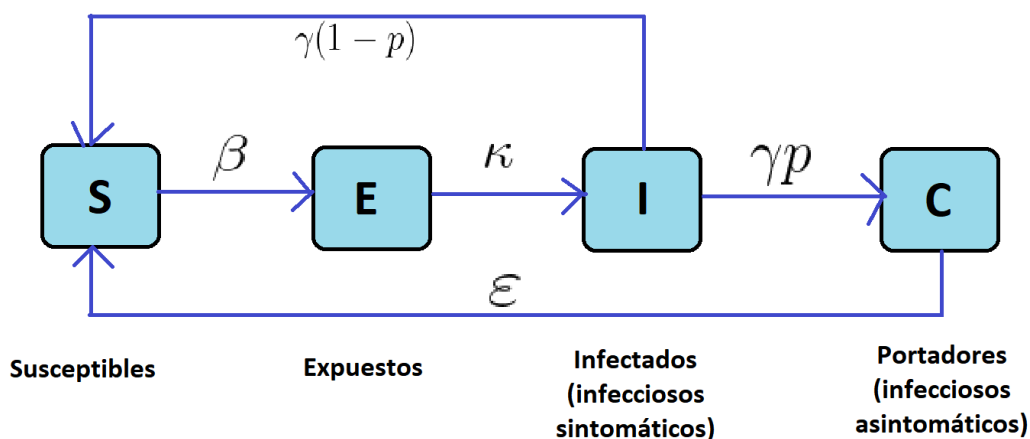


Figura 1

De esta forma, llamaremos modelo SEICS el modelo propuesto con base en las transiciones de los estados representados. El sistema de ecuaciones diferenciales que lo compone se establece a continuación:

$$\begin{aligned}
S'(t) &= -\beta(I + C)\frac{S}{N} + \varepsilon C + \gamma(1 - p)I \\
E'(t) &= \beta(I + C)\frac{S}{N} - \kappa E \\
I'(t) &= \kappa E - \gamma(1 - p)I - \gamma p I \\
C'(t) &= \gamma p I - \varepsilon C,
\end{aligned} \tag{1}$$

el cual considera una población N cerrada, ya que $N'(t) = S'(t) + E'(t) + I'(t) + C'(t) = 0$ y por lo tanto, $N(t) = S(t) + E(t) + I(t) + C(t)$ permanece constante a lo largo del tiempo.

Con respecto a los parámetros del modelo (1), se tiene que β es la tasa de transmisión de la enfermedad, tanto por parte de los infectados como de los portadores. Además, cabe destacar que el término $(I + C)S$ indica el número de contactos entre individuos contagiosos (tanto infectados como portadores) e individuos susceptibles.

En cuanto a κ , esta representa la tasa de transición entre el periodo de latencia y el periodo infeccioso, esto es, indica el tiempo que tarda un individuo en volverse contagioso una vez que fue infectado por la bacteria *Salmonella typhi*. Por otra parte, p corresponde al porcentaje de individuos infecciosos que una vez recuperados se convierten en portadores asintomáticos.

Por otro lado, se tiene que γ es la tasa de recuperación de la enfermedad, siendo a grandes rasgos el recíproco del número de días que conforman el periodo infeccioso. Asimismo, ε indica la tasa de transición de portador asintomático a susceptible, de forma que es el recíproco del número de días que dura la etapa como portador.

3. Antecedentes

Recientemente se han propuesto diferentes modelos compartimentales para la fiebre tifoidea. Entre ellos se encuentra el modelo presentado en [10], en el cual se divide a la población humana en cuatro compartimentos: susceptibles S , infectados I , portadores infectados I y recuperados R . La principal contribución de dicho trabajo es que incorpora tanto la dinámica de transmisión directa (de individuo a individuo) como la indirecta (por contacto con medios contaminados) de la fiebre tifoidea. Para ello, añade el compartimento W que representa la reproducción de las bacterias en el ambiente, suponiendo que estas incrementan su población conforme los infectados y los portadores asintomáticos las excretan.

Similarmente, en el artículo [13] se presenta un modelo que comparte muchas características con el modelo de [10]. Tal modelo también divide a la población humana en cuatro compartimentos: susceptibles S , infectados I , portadores C y recuperados R . De forma análoga, añade un compartimento para la población bacteriana que denomina B_C , sin embargo, la principal aportación de este modelo es que introduce una tasa de ingesta de comida o agua y una probabilidad de que los individuos consuman comida o agua contaminada con la bacteria causante de fiebre tifoidea.

Esto es de suma relevancia puesto que gran parte de las infecciones por fiebre tifoidea se originan por el contacto con bacterias ambientales cuando se ingiere agua o comida contaminada.

Por otro lado, con el objetivo de añadir estrategias de control epidemiológico sobre la propagación de la fiebre tifoidea, en [9] se realiza una extensión del modelo de [10], mientras que en el mismo artículo [13] se propone una extensión del modelo correspondiente. En [9] se añaden al modelo, mediante variables de control, el efecto que tienen las campañas educativas, la prevención a través de saneamiento y la detección de portadores junto con el tratamiento temprano. Asimismo, en la versión del modelo de [13] con control epidemiológico se incorporan tres medidas de control: prevención a través de medidas sanitarias y un higiene adecuado, tratamiento de personas que desarrollaron síntomas de la enfermedad y la detección de portadores.

En los trabajos [3] y [8] se propone un modelo para la fiebre tifoidea en el que uno de los compartimentos corresponde a los individuos vacunados, suponiendo que la vacuna confiere una inmunidad temporal. De acuerdo con [1], las campañas de vacunación en países en desarrollo han sido de gran importancia para combatir la difusión de la fiebre tifoidea, por lo que es clara la relevancia de incluir la vacunación a los modelos de esta enfermedad.

En [3] el modelo propuesto sí posee un compartimento para portadores asintomáticos, definido igual que en [10] y [13]. Caso contrario es el del modelo en [8] que no tiene un compartimento para los portadores asintomáticos, individuos que han resultado ser una fuente importante de nuevas infecciones [12].

Por otra parte, una diferencia clara del modelo de [8] respecto a los demás que se han mencionado en esta sección, es que propone un compartimento T para las personas que están bajo tratamiento para recuperarse de la enfermedad.

Cabe destacar que todos los modelos mencionados consideran una tasa de natalidad y de mortalidad humana. En el caso de [10] y [13] se incluye también una tasa de mortalidad bacteriana. Adicionalmente, en [13] se incluye una tasa de mortalidad a causa de la fiebre tifoidea.

Teniendo como referencia los modelos anteriores para la fiebre tifoidea, se expondrán algunas aportaciones del modelo SEICS propuesto en este proyecto. En primer lugar, a diferencia de todos los artículos mencionados previamente, el modelo SEICS sí incluye una categoría de expuestos, lo cual resulta importante pues se ha estimado que el periodo de latencia puede durar varias semanas [7].

Otro aspecto a resaltar es que en los modelos mencionados que sí incluyen portadores asintomáticos, estos se consideran de distinta manera a cómo se hace en el modelo SEICS. Suponen que una vez que una persona es infectada puede convertirse en infectado o en portador, esto es, los portadores son infectados asintomáticos que sí pueden infectar. No obstante, el caso de Mary Typhoid [2] y algunos estudios como los reportados en [12], han probado que los individuos que se convierten en portadores asintomáticos una vez que se recuperaron son una fuente importante de nuevas infecciones, y no precisamente aquellos que desde que se infectaron son asintomáticos.

Por otro lado, en el modelo SEICS sí se presenta un punto de equilibrio endémico, como se mostró en una sección anterior. Esto parece tener coherencia con la permanencia de la fiebre tifoidea en la población a través de las décadas. En el caso de [13], [8] y [3], los modelos respectivos también poseen un punto de equilibrio endémico.

En contraste, en [10] no se habla de los puntos de equilibrio del modelo, y por las simulaciones mostradas parece que no se presenta un equilibrio endémico.

En dicho modelo de [10], se hace el supuesto de que una vez recuperado se adquiere una inmunidad total, lo cual se ha mostrado que no es así y que las reinfecciones son muy frecuentes [5]. El modelo SEICS sí considera esta característica de la dinámica de la enfermedad y considera como hipótesis que no hay estado de inmunidad temporal puesto que estudios señalan que, en general, el haber estado enfermo no te

protege de infecciones futuras [5].

4. El Modelo Propuesto Está Bien Planteado

Para mostrar que el modelo SEICS está bien planteado, primero, se mostrará que el sistema en (1) tiene solución única en $[0, T]$. Para ello, calculemos la matriz jacobiana asociada al sistema:

$$J(S, E, I, C) = \begin{pmatrix} -\frac{\beta}{N}(I + C) & 0 & -\frac{\beta S}{N} + \gamma(1 - p) & -\frac{\beta S}{N} + \varepsilon \\ \frac{\beta}{N}(I + C) & -\kappa & \frac{\beta S}{N} & \frac{\beta S}{N} \\ 0 & \kappa & -\gamma & 0 \\ 0 & 0 & \gamma p & -\varepsilon \end{pmatrix} \quad (2)$$

Puesto que todas las componentes de la matriz son continuas, se obtiene que $f_1 = -\beta(I + C)\frac{S}{N} + \varepsilon C + \gamma(1 - p)I$, $f_2 = \beta(I + C)\frac{S}{N} - \kappa E$, $f_3 = \kappa E - \gamma(1 - p)I - \gamma p I$ y $f_4 = \gamma p I - \varepsilon C$ son continuamente diferenciables. De esta manera, $f = (f_1, f_2, f_3, f_4)$ es localmente Lipschitz continua. Luego, por el Teorema de Picard y Lindelöf el sistema tiene solución única en $[0, T]$.

Ahora bien, se mostrará que el conjunto

$$\mathbb{R}_+^4 = \{(S, E, I, C) \in \mathbb{R}^4 | S \geq 0, E \geq 0, I \geq 0, C \geq 0\}$$

es no negativo.

Considerando el espacio tridimensional EIC , con $S = 0$, tenemos que:

$$\frac{dS}{dt}|_{S=0} = \varepsilon C + \gamma(1 - p)I \geq 0,$$

por lo que el campo vectorial en el espacio EIC apunta hacia el interior de \mathbb{R}_+^4 , de modo que ninguna solución de \mathbb{R}_+^4 puede escapar a través de dicho espacio.

Ahora, considerando el espacio tridimensional SIC , esto es, donde $E = 0$, tenemos que:

$$\frac{dE}{dt}|_{E=0} = \beta(I + C)\frac{S}{N} \geq 0,$$

obteniendo que el campo vectorial apunta hacia el interior de \mathbb{R}_+^4 y ninguna solución puede escapar a través del espacio SIC .

Por otro lado, al considerar el espacio tridimensional SEC , es decir, donde $I = 0$, tenemos que:

$$\frac{dI}{dt}|_{I=0} = \kappa E \geq 0,$$

resultando que el campo vectorial nuevamente apunta hacia el interior \mathbb{R}_+^4 y ninguna solución puede escapar a través del espacio SEC .

Por último, tomando el espacio SEI (con $C = 0$), encontramos que:

$$\frac{dC}{dt}|_{C=0} = \gamma p I \geq 0,$$

obteniendo que el campo vectorial apunta hacia el interior de \mathbb{R}_+^4 y ninguna solución escapa mediante el espacio SEI .

Se concluye que todas las soluciones que comienzan en \mathbb{R}_+^4 permanecen en \mathbb{R}_+^4 , $\forall t > 0$. Es así, que el modelo propuesto está bien planteado.

5. Encontrando Puntos de Equilibrio

Con el propósito de encontrar los puntos críticos del sistema (1), se debe resolver el siguiente sistema homogéneo:

$$\begin{aligned} -\beta(I+C)\frac{S}{N} + \varepsilon C + \gamma(1-p)I &= 0 \\ \beta(I+C)\frac{S}{N} - \kappa E &= 0 \\ \kappa E - \gamma(1-p)I - \gamma p I &= 0 \\ \gamma p I - \varepsilon C &= 0. \end{aligned} \quad (3)$$

Para solucionarlo, se buscará dejar todas las variables en términos de I . Tomando la tercera y la cuarta ecuación de (3), se tiene que:

$$C = \frac{\gamma p}{\varepsilon} I, \quad E = \frac{(\gamma(1-p) + \gamma p)}{\kappa} I = \frac{\gamma I}{\kappa}, \quad (4)$$

de donde se puede encontrar una expresión de S que sólo dependa de la variable I :

$$S = \frac{\kappa E N}{\beta(I+C)} = \frac{\gamma I N}{\beta(I+C)},$$

y si se supone que **el punto de equilibrio es endémico entonces** $I \neq 0$, de forma que:

$$S = \frac{\gamma I N}{\beta I (1 + \frac{\gamma p}{\varepsilon})} = \frac{\gamma N}{\beta (1 + \frac{\gamma p}{\varepsilon})}.$$

Como el modelo propuesto es de población cerrada, se puede obtener el valor de I definido sólo a través de los parámetros:

$$\begin{aligned} I &= N - S - E - C = N - \frac{\gamma N}{\beta (1 + \frac{\gamma p}{\varepsilon})} - \frac{\gamma I}{\kappa} - \frac{\gamma p I}{\varepsilon} \\ &= \frac{N \left(1 - \frac{\gamma}{\beta (1 + \frac{\gamma p}{\varepsilon})} \right)}{1 + \gamma \left(\frac{1}{\kappa} + \frac{p}{\varepsilon} \right)} = \frac{N (\beta (1 + \frac{\gamma p}{\varepsilon}) - \gamma)}{\beta (1 + \frac{\gamma p}{\varepsilon}) (1 + \gamma (\frac{1}{\kappa} + \frac{p}{\varepsilon}))} \\ &= \frac{N (\beta (1 + \frac{\gamma p}{\varepsilon}) - \gamma)}{\beta (1 + \frac{\gamma p}{\varepsilon}) (1 + \frac{\gamma p}{\varepsilon} + \frac{\gamma}{\kappa})}. \end{aligned} \quad (5)$$

Sustituyendo este valor de I en las ecuaciones (4) para las variables C y E :

$$C = \frac{\gamma p N (\beta (1 + \frac{\gamma p}{\varepsilon}) - \gamma)}{\varepsilon \beta (1 + \frac{\gamma p}{\varepsilon}) (1 + \frac{\gamma p}{\varepsilon} + \frac{\gamma}{\kappa})}, \quad E = \frac{\gamma N (\beta (1 + \frac{\gamma p}{\varepsilon}) - \gamma)}{\kappa \beta (1 + \frac{\gamma p}{\varepsilon}) (1 + \frac{\gamma p}{\varepsilon} + \frac{\gamma}{\kappa})}.$$

De esta manera, el **punto de equilibrio endémico** es:

$$(S_1^*, E_1^*, I_1^*, C_1^*) = \left(\frac{\gamma N}{\beta (1 + \frac{\gamma p}{\varepsilon})}, \frac{\gamma I^*}{\kappa}, I^*, \frac{\gamma p I^*}{\varepsilon} \right), \quad \text{donde } I^* = \frac{N (\beta (1 + \frac{\gamma p}{\varepsilon}) - \gamma)}{\beta (1 + \frac{\gamma p}{\varepsilon}) (1 + \frac{\gamma p}{\varepsilon} + \frac{\gamma}{\kappa})}. \quad (6)$$

Cabe mencionar que al intentar estudiar la estabilidad de este punto de equilibrio mediante la linealización del sistema, se obtienen expresiones algebraicas que resultan ser muy complejas de analizar a pesar de que se simplifiquen.

Si por otro lado, se considera que el **punto de equilibrio es libre de enfermedad**, esto es, **si se supone que** $I = 0$, tomando las ecuaciones en (4) tenemos que $E = 0$ y $C = 0$. Luego, $S = N - E - I - C = N$, resultando que el punto de equilibrio en este caso es $(S_2^*, E_2^*, I_2^*, C_2^*) = (N, 0, 0, 0)$.

Con respecto a este punto de equilibrio se intentará analizar su estabilidad mediante linealización. Tomando la matriz jacobiana calculada previamente en (2) y evaluándola en $(S_2^*, E_2^*, I_2^*, C_2^*)$, se obtiene que:

$$J(N, 0, 0, 0) = \begin{pmatrix} 0 & 0 & -\beta + \gamma(1-p) & -\beta + \varepsilon \\ 0 & -\kappa & \beta & \beta \\ 0 & \kappa & -\gamma & 0 \\ 0 & 0 & \gamma p & -\varepsilon \end{pmatrix}.$$

Tenemos que el primer eigenvalor es $\lambda_1 = 0$, mientras que los otros tres se obtienen de calcular los eigenvalores de la siguiente submatriz:

$$\det \begin{pmatrix} \lambda + \kappa & -\beta & -\beta \\ -\kappa & \lambda + \gamma & 0 \\ 0 & -\gamma p & \lambda + \varepsilon \end{pmatrix},$$

resultando que este determinante es igual a $\lambda^3 + \lambda^2(\kappa\gamma + \kappa\varepsilon + \gamma\varepsilon) + \lambda(\gamma + \kappa + \varepsilon - \beta\varepsilon) - \kappa\beta\gamma p$, cuyas raíces fueron calculadas con el software Maxima y resultaron ser difíciles de analizar al ser expresiones algebraicas muy largas. Sin embargo, existe un teorema que garantiza que **el punto de equilibrio libre de enfermedad es localmente estable si** $\mathcal{R}_0 < 1$. De este modo, a continuación se procederá a calcular dicho \mathcal{R}_0 .

6. Obteniendo el Número Reproductivo Básico

Para calcular el número reproductivo básico del modelo propuesto, mejor conocido como \mathcal{R}_0 , se recurrirá a la obtención del radio espectral de la aproximación de la matriz de siguiente generación.

Identificando que el modelo cuenta con tres estados infecciosos, E, I y C , procedemos a calcular las funciones $\mathcal{F} : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ y $\mathcal{V} : \mathbb{R}^4 \rightarrow \mathbb{R}^3$, representadas vectorialmente como sigue:

$$\mathcal{F} = \begin{pmatrix} \beta(I + C)\frac{S}{N} \\ 0 \\ 0 \end{pmatrix}, \mathcal{V} = \begin{pmatrix} \kappa E \\ -\kappa E + \gamma I \\ -\gamma p I + \varepsilon C \end{pmatrix}.$$

Luego, a partir de las derivadas parciales de estas funciones respecto a los estados infecciosos, y evaluándolas en el punto de equilibrio libre de enfermedad $(S_2^*, E_2^*, I_2^*, C_2^*) = (N, 0, 0, 0)$, obtenemos las siguientes matrices:

$$F = \begin{pmatrix} 0 & \beta & \beta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, V = \begin{pmatrix} \kappa & 0 & 0 \\ -\kappa & \gamma & 0 \\ 0 & -\gamma p & \varepsilon \end{pmatrix}.$$

Para encontrar la matriz con la que se aproximará a la matriz de siguiente generación, se necesita calcular V^{-1} . De esta manera, tenemos que la adjunta de V está dada por:

$$Adj(V) = \begin{pmatrix} \gamma\varepsilon & \kappa\varepsilon & \kappa\gamma p \\ 0 & \kappa\varepsilon & -\kappa\gamma \\ 0 & 0 & \kappa\gamma \end{pmatrix}$$

Por lo tanto, la inversa buscada es la siguiente:

$$V^{-1} = \frac{Adj(V)^T}{det(V)} = \frac{1}{\kappa\gamma\varepsilon} \begin{pmatrix} \gamma\varepsilon & 0 & 0 \\ \kappa\varepsilon & \kappa\varepsilon & 0 \\ \kappa\gamma p & -\kappa\gamma & \kappa\gamma \end{pmatrix} = \begin{pmatrix} \frac{1}{\kappa} & 0 & 0 \\ \frac{1}{\gamma} & \frac{1}{\varepsilon} & 0 \\ \frac{p}{\varepsilon} & -\frac{1}{\varepsilon} & \frac{1}{\varepsilon} \end{pmatrix}$$

Es así, que a continuación se calcula la aproximación a la matriz de siguiente generación:

$$FV^{-1} = \begin{pmatrix} 0 & \beta & \beta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\kappa} & 0 & 0 \\ \frac{1}{\gamma} & \frac{1}{\varepsilon} & 0 \\ \frac{p}{\varepsilon} & -\frac{1}{\varepsilon} & \frac{1}{\varepsilon} \end{pmatrix} = \begin{pmatrix} \frac{\beta}{\gamma} + \frac{\beta p}{\varepsilon} & \frac{\beta}{\gamma} - \frac{\beta}{\varepsilon} & \frac{\beta}{\varepsilon} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Como resultado, tenemos que \mathcal{R}_0 para el modelo propuesto es el radio espectral de FV^{-1} :

$$\mathcal{R}_0 = \rho(FV^{-1}) = \frac{\beta}{\gamma} + \frac{\beta p}{\varepsilon},$$

expresión que tiene una interpretación directa, ya que el primer sumando indica el número de nuevos casos causados por infecciosos sintomáticos durante su periodo infeccioso, mientras que el segundo sumando corresponde a los nuevos casos producidos por portadores durante su etapa como infecciosos asintomáticos.

Utilizando el valor del número reproductivo básico reescrito como $\mathcal{R}_0 = \beta \frac{(\varepsilon + \gamma p)}{\gamma\varepsilon}$, se puede reescribir el valor de I_1^* en el punto de equilibrio endémico encontrado en (6):

$$I_1^* = \frac{\frac{1}{\gamma} \frac{1}{\beta\gamma}}{\frac{1}{\gamma} \frac{\beta\gamma}{\beta\gamma}} \frac{N (\beta (1 + \frac{\gamma p}{\varepsilon}) - \gamma)}{\beta (1 + \frac{\gamma p}{\varepsilon}) (1 + \frac{\gamma p}{\varepsilon} + \frac{\gamma}{\kappa})} = \frac{\beta N (\mathcal{R}_0 - 1)}{\gamma \mathcal{R}_0 (\mathcal{R}_0 + \frac{\beta}{\kappa})} = \frac{\beta N (1 - \frac{1}{\mathcal{R}_0})}{\gamma (\mathcal{R}_0 + \frac{\beta}{\kappa})}. \quad (7)$$

Por lo tanto, el punto de equilibrio endémico del modelo está definido en términos de \mathcal{R}_0 como sigue:

$$(S_1^*, E_1^*, I_1^*, C_1^*) = \left(\frac{N}{\mathcal{R}_0}, \frac{\gamma I_1^*}{\kappa}, I_1^*, \frac{\gamma p I_1^*}{\varepsilon} \right), \text{ donde } I_1^* = \frac{\beta N (1 - \frac{1}{\mathcal{R}_0})}{\gamma (\mathcal{R}_0 + \frac{\beta}{\kappa})}. \quad (8)$$

Esto nos aporta un resultado importante, el cual consiste en que el punto de equilibrio endémico se presenta sólo si $\mathcal{R}_0 > 1$, ya que si $\mathcal{R}_0 < 1$ entonces se obtendría que $S_1^* > N$ y esto contradice que la población considerada en el modelo es cerrada. Si $\mathcal{R}_0 = 1$ el punto de equilibrio endémico coincide con el libre de enfermedad.

Por otro lado, si se intenta nuevamente aplicar el método de linealización del sistema en torno a este punto de equilibrio reescrito, todo se reduce a estudiar las raíces del siguiente polinomio característico:

$$\rho(\lambda) = \lambda \left[\lambda^3 + \lambda^2 \left(\varepsilon + \gamma + \kappa + \frac{\beta(\mathcal{R}_0 - 1)}{\mathcal{R}_0 + \frac{\beta}{\kappa}} \right) + \lambda \left(\varepsilon\gamma + \gamma\kappa + \varepsilon\kappa + (\varepsilon + \gamma + \kappa) \frac{\beta(\mathcal{R}_0 - 1)}{\mathcal{R}_0 + \frac{\beta}{\kappa}} - \frac{\kappa\beta}{\mathcal{R}_0} \right) + \varepsilon\gamma\kappa + (\varepsilon\gamma + \varepsilon\kappa + \kappa\gamma p) \frac{\beta(\mathcal{R}_0 - 1)}{\mathcal{R}_0 + \frac{\beta}{\kappa}} - \kappa(\gamma p - \varepsilon) \frac{\beta}{\mathcal{R}_0} \right],$$

con el que sólo se logra obtener que un eigenvalor es $\lambda_1 = 0$, pero el polinomio de tercer grado que falta por analizar sigue siendo muy complejo. Las raíces correspondientes a este polinomio se obtuvieron a través del software Maxima, teniendo que la expresión de cada una de ellas es extremadamente larga.

7. Simulación de Datos de Incidencia

Para poder generar datos de incidencia por periodos de tiempo, se introdujo la variable $Y(t)$, la cual guarda el número de casos acumulados hasta el instante t y cuya ecuación diferencial está dada por $Y'(t) = \kappa E$. Esta ecuación se añadió al sistema (1) al momento de resolverlo numéricamente con la finalidad de obtener también la solución para Y .

De este modo, se usó el modelo para simular datos de incidencia agregando un ruido Poisson, ya que en el modelo se contabilizan individuos en cada compartimento y no proporciones de ellos. Considerando como unidad de tiempo un día, **los parámetros que fueron tomados como los reales** para generar los datos, son los siguientes: $\gamma = \frac{1}{21} \approx 0.0476$, considerando un periodo de recuperación de 21 días, el cual suele abarcar de 14 a 21 días [6]; $\beta = \frac{3}{21} \approx 0.1428$, que es el triple de γ para que el primer sumando de \mathcal{R}_0 sea de 3; $\kappa = \frac{1}{10} = 0.1$, considerando un periodo de incubación de la bacteria *Salmonella typhi* de 10 días, el cual suele estar entre 7 y 14 días [7]; $\varepsilon = \frac{1}{365} \approx 0.0027$ tomando un periodo de transición de portador asintomático a susceptible de 365 días, el cual se ha encontrado que puede durar varios años [12]; y por último, se tomó $p = 0.04$ ya que del total de recuperados entre el 3 y 5 % pasan a ser portadores asintomáticos [12]. Cabe mencionar, que el \mathcal{R}_0 **asociado** al modelo con estos parámetros es de 5.0857.

Por otro lado, se establecieron $N = 100,000$ como población total, $S_0 = 99,975$, $E_0 = 0$, $I_0 = 25$, $Y_0 = I_0 = 25$ y $C_0 = 0$ como **condiciones iniciales**, y un intervalo de tiempo de 840 días. Este periodo de tiempo es de más de dos años, ya que con el valor elegido de la tasa de transición ε se supuso que el periodo como portador asintomático es de un año y se quiere observar el modelo en un tiempo suficiente para que se presenten transiciones de portadores a susceptibles.

Con estas condiciones iniciales y los parámetros mencionados anteriormente, en este caso el **punto de equilibrio endémico** es $(S_1^*, E_1^*, I_1^*, C_1^*) = (19662.92, 17617.78, 36997.33, 25721.95)$.

Tomando la solución numérica obtenida para $Y(t)$, que nos dice el número de casos acumulados al día t , se puede calcular la incidencia diaria al calcular las diferencias entre valores consecutivos de dicha solución. A partir de los datos de incidencia diaria se pueden generar los de incidencia de cualquier periodo de m días. Para ello, se divide el intervalo original en grupos consecutivos y disjuntos de m días, para después asignarle a cada uno de ellos la suma de los casos correspondientes a esos m días.

Luego, a tales datos de incidencia de m días se les agregó ruido Poisson al generar valores aleatorios de esta distribución usando como media el dato correspondiente.

En la Figura 2 se muestran los **datos de incidencia por semana** ($m = 7$) generados con el procedimiento anterior. En total se observan 120 semanas y se enumeran desde la semana 0, presentándose el pico de la epidemia en la semana 26.

Por otra parte, en la Figura 3 se muestran los **datos de incidencia quincenales** ($m = 15$) generados con el mismo procedimiento pero usando una semilla distinta al caso anterior en el código. En total se observan

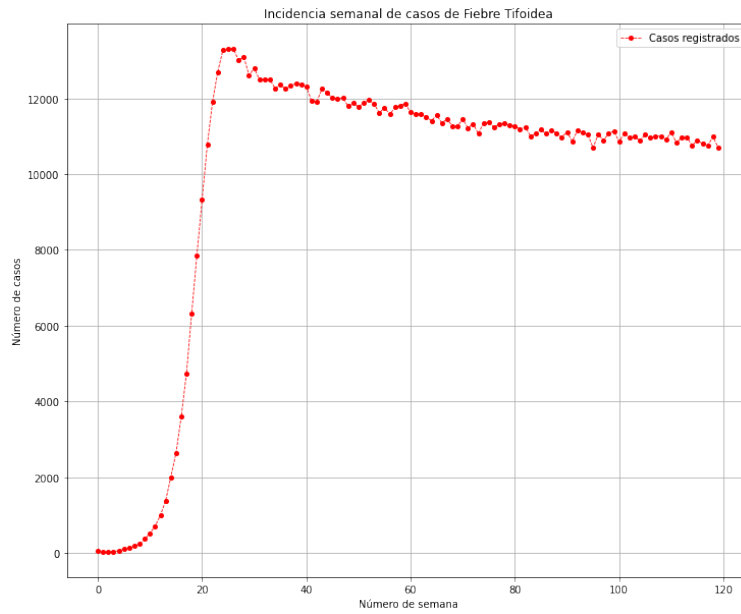


Figura 2: Datos simulados de incidencia semanal.

56 quincenas que se enumeran desde la quincena 0, y en este caso el pico de la epidemia se alcanza en la quincena 12.

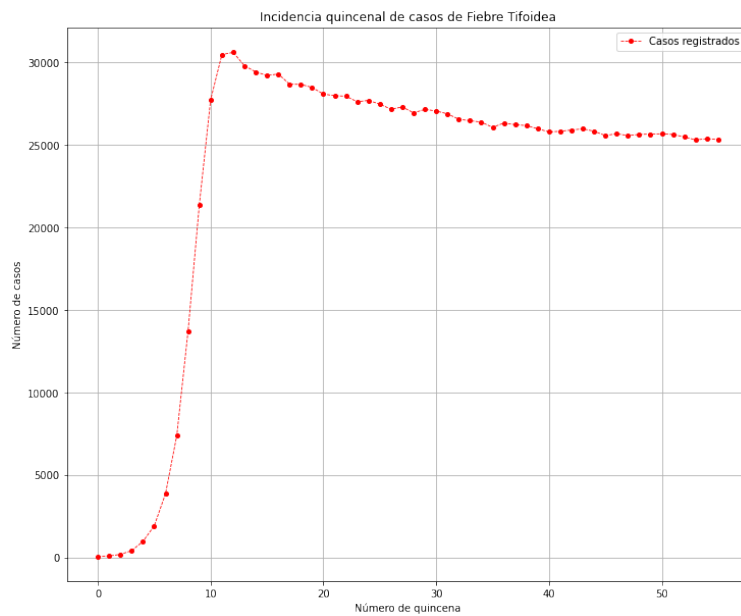


Figura 3: Datos simulados de incidencia quincenal.

En ambas gráficas, se percibe cómo el número de casos de incidencia está tendiendo a estabilizarse poco a poco ya que el modelo tiene un punto de equilibrio endémico, el cual se presenta en $I_1^* = 36,997.33$. No obstante, recordemos que en $I(t)$ están contabilizados todos aquellos infectados que aparecieron en los últimos 21 días. Por lo tanto, el equilibrio para el caso de incidencia semanal debería presentarse alrededor de la tercera parte de I_1^* que son aproximadamente 12,332 casos, mientras que para la incidencia quincenal el equilibrio debería estar alrededor de $\frac{15}{21} = \frac{5}{7}$ partes de I_1^* , esto es, aproximadamente 26,426 casos.

8. Ajuste del Modelo Minimizando la Función de Pérdida Cuadrática

Para realizar un ajuste tanto a los datos de incidencia semanal como a los de incidencia quincenal, se utilizó la función de pérdida cuadrática dada por la siguiente ecuación:

$$\sum_{j=0}^{n-1} (X_j - \hat{X}_j)^2, \quad (9)$$

donde $\{X_j\}_{j=0}^{n-1}$ son los datos sintéticos de incidencia generados previamente mientras que $\{\hat{X}_j\}_{j=0}^{n-1}$ son los datos de incidencia estimados con el modelo. De esta forma, se buscan los parámetros $\hat{\beta}, \hat{\kappa}, \hat{\gamma}, \hat{\varepsilon}$ y \hat{p} con los que la solución del modelo minimice la función de error (9).

Para poder estimar dichos parámetros, se utilizó el método de optimización de Levenberg-Marquardt mediante la función *minimize* de la paquetería *lmfit* de Python 3.

Primero, se hizo el ajuste del modelo a los datos de incidencia semanal, brindándole al algoritmo los siguientes valores iniciales y rangos de búsqueda para cada parámetro:

Parámetro	Valor inicial	Rango de búsqueda
β	0.2	$[0, 0.5]$
κ	0.2	$[0, 0.5]$
γ	0.1	$[0, 0.5]$
ε	0.01	$[0, 0.1]$
p	0.1	$[0, 0.2]$

De esta manera, los **valores estimados de los parámetros usando los datos de incidencia semanal** son: $\hat{\beta} = 0.1321$, $\hat{\kappa} = 0.1179$, $\hat{\gamma} = 0.0458$, $\hat{\varepsilon} = 0.0030$ y $\hat{p} = 0.0435$. Estos parámetros no difieren mucho de los parámetros reales, lo cual también se refleja en que el \mathcal{R}_0 calculado a partir de estos parámetros estimados es de 4.7758, que no está muy alejado del obtenido con los parámetros reales y que es igual a 5.0857.

En efecto, al observar el resultado obtenido de forma gráfica en la Figura 4, la curva obtenida a través del modelo se ajusta muy bien sobretodo en las semanas situadas antes del pico. Incluso es capaz de alcanzar el pico casi al mismo tiempo que los datos simulados, aunque sobreestima un poco el número máximo de casos.

Ahora bien, al hacer el ajuste del modelo a los datos de incidencia quincenal, dándole al algoritmo los mismos valores iniciales y rangos de búsqueda que en el caso anterior, **se estimaron los siguientes**

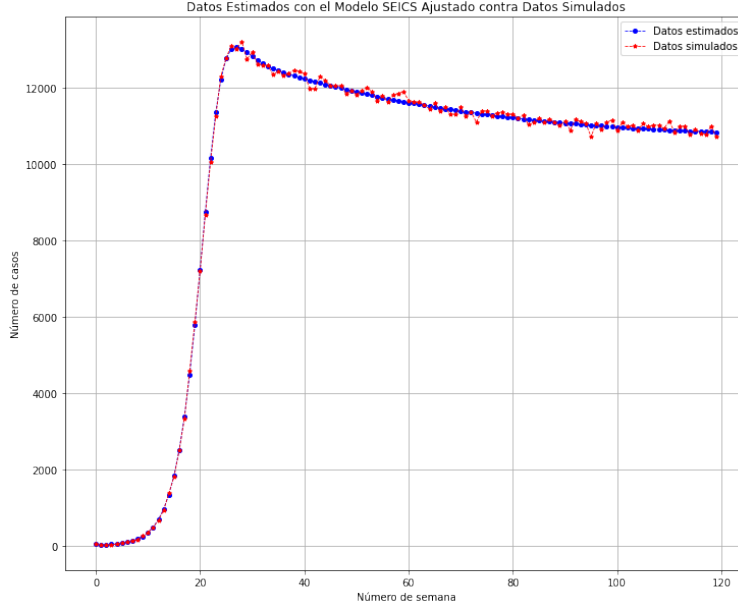


Figura 4: Modelo ajustado para incidencia semanal.

parámetros $\hat{\beta} = 0.1501, \hat{\kappa} = 0.0921, \hat{\gamma} = 0.0503, \hat{\varepsilon} = 0.0031$ y $\hat{p} = 0.04609$. Al igual que en el caso de incidencia semanal, al comparar estos valores estimados de los parámetros con los reales se observa una diferencia pequeña. No obstante, en este caso el \mathcal{R}_0 es igual a 5.1836 que **está aún más cerca del \mathcal{R}_0 original** que el obtenido con el ajuste realizado usando la incidencia semanal.

Al visualizar la gráfica del resultado en la Figura 5, se aprecia un ajuste bastante bueno que **parece ser mejor que el anterior**. En este caso el pico de la curva ajustada sí coincide con el de los datos simulados, tanto en el instante en que se presentó como en la cantidad de casos.

Para constatar que el utilizar datos de incidencia quincenal mejoró el ajuste obtenido se recurrió al Error Absoluto Medio Porcentual (MAPE por sus siglas en inglés), entre los datos sintéticos de incidencia X y los datos predichos con el ajuste \hat{X} . Este error está definido de la siguiente manera:

$$MAPE(X, \hat{X}) = \frac{100}{n} \sum_{j=0}^{n-1} \left| \frac{X_j - \hat{X}_j}{X_j} \right|. \quad (10)$$

En el caso del ajuste de los datos de incidencia semanal el error calculado con (10) fue de 0.01699, mientras que en el ajuste de los datos de incidencia quincenal fue de 0.007979. Puesto que dicho error disminuyó al usar la incidencia quincenal, se puede concluir que aumentando el periodo de incidencia a 15 días los resultados mejoraron.

Por lo tanto, en la siguiente sección, donde se realizará la estimación de los parámetros, pero ahora usando inferencia bayesiana, **se optará por los datos de incidencia quincenal**.

Cabe mencionar que todo el código utilizado para la creación de los datos sintéticos así como el ajuste mediante la minimización de la función de pérdida cuadrática, se encuentra en el archivo *ajuste_seics_roman.py*, el cual se encuentra en el repositorio de GitHub de este proyecto.

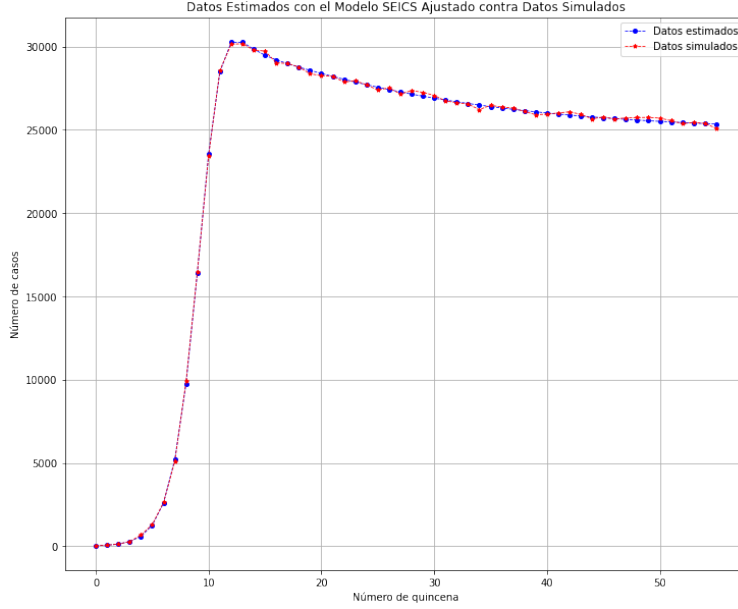


Figura 5: Modelo ajustado para incidencia quincenal.

9. Infiriendo los Parámetros del Modelo Mediante Métodos MCMC

9.1. Modelo de los Datos

Supongamos que los datos observados de incidencia quincenal son X_0, X_1, \dots, X_{n-1} , correspondientes a los intervalos de tiempo $[t_0, t_1), [t_1, t_2), \dots, [t_{n-1}, t_n)$, donde cada intervalo de tiempo abarca una quincena.

Queremos calcular el número de individuos que pasan del compartimento E al compartimento I entre los instantes t_i y t_{i+1} , para $i = 0, \dots, n-1$. Para ello, como ya se había mencionado en las secciones anteriores, agregamos al modelo SEICS la ecuación $Y'(t) = \kappa E(t)$, de forma que con la variable Y se almacena el número de casos infectados acumulados. De esta manera, la incidencia estimada a través del modelo está dada por $\hat{Y}_i(\theta) = \int_{t_i}^{t_{i+1}} \kappa E(t) dt$, donde $\theta = (\beta, \kappa, \gamma, \varepsilon, p)$.

Por lo tanto, consideraremos el número de casos reportados de nuevos infectados por quincena como una variable aleatoria de conteo que tiene como media a $\hat{Y}_i(\theta)$, tal y como se consideró al generar los datos sintéticos. De esta forma, el modelo de datos a utilizar será una distribución Poisson, ya que además de que los datos son de conteo, tenemos que el tiempo que tarda un individuo en pasar del compartimento E al compartimento I es exponencial.

Así, tenemos que $X_i \sim \text{Poisson}(\hat{Y}_i(\theta))$, $i = 0, \dots, n-1$, con $\theta = (\beta, \kappa, \gamma, \varepsilon, p)$ y donde $\hat{Y}_i(\theta)$ es la incidencia de la i -ésima quincena aproximada con el modelo SEICS que tiene a θ como sus parámetros.

Con base en esto, tenemos que la verosimilitud de las observaciones está dada por la siguiente expresión:

$$L(\underline{X}|\theta) \propto \prod_{i=0}^{n-1} e^{-\hat{Y}_i(\theta)} (\hat{Y}_i(\theta))^{X_i}.$$

9.2. Elección de Distribuciones a Priori

Como la distribución $Gamma(\alpha^*, \beta^*)$ tiene soporte $[0, \infty)$, es muy flexible en cuanto a las formas que puede presentar y a los intervalos en los que se puede concentrar la mayor parte de su densidad, y además es una a priori conjugada de la distribución Poisson (aunque en este caso por cada observación se tiene una media $\lambda = \hat{Y}_i(\theta)$ distinta), se elegirá este tipo de distribución como a priori para los parámetros: β, κ, γ y ε .

Por otro lado, como el parámetro p es un porcentaje e incluso se puede interpretar como la probabilidad de convertirse en portador asintomático, se usará una distribución $Beta(\alpha^*, \beta^*)$ como a priori, que precisamente tiene soporte $[0, 1]$ y es conjugada de parámetros que representan probabilidades.

Asimismo, cabe destacar que supondremos que las cinco distribuciones a priori son independientes, por lo que las distribuciones marginales posteriores también lo serán.

Para determinar los hiperparámetros de cada una de estas distribuciones a priori, nos basaremos en la metodología propuesta en [4].

En general, supongamos que los hiperparámetros que queremos determinar son α^* y β^* . Sea φ el parámetro al que le estamos asignando la distribución a priori y $[L_\varphi, U_\varphi]$ un intervalo de incertidumbre sobre cuál es el valor de φ . Entonces se propone que:

$$\mathbb{E}(\varphi|\alpha^*, \beta^*) = \frac{L_\varphi + U_\varphi}{2}, \quad Var(\varphi|\alpha^*, \beta^*) = \left(\frac{U_\varphi - L_\varphi}{2z} \right)^2, \quad (11)$$

donde $z \in [1.2, 3.5]$, y de acuerdo a la desigualdad de Chebyshev $2z$ veces la desviación estándar cubre la mayor parte del rango de variabilidad de φ .

Por otro lado, la extensión de la desigualdad de Chebyshev establece que si la distribución de φ es unimodal, entonces:

$$P(\mu_\varphi - z\sigma_\varphi < \varphi < \mu_\varphi + z\sigma_\varphi) \geq 1 - \frac{1}{2.25z^2},$$

donde μ_φ y σ_φ son la media y la desviación estándar de φ . De esta manera, si los hiperparámetros α^* y β^* satisfacen las ecuaciones en (11), entonces al menos el $(1 - \frac{1}{2.25z^2})\%$ de la distribución a priori con dichos hiperparámetros estará contenida en el intervalo $[L_\varphi, U_\varphi]$.

Para el caso particular de la distribución a priori $Gamma(\alpha^*, \beta^*)$ establecida como sigue:

$$P(\varphi) = \frac{(\beta^*)^{\alpha^*}}{\Gamma(\alpha^*)} \varphi^{\alpha^*-1} e^{-\beta^*\varphi}, \quad \text{con } \alpha^* > 0, \beta^* > 0,$$

tenemos que, dado el intervalo de incertidumbre $[L_\varphi, U_\varphi]$, los hiperparámetros deben cumplir las siguientes ecuaciones:

$$\frac{\alpha^*}{\beta^*} = \frac{L_\varphi + U_\varphi}{2}, \quad \frac{\alpha^*}{(\beta^*)^2} = \left(\frac{U_\varphi - L_\varphi}{2z} \right)^2.$$

Resolviendo dichas ecuaciones, encontramos que los hiperparámetros buscados son los siguientes:

$$\beta^* = \frac{2z^2(L_\varphi + U_\varphi)}{(U_\varphi - L_\varphi)^2}, \quad \alpha^* = \left(\frac{z(L_\varphi + U_\varphi)}{U_\varphi - L_\varphi} \right)^2 \quad (12)$$

Por otro lado, se tiene el caso de la distribución $Beta(\alpha^*, \beta^*)$ que cabe mencionar que también es unimodal. Sus hiperparámetros deben cumplir las siguientes ecuaciones:

$$\frac{\alpha^*}{\alpha^* + \beta^*} = \frac{L_\varphi + U_\varphi}{2}, \quad \frac{\alpha^* \beta^*}{(\alpha^* + \beta^*)^2 (\alpha^* + \beta^* + 1)} = \left(\frac{U_\varphi - L_\varphi}{2z} \right)^2.$$

De este manera, los hiperparámetros que resuelven las ecuaciones anteriores están dados por:

$$\alpha^* = \frac{z^2(2 - L_\varphi - U_\varphi)(L_\varphi + U_\varphi)^2 - (U_\varphi - L_\varphi)^2(L_\varphi + U_\varphi)}{2(U_\varphi - L_\varphi)^2}, \quad \beta^* = \alpha^* \left(\frac{2 - L_\varphi - U_\varphi}{L_\varphi + U_\varphi} \right). \quad (13)$$

Para proponer los intervalos de incertidumbre para cada uno de los parámetros del modelo SEICS, consideramos el ajuste numérico realizado previamente con el algoritmo de optimización Levenberg-Marquardt. De esta forma, se proponen los siguientes intervalos que contienen los estimadores $\hat{\beta} = 0.15012264$, $\hat{\kappa} = 0.09214103$, $\hat{\gamma} = 0.05035646$, $\hat{\varepsilon} = 0.00314205$ y $\hat{p} = 0.04609739$ encontrados con dicho ajuste numérico:

$$[L_\beta, U_\beta] = [0, 0.3], \quad [L_\kappa, U_\kappa] = [0, 0.25], \quad [L_\gamma, U_\gamma] = [0, 0.2], \quad [L_\varepsilon, U_\varepsilon] = [0, 0.05], \quad [L_p, U_p] = [0, 0.1].$$

Usando las ecuaciones (12) y (13), y un valor de $z = 2$, las distribuciones a priori a utilizar son: $\beta \sim Gamma(4, 26.66)$, $\kappa \sim Gamma(4, 32)$, $\gamma \sim Gamma(4, 39.99)$, $\varepsilon \sim Gamma(4, 159.99)$ y $p \sim Beta(3.75, 71.24)$. Puesto que $z = 2$, al menos el 88.88 % de cada distribución a priori se encuentra en el intervalo de incertidumbre correspondiente.

En la Figura 6 se muestra la función de densidad de cada una de las cinco distribuciones a priori propuestas.

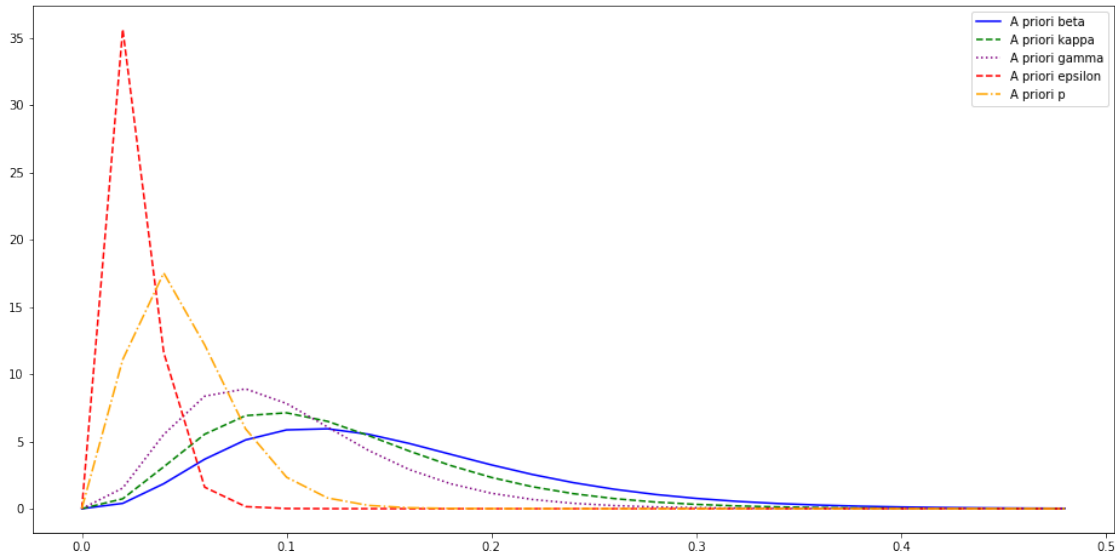


Figura 6

9.3. Simulaciones de la Distribución Posterior Usando *t-walk*

Una vez definidas las a priori y la verosimilitud, la distribución posterior es proporcional al producto de las cinco a prioris (pues se supuso que son independientes) y la verosimilitud. Para obtener simulaciones de la distribución posterior se aplicó el algoritmo tipo MCMC llamado *t-walk*, disponible para Python en la paquetería *pytwalk* creada por Christen, J.A. & Fox, C. (2010).

Se realizaron 400,000 iteraciones y se estableció un periodo de burn-in de 60,000. Del total de iteraciones, se obtuvo una tasa de aceptación de 0.01369 de los valores propuesto para θ .

Las cadenas encontradas para cada parámetro después de eliminar el periodo de burn-in, se muestran en la Figura 7, donde la primera corresponde a β , la segunda a κ , la tercera a γ , la cuarta a ε y la última a p .

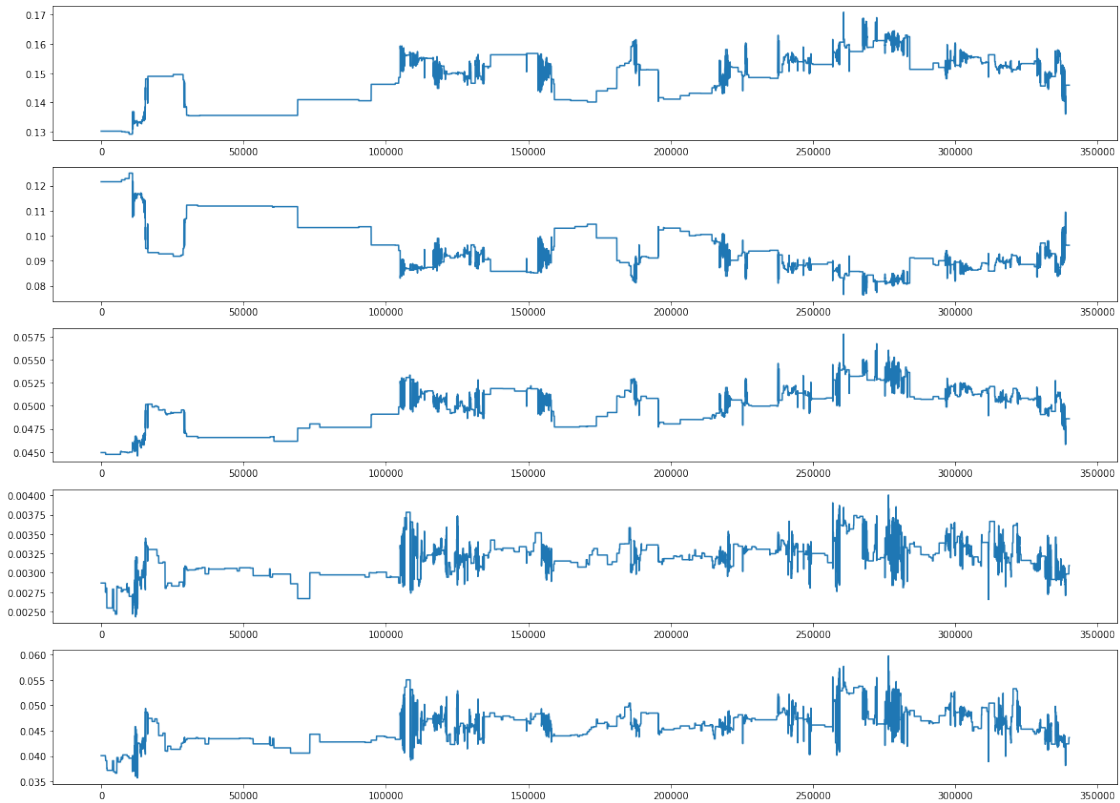


Figura 7: Cadenas obtenidas al desechar las primeras 4800 simulaciones.

Luego de explorar diferentes valores de adelgazamiento de la cadena, se observó que con todos el valor de autocorrelación seguía siendo alto de principio a fin. Por lo tanto, el criterio para elegir el valor de salto para el adelgazamiento fue el Tamaño de Muestra Efectivo (Effective Sample Size), que mide el cómo la autocorrelación dentro de las cadenas aumenta la incertidumbre en las estimaciones. En cierto sentido, se refiere a la cantidad de muestras independientes contenidas en la cadena, por lo que entre mayor sea el valor del Tamaño de Muestra Efectivo la cadena adelgazada es mejor.

De esta manera, se eligió un adelgazamiento con saltos de 20. Las cadenas resultantes se muestran en la

Figura 8, mientras que los histogramas correspondientes a las simulaciones conservadas de la distribución posterior se observan en la Figura 9.

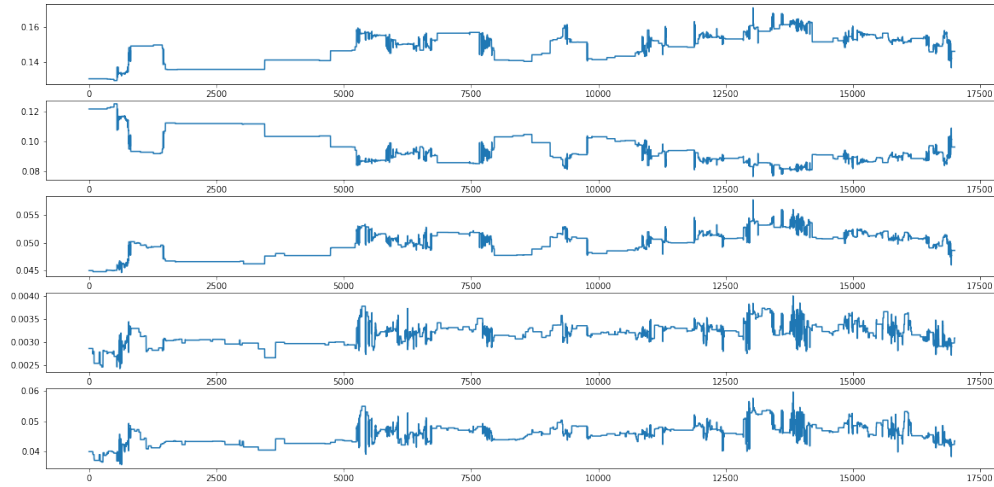


Figura 8

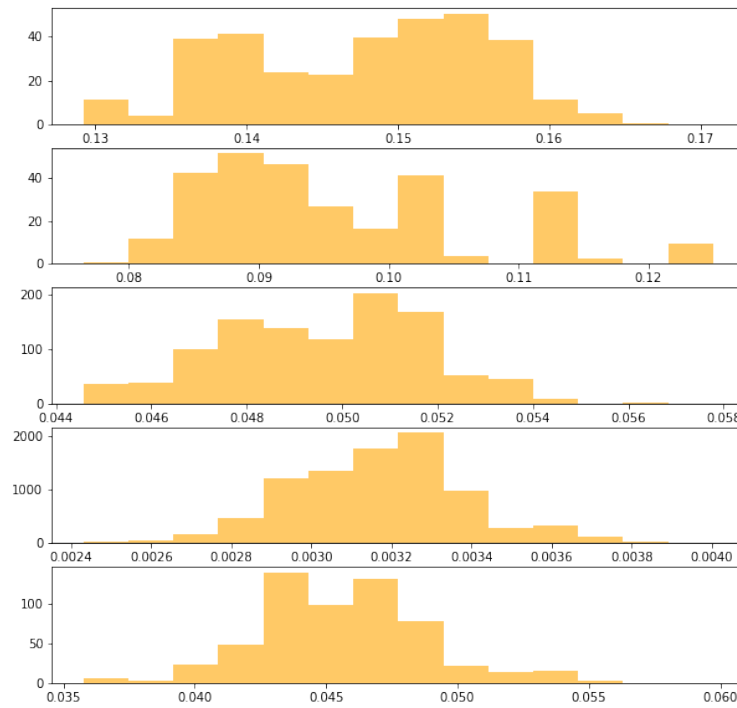


Figura 9

9.4. Intervalos de Densidad Posterior y Estimación de Parámetros

A través de las simulaciones de la distribución posterior presentadas en los histogramas de la Figura 9, podemos calcular intervalos del 90 y 95 % de densidad posterior mediante los cuantiles 0.025, 0.05, 0.95 y 0.975. Tales intervalos se muestran en la siguiente tabla:

Parámetro	Intervalo al 90 %	Intervalo al 95 %
β	[0.1354, 0.1591]	[0.1302, 0.1611]
κ	[0.0840, 0.1122]	[0.0816, 0.1215]
γ	[0.0461, 0.0530]	[0.0449, 0.0536]
ε	[0.0028, 0.0035]	[0.0026, 0.0036]
p	[0.0405, 0.0516]	[0.0399, 0.0532]

De esta tabla, se puede destacar que todos los intervalos son estrechos y que la mayoría contiene el valor real del parámetro correspondiente con el que se generaron los datos sintéticos. Los únicos que no lo contienen son los intervalos del 90 % de densidad posterior de ε y de p , donde el valor real es muy cercano a los límites inferiores de dichos intervalos.

A partir de los intervalos del 90 % de densidad posterior, se obtuvieron las curvas de incidencia resultantes de asignarle a cada parámetro diferentes valores dentro de su intervalo. Así, se crearon las bandas en color gris mostradas en la Figura 10, que cabe mencionar que contienen los datos de incidencia sintéticos a lo largo de todo el periodo considerado.

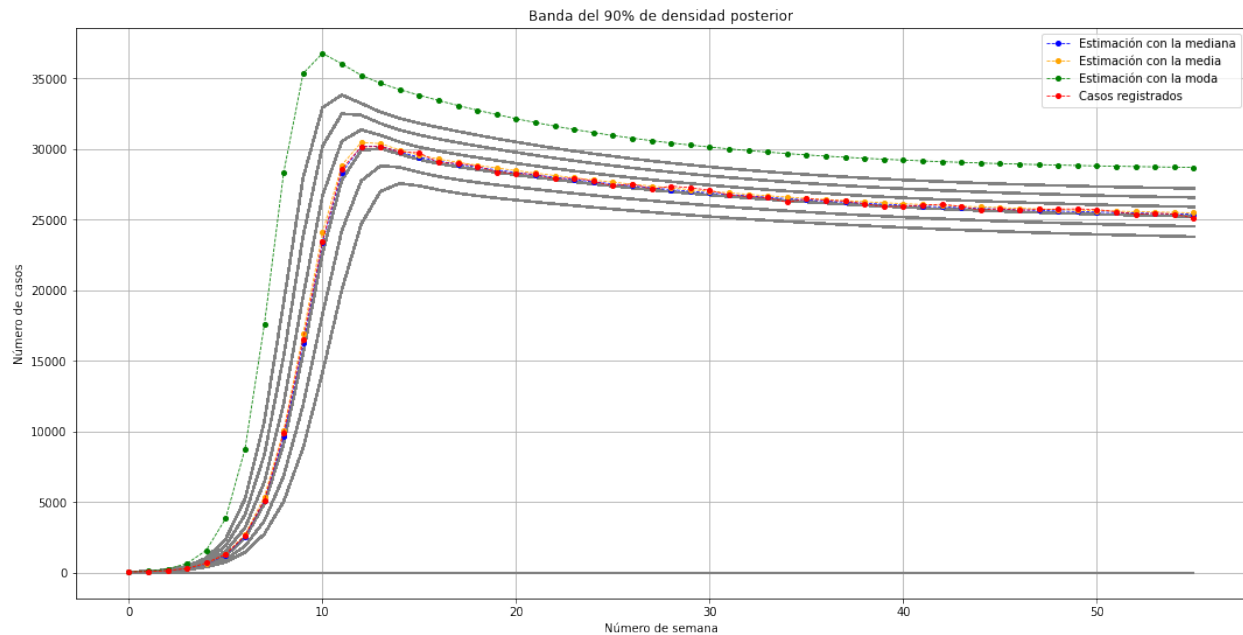


Figura 10

Por otro lado, a continuación se presentan los estimadores puntuales de cada parámetro calculados al obtener la media, la mediana y la moda de las simulaciones de la respectiva distribución marginal posterior. Además, en la tabla se comparan con el valor real del parámetro.

Parámetro	Media	Moda	Mediana	Valor real
β	0.1476	0.1708	0.1489	0.1428
κ	0.09605	0.1249	0.09273	0.1
γ	0.04962	0.05777	0.04997	0.0476
ε	0.003169	0.004004	0.003182	0.002739
p	0.04562	0.05973	0.04577	0.04

Podemos ver cómo las estimaciones de los parámetros son buenas y que, como era de esperarse, las estimaciones que más difieren del valor real son las obtenidas con la moda, pero aún así se mantienen cercanas. En la Figura 11, se muestran las curvas de incidencia usando las estimaciones de los parámetros con la media, la moda y la mediana, y comparándolas con los datos simulados.

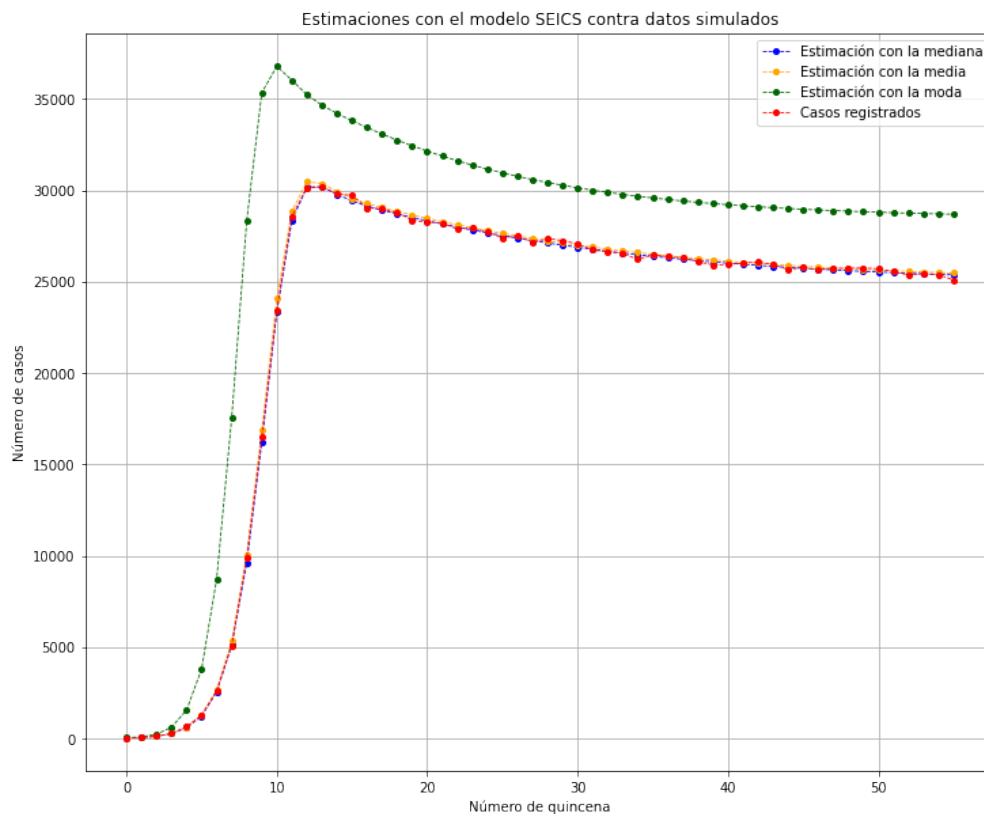


Figura 11

En la gráfica se ve claramente que las estimaciones realizadas con la media y la mediana son muy

buenas (de hecho las dos curvas correspondientes se sobreponen), mientras que la curva generada con las estimaciones hechas con la moda sí posee una diferencia significativa con los datos simulados.

El código implementado para la aplicación del *t-walk* y la obtención de los resultados expuestos sobre la inferencia bayesiana de los parámetros se encuentra en el archivo *seics_twalk_roman.py* del repositorio de GitHub correspondiente a este proyecto.

10. Ventajas y Desventajas del Modelo SEICS

Comparando el modelo SEICS con algunos modelos recientes de la fiebre tifoidea se tiene el modelo presentado en [10], cuya principal contribución es **incorporar la dinámica de transmisión indirecta** (por contacto con medios contaminados) al añadir el compartimento W para la población bacteriana. Asimismo, se tiene el modelo propuesto en el artículo [13], el cual comparte muchas características con el modelo de [10] y que también agrega un compartimento para la **población bacteriana** que denomina B_C . No obstante, su principal aportación es que introduce una tasa de ingesta de comida o agua y una **probabilidad de que se consuma comida o agua contaminada** con la bacteria causante de fiebre tifoidea.

Esto es de suma relevancia puesto que gran parte de las infecciones por fiebre tifoidea se originan por el contacto con bacterias ambientales cuando se ingiere agua o comida contaminada. Incluso en el mismo artículo [10] se resalta que en ningún modelo que antecede a su propuesta se consideran ambas formas de transmisión. De esta forma, los compartimentos W en [10] y B_C en [13] representan una gran ventaja respecto al modelo SEICS propuesto en el presente proyecto, ya que en (1) sólo se consideran las infecciones causadas por interacciones directas entre individuos.

No obstante, el modelo SEICS tiene varios puntos a favor. Primero, a diferencia de [10] y [13] **sí incluye una categoría de expuestos**, lo cual resulta importante pues el periodo de latencia sí es significativo [7].

Por otro lado, en los dos modelos mencionados que se están tomando como punto de comparación se consideran los portadores asintomáticos de distinta manera a cómo se hace en el modelo SEICS. Los autores suponen que una vez que una persona es contagiada puede convertirse en infectado (I) o en portador (I_C en [10] y C en [13]). Sin embargo, hechos históricos como el caso de Mary Typhoid [2] y estudios realizados como los reportados en [12], han mostrado que los individuos que son una fuente importante de nuevas infecciones son aquellos que se convierten en **portadores asintomáticos una vez que se recuperaron** y no aquellos que desde que se infectaron son asintomáticos.

Otro aspecto a destacar es que en el modelo SEICS sí se presenta un **punto de equilibrio endémico**. Esto parece tener coherencia con el comportamiento que ha manifestado la fiebre tifoidea a lo largo de los años, ya que a pesar de que transcurrió más de un siglo desde que la bacteria causante fue aislada, la enfermedad ha permanecido y sigue ocasionando numerosos brotes en diferentes partes del mundo.

En el caso de [13] el modelo también posee un punto de equilibrio endémico debido a que sí se consideran transiciones entre el estado recuperado y susceptible, suponiendo que la inmunidad es temporal. En contraste, en [10] no se habla de los puntos de equilibrio del modelo, y por las simulaciones mostradas parece que no posee un equilibrio endémico. De hecho, en el modelo de [10] se toma como supuesto que una vez recuperado se adquiere una inmunidad total, lo cual se ha mostrado que no es verdad y las reinfecciones son bastante comunes [5].

De esta manera, el modelo SEICS sí considera esta característica de la dinámica de la enfermedad y se toma como supuesto que **no hay estado de inmunidad temporal** ya que, en general, el haber estado

enfermo no te protege de infecciones futuras [5].

Puesto que el periodo que un individuo puede durar como portador asintomático puede abarcar varios años, el suponer que la **población es fija**, como se hizo en el modelo SEICS, es una **limitante** para poder capturar en forma apropiada la dinámica poblacional real en periodos de tiempo largos. En cambio, tanto el modelo de [10] como el de [13] consideran una tasa de natalidad humana y también tasas de mortalidad natural humana y bacteriana. Adicionalmente, en [13] se incluye una tasa de mortalidad a causa de la fiebre tifoidea.

Cabe mencionar que al incluir estas características de la dinámica poblacional cada uno de estos dos modelos contemplan más de diez parámetros en total, mientras que en el modelo SEICS sólo se consideran cinco. Esto resulta importante al momento de efectuar la estimación de parámetros mediante métodos MCMC, pues no sólo se deben generar y evaluar propuestas para cada parámetro sino que el **costo computacional** de evaluar el problema directo aumenta conforme **más parámetros** se vayan a estimar. Por lo tanto, siempre es preferible dejar el número de parámetros tan pequeño como sea suficiente para describir adecuadamente la dinámica de la epidemia.

Por otra parte, en [9] se presenta una versión extendida del modelo de [10], mientras que en el caso de [13] ahí mismo se hace una extensión del modelo correspondiente, ambas con el propósito de incorporar el efecto de **estrategias de control epidemiológico** sobre la difusión de la fiebre tifoidea. Estas versiones extendidas de los modelos representan una herramienta muy útil para la evaluación y el planteamiento de escenarios que permita llevar a cabo estrategias de control epidemiológico pertinentes, que reduzcan al máximo el número total de infectados durante el brote.

Así, representan una gran ventaja respecto del modelo propuesto SEICS, ya que para que el modelo sea más útil en la toma de decisiones en salud pública sí es importante capturar el impacto que las diferentes medidas de prevención tienen sobre el desarrollo de la epidemia.

Por último, en los artículos [3] y [8] se propone un modelo en el que uno de los compartimentos corresponde a los individuos **vacunados**, suponiendo que la vacuna confiere una inmunidad temporal. De este modo, se introduce una tasa de transición de vacunados a susceptibles. De acuerdo con [1] las campañas de vacunación en países en desarrollo han sido de gran relevancia para combatir la difusión de la fiebre tifoidea, pero aún se está trabajando en desarrollar mejores vacunas y que sean capaces de proteger a la población de diversas y nuevas variantes de la *Salmonella typhi*. De aquí la importancia de incluir la vacunación a los modelos de fiebre tifoidea, lo cual no se hizo en el modelo SEICS.

Respecto a dichos modelos presentados en [3] y [8], las principales ventajas que posee el modelo SEICS son la inclusión del compartimento de expuestos y la forma en que considera el compartimento de portadores asintomáticos C como aquellos que lo son una vez que se recuperaron y no aquellos que lo son desde el momento en que se contagian. De hecho, en el caso de [8] ningún compartimento corresponde a los portadores asintomáticos, los cuales han mostrado ser una fuente importante de nuevas infecciones [12][2].

11. ¿Cómo se Puede Mejorar el Modelo?

A diferencia de enfermedades como la influenza o la COVID-19 para las que existen reportes diarios o semanales, el gobierno mexicano suele reportar la incidencia anual de la fiebre tifoidea, tal y como se puede consultar en [11]. A pesar de que la solución del modelo mejoró aumentando el periodo de incidencia, el contar con un número de datos tan limitado como lo son los datos anuales no sería suficiente para encontrar

un buen ajuste o un ajuste que realmente sea informativo respecto a la propagación de la enfermedad.

Otro reto importante en la recolección de datos, es que es difícil tener una buena aproximación del número de portadores asintomáticos de la bacteria *Salmonella typhi* dentro de la población, precisamente porque no manifiestan síntomas. Esto se podría lograr si se le diera un seguimiento a todos los recuperados de fiebre tifoidea por al menos un año, para aplicarles la prueba correspondiente y detectar si son portadores o no.

Ahora bien, respecto a las limitaciones del modelo, se han detectado diferentes variantes importantes de la bacteria *Salmonella typhi* [12], las cuales tienen asociadas tasas de transmisión y recuperación distintas. Por lo tanto, se podrían añadir más compartimentos y parámetros para hacer la distinción entre infectados causados por una variante u otra.

Por otro lado, como los periodos de tiempo estudiados a través del modelo son grandes para poder capturar la transición de los portadores asintomáticos a susceptibles, la población puede variar significativamente. De este modo, sería útil suponer que la población no es cerrada y añadir tasas de nacimiento y de muerte natural al modelo, como se propone en [10], [13], [8] y [3].

Asimismo, se podría extender el modelo incorporando un parámetro que indique la disminución de susceptibles acorde al impacto de las medidas de prevención sanitarias como se plantea en [9] y [13], aunque también se podría capturar esto en gran medida al reducir la tasa de transmisión.

Lo que representaría una mejora significativa en el modelo sería la inclusión de un compartimento que describa la concentración y el crecimiento de la bacteria en el ambiente (principalmente en alimentos y en contenedores de agua), tal y como se propone en [10] y en [13]. Así, se podrían contabilizar todas las infecciones causadas por agua y alimentos contaminados.

Para finalizar, como actualmente en muchos lugares se realizan campañas de vacunación efectivas contra la fiebre tifoidea [1], sería útil agregar un parámetro de vacunación que reduzca el número de susceptibles y crear un compartimento V que contenga a todos aquellos individuos a los que se les ha conferido una inmunidad temporal a través de la vacuna. Esto con base en [3] y [8], pues se ha mostrado que la inmunidad no es permanente y se recomienda aplicarse una dosis de refuerzo cada dos años [1].

12. Repositorio del Proyecto en GitHub

Se puede consultar todo el código utilizado para el desarrollo de este proyecto en la liga que se muestra a continuación, en donde se incluyen los archivos de código Python mencionados previamente en este proyecto:

- https://github.com/romanzml3/Modelo_SEICS_Roman

Referencias

- [1] Ali, M., Yahaya, A., Nas, F., Anas, A., & Ibrahim, I. (2017). Trends in Treatment and Vaccine Development of Typhoid Fever: A Review. *Asian Journal of Biotechnology and Bioresource Technology*, 2(4), 1-7.
- [2] Brooks, J. (1996). The sad and tragic life of Typhoid Mary. *CMAJ: Canadian Medical Association Journal*, 154(6), 915.

- [3] Edward, S., & Nyerere, N. (2016). Modelling typhoid fever with education, vaccination and treatment. *Eng. Math*, 1(1), 44-52.
- [4] Gutiérrez Pulido, H., & Gutiérrez González, P. (2013). Fundamentos y aplicaciones de la estadística Bayesiana. *México: Universidad de Guadalajara*.
- [5] Im, J., Islam, M. T., Kim, D. R., Ahmmed, F., Chon, Y., Zaman, K., ... & Clemens, J. D. (2020). Protection conferred by typhoid fever against recurrent typhoid fever in urban Kolkata. *PLoS neglected tropical diseases*, 14(8), e0008530.
- [6] Paul, U. K. (2017). Typhoid Fever–Recent Management. *API-Medicine Updat*, 64-7.
- [7] Paul, U. K., & Bandyopadhyay, A. (2017). Typhoid fever: a review. *Int J Adv Med*, 4(2), 300-306.
- [8] Peter, O. J., Ibrahim, M. O., Akinduko, O. B., & Rabi, M. (2017). Mathematical model for the control of typhoid fever. *IOSR J. Math*, 13(4), 60-66.
- [9] Peter, O. J., Ibrahim, M. O., Edogbanya, H. O., Oguntolu, F. A., Oshinubi, K., Ibrahim, A. A., ... & Lawal, J. O. (2021). Direct and indirect transmission of typhoid fever model with optimal control. *Results in Physics*, 27, 104463.
- [10] Peter, O. J., Ibrahim, M. O., Oguntolu, F. A., Akinduko, O. B., & Akinyemi, S. T. (2018). Direct and indirect transmission dynamics of typhoid fever model by differential transform method.
- [11] SEMARNAT (5 de enero de 2022). *Casos nuevos reportados de Fiebre Tifoidea (Número de casos)*. http://dgeiawf.semarnat.gob.mx:8080/ibi_apps/WFServlet?IBIF_ex=D1_SAMBIENTAL01_11&IBIC_user=dgeia_mce&IBIC_pass=dgeia_mce&NOMBREENTIDAD=* &NOMBREANIO=*
- [12] Senthilkumar, B., Senbagam, D., & Rajasekarapandian, M. (2014). An epidemiological surveillance of asymptomatic typhoid carriers associated in respect to socioeconomic status in India. *Journal of Public Health*, 22(3), 297-301.
- [13] Tilahun, G. T., Makinde, O. D., & Malonza, D. (2017). Modelling and optimal control of typhoid fever disease with cost-effective strategies. *Computational and mathematical methods in medicine*, 2017.