

# A Hidden Markov Model for Surveilling the Transmission of Lassa Fever Epidemic Disease

## Abstract

Lassa fever is an infectious viral disease that is endemic in Nigeria and other West African countries. Early detection and response to outbreaks of the disease are critical to prevent its spread and reduce illness and death. Finding mathematical patterns that explain the mechanisms of Lassa fever transmission, as well as a thorough understanding of the biological factors affecting the disease, are necessary in putting in place a surveillance system and preventing further disease spread. In this study, we develop a Hidden Markov Model (HMM) approach to surveil the transmission of Lassa fever virus in Nigeria. The HMM model was developed using the **SIR model** to formulate the transition matrix and data from past outbreaks of the disease to compute the observations. Our results showed the dry season as the peak period for Lassa fever and records its lowest numbers during the rainy season. The transition matrix showed a 98% chance of transitioning to the infected state from being susceptible and a 96% chance of remaining infected. The stable probability resulted in a 97.9% probability of transitioning to the infected state and a 1.7% chance of transitioning to the susceptible state. The Empirical analysis using the proposed HMM approach does not only provide a valuable tool for public health officials to track and respond to outbreaks of Lassa fever, leading to more effective disease control strategies but also, establishes an efficient structure for other infectious diseases modeling to aid in early detection and response to outbreaks.

**Keywords:** Lassa fever, Hidden Markov model, infectious diseases, transition probability matrix, emission matrix, Viterbi algorithm

## 1. Introduction

Lassa fever is a viral illness spread by infected rodents known as multimammate rats (*Mastomys natalensis*). The main transmission of the virus is through person-to-person transmission, direct or indirect contact with food or objects contaminated with urine or faeces of infected multimammate rats (**Ibrahim and Denes (2021), Hamblion, et al. (2018)**). These rats are abundant in rural areas of parts of some West African countries, including Nigeria, Sierra Leone, Liberia, and Guinea (WHO, 2002). The low standards of living and poor sanitation are one reason this virus is dominant in these rural communities.

Consequently, Lassa fever has been endemic in these West African countries where the multimammate rats are present. The number of infections per year of Lassa fever is estimated between 100,000 and 300,000, with approximately 1% leading to death (Ibrahim and Denes (2021)).

In Nigeria, Lassa fever is endemic which is typically detected during the dry season (December–April) with hundreds of laboratory-confirmed cases per month. Approximately 90–95% of these persons are infected by direct or indirect contact with household items contaminated by the excretions of infected *Mastomys* rats. (WHO, 2022). Numerous preventive and control tactics

have been recommended by the World Health Organization (WHO) and the Nigerian Government, through the Centre for Disease Control (NCDC), for the elimination of Lassa fever outbreaks from endemic communities such as Nigeria. Preventive measures such as improved sanitation, proper storage of foodstuffs, environmental sanitation, and keeping cats can reduce the ailment. Animal products should be meticulously cooked, early detection and proper treatment of the virus. Seclusion of infected persons and provision of treatment reduces the risk of person-to-person transmission in health care facilities.

Dissemination of guidelines for proper case management, and infection prevention and control (IPC); Enhanced surveillance activities for Lassa fever-affected areas to increase discovery of cases; Provision of special treatment centres for clinical management in affected areas; Increased laboratory capacity to ensure timely processing of samples; Risk communications and community engagement activities through television, radio, print, social media, and other strategies are strongly endorsed to lessen the fatality rate.

Mathematical modeling is a theoretical approach that has been used extensively and successfully to study the dynamics and control of infectious diseases like Lassa fever. These models, the one considered here, can assist in existing measures being more effective. Though Lassa fever appears in WHO's lists of prioritizing diseases for research and development in emergency contexts, only a few theoretical studies are available. Some findings on Lassa fever using mathematical models are summarized below.

## **2. Related Literature**

Numerous mathematical modeling with their methodologies and findings have been conducted to enlighten and provide more information on the transmission dynamics and to control the endemic disease (see Onah and Collins, (2020), Bakare et al., (2020)).

A multiple-patch model was developed by Onah and Collins, (2020) to examine the effects of socioeconomic class on Lassa fever. The author showed a sensitivity analysis and a numerical illustration of the effect of parameter models for spread of disease and incidence. Their results reveal that humans' socioeconomic status has a significant influence on the dynamics of LF transmission. The study recommends that human socioeconomic classes be considered in order to attain complete LF elimination in communities where the virus is dominant. A study titled "Evaluation of rodent control to fight Lassa fever based on field data and mathematical modelling" was presented in (Marien, et al., (2019)). In order to eradicate LF in rural areas, the authors used a mathematical model to experiment numerous control approaches in rural upper Guinea to determine how long and frequently control strategies should be performed. The mathematical model suggests that the best strategy for eradicating LF is continuous control or rodent vaccination. A spatial analysis of Lassa fever data from human cases and infected rodents from 1965 to 2007 was performed in Fichet-Calvet and Rogers (2009), to describe the LF risk maps in West Africa. The authors researched on the impact of environmental variables that are extrinsic such as temperature, vegetation, and rainfall on the transmission dynamics of LF in Cameroon. The study, shows that rainfall has strong effect in defining high risk areas, while temperature has a lesser effect in high-risk areas. Additionally, the risk maps revealed that the most unsafe region is located between Guinea and Cameroon.

By using a mathematical model, Ojo, et al (2021) discovered that any control strategy that reduces rodent populations and the risk of transmission from rodents to humans will assist in achieving Lassa fever elimination in Nigeria. Musa et al. (2020) used mechanistic modelling that takes into account quarantine, isolation, and hospitalization processes of Lassa fever epidemics in Nigeria from 2016–2019. Particularly, the similarities in the transmission dynamics driving three major Lassa fever outbreaks from 2016–2019 in Nigeria were outlined by their study. Ndenda et al. (2022) used fractional-order dynamic modelling to study the effect of environmental viral load, interpersonal contact, and infected rodents on Lassa fever transmission dynamics. They discovered that with multiple interventions and control measures such as environmental sanitation, can help in eradicating infections significantly. Marien et al. (2019) applied a mathematical model to study the impact of rodents' control to fight Lassa fever. Their study shows that rodent vaccination is an approach that could eliminate Lassa virus. Abdulhamid et al. (2022) used a deterministic mathematical model to study Lassa fever dynamics. Their study reveals that the existence of backward bifurcation in the model makes the control of Lassa fever more difficult. Zhao et al. (2020) used a mathematical model to investigate the effects of rainfall on Lassa fever epidemics in Nigeria by quantifying the association between reproduction number and rainfall for several locations in Nigeria.

Innocent and Omo (2017) developed a mathematical model for investigating the dynamics of the disease, and recommended avoiding contact with species that carried viruses and introducing vaccines for humans. Akinpelu and Akinwande (2018) developed a mathematical model for sensitivity analysis of Lassa fever. This model is divided into five compartments of susceptible (S), latent (L), infected (I), isolated (I), and recovered (R) by using the next-generation method and obtained  $R_0$ , showing that disease-free equilibrium was locally and globally asymptotically stable. James et al., (2015) analyzed stability for Lassa fever and recommended quarantines and making strategies for permanent immunity. Bakare et al. (2020) worked on the transmission dynamics of the disease and derived a nonlinear ordinary differential equation by introducing the seasonal parameters.

Salim (2020) studied a Markov chain in place of a time series of lung TB infections. To set the problem according to the Markov chain, he made hypotheses regarding the number of infections and identified the chain that turned out to be non-ergodic. He then calculated the expected absorbing time and probability for the transient state.

Adigun, et al. (2019) examined the environmental elements that contribute to the transmission of infectious diseases. The study calculated the diseases' patterns of transition, testing the Markovian property and how stationary the process is over the study period, the research concluded that the past history of infectious diseases will have an impact on the present through the current condition and recommended that more needs to be done by the government in the areas of sensitization and the fight against infectious diseases.

Also, Inegbedion (2022) in an attempt to estimate the population's susceptibility to the COVID-19 pandemic as well as the proportions of infections, recoveries, and fatalities that will result from it, designed a longitudinal study of COVID-19 using information from the daily updates of the NCDC for the period of 1 May to 23 August 2020. On the data, a Markov chain analysis was done and the findings shows that over the long term, 8.4% of the population will be at risk for

COVID-19 infections, 26.4% of them will become infected, 61.2% of those who become infected will recover, and 4% of those who become infected will die.

Nkemnole and Osunkeye (2016) investigated the prevalence of endemic diseases in a population and their duration of resistance using a stochastic model like the Markov chain, which has a continuous time and discrete state space and necessitates the Monte Carlo simulation to produce the desired results. They evaluated a few areas of active simulation research in the health sector, including the impact of gender on the typical number of days a disease lasts. Findings of the empirical analysis revealed a significant correlation between gender and the duration of endemic disease persistence in a population. Nkemnole and Udoh (2022) built the transition and emission probabilities of Covid-19 cases and created a hidden Markov model. The outcome revealed that between the study periods, the estimated case fatality rate for Covid-19 in Lagos State was 4.35%. Moving from an infected state to one of recovery has a 25% chance, moving from an infected state to death has a 50% chance, and the chance of still being infected has a 25% chance.

These studies have made significant progress in the dynamics and control of Lassa fever. Nevertheless, finding mathematical patterns that explain the mechanisms of Lassa fever transmission, as well as incorporating multiple control measures together with real data to study and make predictions of the possible future dynamics of the Lassa fever epidemic in Nigeria are necessary in putting in place a surveillance system and preventing further disease spread. As a result, it is important to examine the trends of the seasonality of Lassa fever and employ the Hidden Markov Model to model the occurrence of Lassa fever in Nigeria.

The findings from this study should aid both researchers and policymakers in developing better control strategies for effective control and management of seasonal Lassa fever outbreaks in endemic areas.

The remaining of the article is organized as follows: In Section 3, we present the formulation of the Lassa fever HMM model, performed a mathematical analysis on the formulated model, data fitting and parameter estimation. Numerical results and discussion of the analytical findings and the study conclusion are presented in Sections 4 and 5 respectively.

### **3. Methods**

#### **3.1 Epidemiology**

Epidemiology is the study of how diseases and their causes are distributed throughout populations. Understanding the trends, causes, and effects of health and disease in populations as well as creating and assessing interventions to promote health and prevent disease make up this important field of public health (WHO, 2021). Some of the epidemiological models are SIR (Susceptible-Infected-Recovered), SEIR (Susceptible Exposed-Infected-Recovered) and SEIRD (Susceptible-Exposed-Infectious-Recovered-Deceased). This study discusses the SIR model as the hidden state for developing the HMM for Surveilling the Transmission of Lassa Fever Virus

#### **Epidemic Surveillance**

Epidemic surveillance is a process of collecting, analyzing, and disseminating data to detect and respond to outbreaks or other public health events. A surveillance system can be classified into two main types: passive and active.

Passive surveillance systems rely on the reporting of health events by healthcare providers, laboratories, or individuals. These systems are useful for detecting outbreaks after they have started, but they have limitations in detecting outbreaks in real-time. Active surveillance systems, seek out cases of a disease or health event by actively searching for cases in the community. This type of surveillance is typically more resource-intensive but can detect outbreaks earlier and provide more detailed information about the outbreak. There are also hybrid systems that combine elements of passive and active surveillance, taking advantage of the strengths of both types (Petersen & Fullerton, 2018).

Surveillance systems for Lassa fever typically collect data on the incidence and prevalence of Lassa fever cases, as well as the demographics, comorbidities, and treatment trends. This data is then analyzed to detect trends, patterns, and outbreaks, and to inform public health action (WHO, 2017). Surveillance systems for Lassa fever can be electronic such as electronic medical records or manual such as paper-based reporting and can be implemented at various levels of the health care system, including the national, regional, and local levels

### 3.1.1 SIR model:

The SIR model is a classical mathematical model used in epidemiology to study the spread of infectious diseases (Hethcote, 2000). It stands for Susceptible-Infected-Recovered and represents the three distinct populations in a population with a disease with Susceptible (S) being individuals who are susceptible to the disease but have not yet contracted it, Infected (I) referring to individuals who have contracted the disease and are capable of spreading it to susceptible individuals and Recovered (R) for individuals who have either recovered from the disease or died from it and are no longer capable of spreading the disease.

The model is typically described by a set of ordinary differential equations (ODEs), which represent the rate of change of the size of each population over time:

$$dS/dt = -\beta S(t)I(t) \quad (1)$$

$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t) \quad (2)$$

$$\frac{dR}{dt} = \gamma I(t) \quad (3)$$

where  $\beta$  is the transmission rate and  $\gamma$  is the recovery rate. The solution to these equations gives the evolution of the number of individuals in each state over time (William & McKendrick, 1927).

The formulation of the SIR model for Lassa fever transmission between states were established. The transmission states in the SIR model were used to generate the states of a hidden Markov model. The Viterbi algorithm and Expectation-Maximization algorithms were used respectively to determine the most likely hidden sequences of the states and estimate the model's parameters.

### 3.1.2 SIR Model for the Transmission of Lassa Fever in Nigeria

The total human population at time  $t$ , denoted by  $N_h(t)$  is further divided into susceptible ( $S_h$ ), infectious ( $I_h$ ) and removed ( $R_h$ ). Hence, the total human population at time  $t$  is given as:

$$N_h(t) = S_h + I_h + R_h \quad (4)$$

The susceptible rodent population at any time  $t$  is represented as  $S_r(t)$  and the infected rodent population at time  $t$  is given as  $I_r(t)$ ; the total rodent population at time  $t$  is given as:

$$N_r(t) = S_r(t) + I_r(t) \quad (5)$$

According to their disease status, each subpopulation's progression from one class to another is modeled. The recruitment rate,  $\Lambda_h$ , populates the susceptible human population through birth or immigration and from recovered subpopulation due to  $\gamma$  rate of immunity loss (Abioye et al., 2020). The dynamics of Lassa fever in the population are described by the deterministic system of nonlinear differential equations given as:

$$\frac{dS_h(t)}{dt} = \Lambda_h - \frac{\alpha_1 \alpha_2 S_h(t) I_r(t)}{N_h} + \gamma R_h(t) + \tau_{nc} I_h(t) - \mu_h S_h(t), \quad t \geq 0 \quad (6)$$

$$\frac{dI_h(t)}{dt} = \frac{\alpha_1 \alpha_2 S_h(t) I_r(t)}{N_h} - \tau_c I_h(t) - r_c I_h(t) - \tau_{nc} I_h(t) - \delta I_h(t) - \mu_h I_h(t), \quad t \geq 0 \quad (7)$$

$$\frac{dR_h(t)}{dt} = \tau_c I_h(t) + r_c I_h(t) - \gamma R_h(t) - \mu_h R_h(t), \quad t \geq 0 \quad (8)$$

$$\frac{dS_r(t)}{dt} = \Lambda_r - \frac{\alpha_1 \alpha_3 S_r(t) I_h(t)}{N_h} - \mu_r S_r(t), \quad t \geq 0 \quad (9)$$

$$\frac{dI_r(t)}{dt} = \frac{\alpha_1 \alpha_3 S_r(t) I_h(t)}{N_h} - \mu_r I_r(t), \quad t \geq 0 \quad (10)$$

where  $S_h(0) \geq 0, I_h(0) \geq 0, R_h(0) \geq 0, S_r(0) \geq 0, I_r(0) \geq 0$ .

The model parameters are defined as follows:

**Table 1: The parameter definitions of the SIR model**

Parameters	Description
$N_h$	Number of human populations
$\alpha_1$	Infection rate of rodents
$\alpha_2$	The force of infection
$\alpha_3$	Infection rate of human interact
$\tau_c$	The connection of humans with drugs
$\tau_{nc}$	The rate at which humans do not have a relationship with drugs
$r_c$	Rate of awareness
$\gamma$	Loss of immunity
$\Lambda_h$	Birthrate of humans
$\mu_h$	Deathrate of humans
$\delta$	Rate of mortality of an infectious class
$\Lambda_r$	The birthrate of rodents
$\mu_r$	The deathrate of rodents

Some of the model's characteristics must be investigated in order for it to accurately represent the transmission of Lassa fever; few of those characteristics are given below.

### 3.1.3 Transition Probabilities from the SIR Model

Given the earlier defined SIR model, the transition probabilities from one hidden state to another of transmission of Lassa fever can be estimated using the model parameters. In the SIR model, the assumption is that transitions are made from the suspected state to the infected state, from the infected state to recovery and from recovery to susceptible (Hethcote, 2000).

The probability of transitioning to infected state is given as:

$$p_{(s-1, i+1) \leftarrow (s, i)}(\Delta t) = \frac{\beta si}{N} \Delta t \quad (11)$$

The probability to make the transition from the infectious class to recovered class is :

$$p_{(s, i-1) \leftarrow (s, i)}(\Delta t) = \gamma i \Delta t \quad (12)$$

Since a constant population is assumed, the probability that the number of infectious remains unchanged after a time step is:

$$p_{(s, i) \leftarrow (s, i)}(\Delta t) = 1 - \left[ \frac{\beta si}{N} - \gamma i \right] \Delta t \quad (13)$$

where  $\beta$  and  $\gamma$  are transmission and recovery rate respectively.

The probabilities involved in transitions are then estimated for each time step  $t$  using the values of the rate parameters  $\beta$  and  $\gamma$ . The probabilities gotten makes up a Markov chain transition matrix containing the transition from one hidden state to another.

## 3.2 Hidden Markov Model (HMM)

The HMM is composed of two sets of variables: the hidden states and the observations. The hidden states are the underlying variables that define the internal state of the system, and the observations are the variables that are directly observable. The hidden states and the observations are related through a set of probabilistic relationships, and the aim of an HMM is to estimate the hidden states given a sequence of observations.

### 3.2.1 Assumption of Hidden Markov Model

The HMM is based on three major assumptions:

**Markov Assumption:** The current hidden state  $X_t$  depends solely upon the previous state of the hidden variable i.e.

$$\Pr[X_t = x_t | X_{t-1} = x_{t-1}, O_t = o_t] = \Pr[X_t = x_t | X_{t-1} = x_{t-1}] \quad \forall x_t \in \mathbb{X} \quad (14)$$

**Output Independence:** The current observed state  $O_t$  depends solely upon the current state of the unobserved variable, i.e

$$\Pr[O_t = o_t | X_t = x_t, O_{t-1} = o_{t-1}] = \Pr[O_t = o_t | X_t = x_t] \quad \forall o_t \in \mathbb{O} \quad (15)$$

**Stationarity:** The transition probabilities are independent of time, i.e.  $\forall t \geq 0, s \in 0, \dots, t$

$$\Pr[X_t = j | X_{t-1} = i] = \Pr[X_{t+s-1} = i] \quad \forall i, j \in \mathbb{X} \quad (16)$$

### 3.2.1 Parameter Estimation

Parameter estimation in HMM involves finding the values of the parameters (i.e. transition probabilities, emission probabilities, initial state probabilities) that maximize the likelihood of

observing the data given the model. The parameter estimation is to find the values that best describe the underlying process that generated the data. The parameter estimates obtained from this process are then used to make predictions about future events.

The Baum-Welch algorithm, a special case of the Expectation-Maximization (EM) algorithm, is one method for estimating the parameters of an HMM. The Baum-Welch algorithm for the estimation of the parameters of a HMM are described as follows:

(i) Set the initial values for the transition probability matrix  $A$  and the emission probability matrix  $B$ . The initial values for the state probabilities ( $\pi$ ) are calculated based on the number of hidden states.

(ii) Compute the forward probability matrix  $\alpha$  and the backward probability matrix  $\beta$  using the forward algorithm and the backward algorithm.

(iii) For each state  $i$  and each observation  $t$ , calculate the intermediate quantity  $\gamma(i, t)$  using the formula:

$$\gamma(i, t) = \frac{\alpha(i, t)\beta(i, t)}{\sum_j \alpha(j, t)\beta(j, t)} \quad (17)$$

(iv) Re-estimate the transmission probabilities  $a(i, j)$  using:

$$a(i, j) = \frac{\sum_t \gamma(i, t)a(i, j)b(j, y(t+1))\beta(j, t+1)}{\sum_t \gamma(i, t)} \quad (18)$$

(v) Re-estimate the emission probabilities  $b(i, k)$ :

$$b(i, k) = \frac{\sum_t \gamma(i, t)I(y(t)=k)}{\sum_t \gamma(i, t)} \quad (19)$$

Repeat steps (ii) to (v) until the difference between the previous iteration and the current iteration is small or until a maximum number of iterations is reached.

### 3.2.2 Estimating the stages of the virus with HMM

With the SIR model already defined, the HMM will be used to estimate the unobserved infection stages of the virus based on the trend of confirmed cases recorded per week. The three classes of the SIR model were used as the hidden states of the model built.

The vector  $S = \{s_1, s_2, s_3\}$  represents the hidden states of the model where  $s_1$  is the susceptible state,  $s_2$  is the infected state and  $s_3$  is the removed state. Likewise,  $Q = \{q_1, q_2, q_3\}$  is the set of observations with  $q_1$  representing increasing number of confirmed cases (IN),  $q_2$  for steady number of confirmed cases (ST) and  $q_3$  for decreasing number of confirmed cases (DC).

#### 3.2.2.1 Transition and Emission Matrix

$A$  is the transition probability matrix of  $3 \times 3$  dimension. It stores the probability  $a_{ij}$  of transitioning from one of the hidden states  $i$  known to be either susceptible, infected or recovered to another one of the state  $j$ .

$$\begin{matrix} & \mathbf{S} & \mathbf{I} & \mathbf{R} \\ \mathbf{S} & a_{11} & a_{12} & a_{13} \\ \mathbf{I} & a_{21} & a_{22} & a_{23} \\ \mathbf{R} & a_{31} & a_{32} & a_{33} \end{matrix}$$

$$a_{ij} = P(S_{t+1} = s' | S_t = s), \forall s', s \in S \quad (20)$$

The emission probability matrix  $B$  is also  $3 \times 3$  which contains the probability  $b_{jk}$  of having any of the decreasing, increasing or steady observation  $k$  given any hidden state  $j$ .

$$\begin{matrix} \mathbf{S} \\ \mathbf{I} \\ \mathbf{R} \end{matrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$b_{jk} = P(Q_t = q | S_t = s) \forall s \in S, q \in Q \quad (21)$$

The initial probability vector  $\pi = \{\pi_1, \pi_2, \pi_3\}$  represents the probability of starting in from the susceptible, infected or recovered state of Lassa fever. The hidden Markov model  $\theta$  built from the three parameters stated is then given as:

$$\theta = \{\pi, A, B\} \quad (22)$$

The hidden Markov model  $\theta = \{\pi, A, B\}$  is then used to solve two problems:

- 1. Finding the Hidden Sequence:** The transition matrix  $A$ , emission matrix  $B$  and a sequence of given observations  $Q = (q_1, q_2, q_3, \dots, q_T)$  is used to find the hidden sequence  $S = (s_1, s_2, \dots, s_t)$  that is most likely to generate  $Q = (q_1, q_2, q_3, \dots, q_T)$ , i.e

$$S^* = \operatorname{argmax}_S P(Q|S, A, B) \quad (23)$$

- 2. Estimating the Parameters:** The transition matrix  $A$  and emission matrix which are the parameters of the models are estimated. The estimated parameters represent the values that are most likely to generate any given sequence  $Q$  i.e

$$A^*, B^* = \operatorname{argmax}_{A, B} P(Q|A, B) \quad (24)$$

The problems above can be solved using the Viterbi algorithm and Baum-Welch algorithm.

### Solution to Finding the Hidden Sequence

The sequence of observations  $Q = (q_1, q_2, q_3, \dots, q_T)$  which corresponds to the decreasing, steady or increasing number of cases of Lassa fever per time and known transition and emission probability matrices  $A$  and  $B$  respectively are used to compute the most likely hidden sequence of the set of observations in  $Q$ . For each hidden sequence  $S$  of the state of Lassa fever at each time  $t$ , the joint probability that both  $O$  and  $S$  happen is:

$$P(Q, S|A, B) = P(Q|S, A, B) \cdot P(Q|A, B) \quad (25)$$

$$= \prod_{i=1}^T P(q_i | s_i) \cdot \prod_{i=1}^T P(s_i | s_{i-1}) \quad (26)$$

$$= P(q_1 | s_1) \cdot P(q_2 | s_2) \cdot P(q_3 | s_3) \cdots P(q_T | s_T) \cdot P(s_1 | s_0) \cdot P(s_2 | s_1) \cdot P(s_3 | s_2) \cdots P(s_T | s_{T-1})$$

$$= B(s_1, q_1) \cdot B(s_2, q_2) \cdots B(s_T, q_T) \cdot B(s_0, s_1) \cdot B(s_2, s_1) \cdots B(s_{T-1}, s_T)$$

$$= \prod_{i=1}^T B(s_i, q_i) \cdot \prod_{i=1}^T A(s_{i-1}, s_i) \quad (27)$$

The sequence  $S$  that maximizes the equation above is:

$$S^* = \operatorname{argmax}_S P(O, S|A, B) \quad (28)$$

The Viterbi algorithm simplifies and answer the problem of finding the most likely hidden sequence. It follows the following steps:

- Initialize: For each hidden state  $s$  of the state of Lassa fever at each time  $t$ :

$$g[1, s] \leftarrow B(s, o_1) \cdot A(s_0, s)$$

- For  $t = 2$  to  $T$ : For each hidden state  $s$ ,

$$g[t, s] \leftarrow \max_{s'} g[t-1, s'] \cdot A(s', s) B(s, o_t)$$

$$h[t, s] \leftarrow \operatorname{argmax}_{s'} g[t-1, s'] \cdot A(s', s) B(s, o_t)$$

- iii) Follow  $h[t, s]$  to find  $s_T^*, s_{T-1}^*, \dots, s_1^*$ . Starting at  $t = T$   
 $s_T^* \leftarrow \operatorname{argmax}_s g[T, s]$   
 $s_t^* \leftarrow h[t + 1, s_{t+1}^*]$  for  $t = T - 1, T - 2, \dots, 1$

### 3.2.3 Estimating the parameters of the model

The study used the Baum-Welch algorithm (Expectation-Maximization procedure) to obtain the estimated values of the transition matrix  $A$  and emission matrix  $B$ .  $A$  and  $B$  are initially initialized at random, then their values are repeatedly updated up to 1000 iterations. The Expectation step and the Maximization step are the first two steps in each updating iteration.

#### Expectation step

Since the transition matrix  $A$  and emission matrix  $B$  are known, we make use of the expectation step to compute the following:

$$\gamma[t, s] = P(s_t = s | Q, A, B) \quad (29)$$

$$\xi[t, s', s] = P(s_{t-1} = s', s_t = s | Q, A, B) \quad (30)$$

$\gamma[t, s]$  counts how many times does the  $t^{th}$  states of the hidden sequence of Lassa fever state equal  $s$  and  $\xi[t, s', s]$  counts how many time  $(s', s)$  happens at the  $(t - 1)^{st}$  step and the  $t^{th}$  step in the hidden sequence of Lassa fever, both up to normalization by a partition function.

$\gamma[t, s]$  and  $\xi[t, s', s]$  can be referred to as the pseudo counts.

Given the partial joint probabilities:

$$\alpha[t, s] = P(q_1, q_2, \dots, q_t, s_t = s | A, B)$$

$$\beta[t, s] = P(q_{t+1}, q_{t+2}, \dots, q_T | s_t = s, A, B)$$

then  $\gamma[t, s]$  and  $\xi[t, s', s]$  can be computed as follows:

$$\begin{aligned} \gamma[t, s] &= P(s_t = s | Q, A, B) \\ &= \frac{P(s_t = s | Q, A, B)}{P(Q | A, B)} \\ &= \frac{\alpha[t, s] \cdot \beta[t, s]}{\sum_{s'} \alpha[t, s'] \cdot \beta[t, s']} \\ \xi[t, s', s] &= P(s_{t-1} = s', s_t = s | Q, A, B) \\ &= \frac{P(s_{t-1} = s', s_t = s | Q, A, B)}{P(Q | A, B)} \\ &= \frac{\alpha[t-1, s'] \cdot A(s', s) \cdot \beta[t, s]}{\sum_{s'} \alpha[t, s'] \cdot \beta[t, s']} \quad (31) \end{aligned}$$

#### Maximization step

With  $\gamma[t, s]$  and  $\xi[t, s', s]$  obtained from the expectation step, we make use of the maximum likelihood estimator to derive the updated values for the parameters  $A$ , the transition matrix and  $B$ , the emission matrix, in the maximization step as follows:

$$\hat{A}(s', s) = \frac{\text{No of times } s \text{ follows } s'}{\text{No of times anything follows } s'} = \frac{\sum_{t=1}^{T-1} \xi[t, s', s]}{\sum_{s''} \sum_{t=1}^{T-1} \xi[t, s', s'']} \quad (32)$$

$$\hat{B}(s, o) = \frac{\text{No of times } o \text{ is observed given } s}{\text{No of times anytime is observed given } s} = \frac{\sum_{t=1}^T 1[o_t = o] \gamma[t, s]}{\sum_{t=1}^T \gamma[t, s]} \quad (33)$$

## 4. Result

### 4.1 Data

The dataset for this study was gotten from the weekly Lassa fever outbreak report on the National Centre for Disease Control and prevention (NCDC) website for the period between the 7<sup>th</sup> January 2021 and 29<sup>th</sup> January 2023 spanning 108 weeks. The fields for the dataset are suspected, confirmed, probable and death cases.

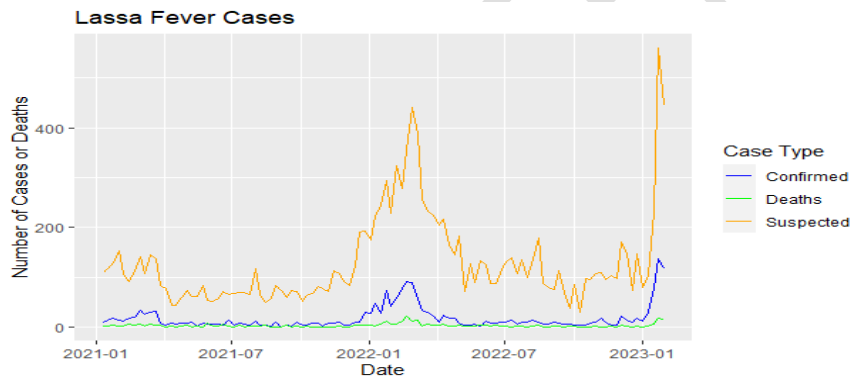
The incidence case count of suspected, confirmed and death cases of Lassa fever in Nigeria from January 2021 to January 2023 are described in Table 2 below:

**Table 2: Descriptive Statistics**

Statistic	Suspected	Confirmed	Deaths
Min	29.00	0.00	0.00
1st quart	72.75	5.00	0.00
Median	103.50	9.00	1.00
Mean	130.06	17.56	2.63
3 <sup>rd</sup> quart	145.25	17.25	3.00
Max	560.00	137.00	21.00

Table 2 reveals that there were an average of 130 suspected cases, with a minimum of 29 and a maximum of 560. With a mean of about 18 confirmed cases, the number of confirmed cases ranged from 0 to 137. A mean of about 3 deaths occurred; the number of fatalities ranged from 0 to 21. The summary indicates that there were relatively many suspected cases and fatalities, but only a small number of confirmed and likely cases. This could mean that identifying all cases of the disease or making an accurate diagnosis will be difficult.

The times series representation of the data is given below:



**Figure 1: Time Series plot of weekly incidence of Lassa fever cases in Nigeria from January 2021 to January 2023**

The plot shows a clear seasonal pattern in the incidence of Lassa fever, with the highest number of cases occurring from December through February and the lowest number of cases occurring between June and August. This pattern is consistent across all three trends - suspected cases, confirmed cases, and deaths.

The plot also shows a significant peak in the number of cases in the recent time, which is most pronounced in January 2023. This peak coincides with the highest number of suspected and confirmed cases, and is also reflected in the trend for deaths, although to a lesser extent.

**Table 3: Initial State Probabilities**

Susceptible	Infected	Recovered
0.924	0.076	0.000

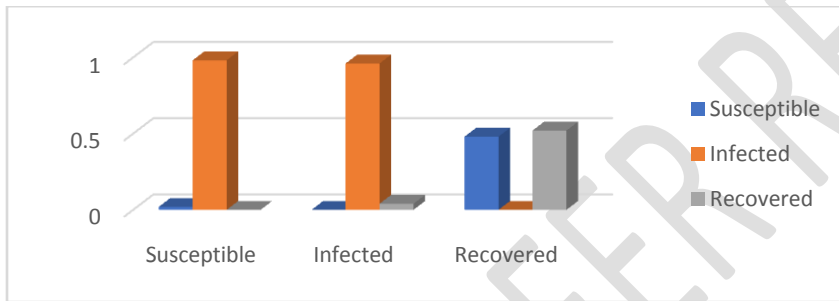
The initial state probability shows a 0.924(92.4%) probability of starting at the susceptible state, 0.076 (7.6%) probability of starting on the infected state and the probability of starting at recovery state is 0.

The transmission matrix gotten from the solution to the transition of the SIR model is as below:

**Table 0: Transition Probability Matrix**

	Susceptible	Infected	Recovered
Susceptible	0.019	0.981	0.000
Infected	0.000	0.960	0.040
Recovered	0.480	0.000	0.520

$$P(S_{t+1} = s' | S_t = s), \forall s', s \in S$$



**Figure 2: Graphical Representation of Transition Probability Matrix**

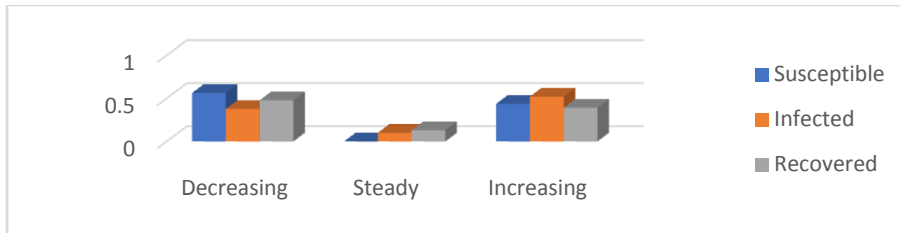
From the transition matrix, if the current state is infected, there is a 0.960 probability of staying infected, a 0.04 probability of moving to the recovery state, and a 0 probability of moving to the susceptible state.

The emission matrix B is given as:

**Table 5: Emission Probability Matrix**

	Decreasing	Steady	Increasing
Susceptible	0.563	0.000	0.437
Infected	0.377	0.101	0.522
Recovered	0.478	0.130	0.391

$$P(Q_t = q | S_t = s) \forall s \in S, q \in Q$$



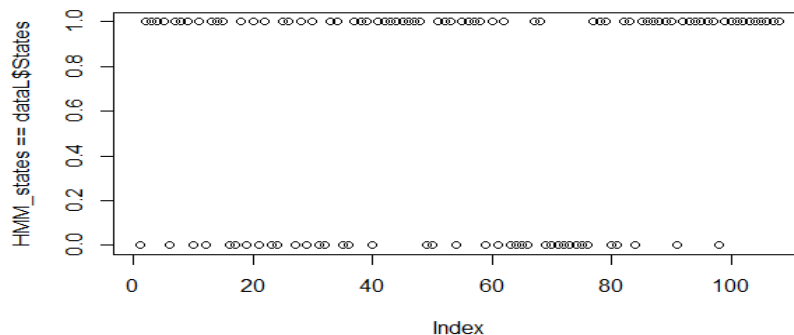
**Figure3:Graphical Representation of Emission Probability Matrix**

The stable state probability defined by  $\pi \times A$  is gotten as [0.017 0.979 0.0034]. There is a 98% chance of continuing in the infected state.

#### 4.2 Decoding the Most Likely Hidden Sequence Using Viterbi Algorithm

The Viterbi algorithm was applied to the HMM to infer the most likely sequence of hidden states given the observed sequence. The resulting sequence of hidden states which provides an estimate of the underlying process that generated the observed data is given below:

```
[1] "Susceptible" "Infected" "Infected" "Infected" "Infected" "Infected" "Infected"
[8] "Infected" "Infected" "Infected" "Infected" "Infected" "Infected" "Infected"
[15] "Infected" "Infected" "Infected" "Infected" "Infected" "Infected" "Infected"
[22] "Infected" "Infected" "Infected" "Infected" "Infected" "Infected" "Infected"
[29] "Infected" "Infected"
```



**Figure 4: Comparison of Actual Hidden Sequence to Predicted Hidden Sequence**

To obtain the proportion of predicted hidden sequence that tallies with the actual data, we use:  
`sum(HMM_states == dataL$States)/length(dataL$States)`

```
[1] 0.6422018
```

By comparing the predicted hidden sequence to the actual data, we got a 0.648 value which means that 64.8% of the hidden states are responsible for the observations.

#### 4.3 Estimating the Parameters of the Model

The Baum-Welch algorithm was applied to the HMM that we created to obtain the estimated values for the parameters of the model. These values are crucial for comprehending the HMM's behavior and making predictions based on the symbols that have been observed. The estimated values for the transition and emission matrices are presented in the sections that follow:

The general HMM is represented by the equation:

$$\theta = \{\pi, A, B\}$$

After 1000 algorithms of the Baum-Welch algorithm, the estimated HMM  $\hat{\theta} = \{\hat{\pi}, \hat{A}, \hat{B}\}$  has the following values:

$$\hat{\pi} = \{1.000, 0.000, 0.000\}$$

The initial probability vector in this instance has values of (1, 0, 0), indicating that there are no individuals in the infected or recovered states at the start of the system and that all individuals are in the susceptible state.

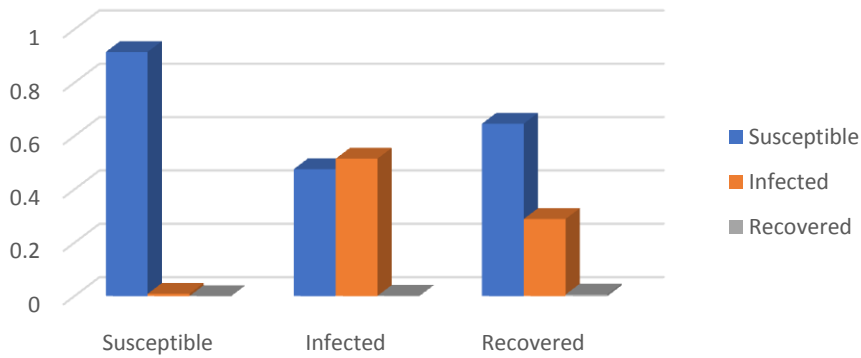
This interpretation is in line with the SIR model, according to which the disease can only start in a state that is susceptible. In this instance, the estimated initial probability vector suggests that everyone is susceptible to the disease at the start of the disease outbreak, as there are neither infected nor recovered individuals.

**Table 6: Estimated Transition Probabilities**

	Susceptible	Infected	Recovered
Susceptible	0.915	0.0085	0.000
Infected	0.476	0.515	0.001
Recovered	0.646	0.289	0.006

$$P(S_{t+1} = s' | S_t = s), \forall s', s \in S$$

From the table above, the estimated probability of staying in the susceptible state is high at 91.5%, while the estimated chance of transitioning to the from the susceptible state to the infected state is low at 0.85%. The estimated probability of staying infected is relatively high 51.5%.



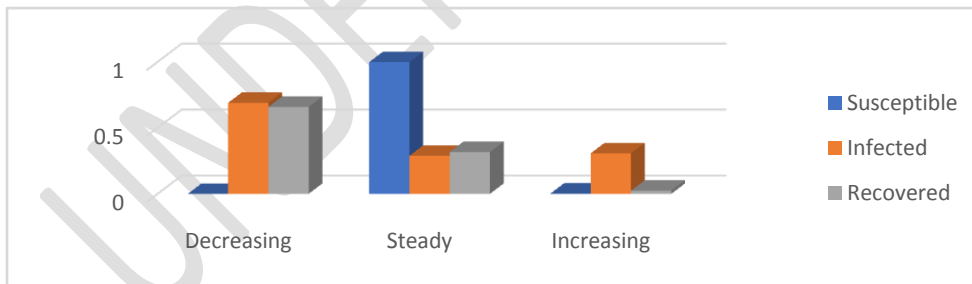
**Figure 5: Graphical Representation of Estimated Transition Probabilities**

Table 7: Estimated Emission Probabilities

	Decreasing	Steady	Increasing
<b>Susceptible</b>	0.000	0.998	0.002
<b>Infected</b>	0.690	0.288	0.307
<b>Recovered</b>	0.659	0.317	0.024

$$P(Q_t = q | S_t = s) \forall s \in S, q \in Q$$

The probability of observing the emission Infected given that the HMM is in the Decreasing state is 0.690, while the probability of observing Recovered given that the HMM is in the Increasing state is 0.024.



**Figure 6: Graphical Representation of the Estimated Emission Probabilities**

## 5 Discussion and Conclusion

This study surveilled the occurrence and state transmission of Lassa fever in Nigeria. The time plot showed a peak period of Lassa fever to during the dry season between January and March and the lowest number of confirmed cases to be during the raining season between June and August. Furthermore, we used the SIR model to define the hidden states of the hidden Markov model and thereby compute the transition probability matrix of the hidden Markov model. The transition probability matrix estimated showed a 98% chance of transitioning to the infected state from being susceptible and also gave a 96% chance of staying in infected.

The HMM was then applied to the observations from the Lassa fever dataset available from January 2021 to January 2023 on the Nigeria Centre for Disease Control and Prevention (NCDC) website. The model then depicted the most likely hidden sequence of Lassa fever given the observation of the confirmed cases of Lassa fever. The result showed that the hidden sequence was responsible 65% of the observed states. The stable state probability estimated a 97.9% probability of transitioning to the infected state, 1.7% chance of transition to the susceptible state.

Conclusively, this research work demonstrates the effectiveness of using HMM to surveil the occurrence of Lassa fever in Nigeria. Our analysis of Lassa fever data in Nigeria showed that the HMM model was able to accurately capture the dynamics of the disease transmission process, showing the most probable state as the infected state thereby proving a valuable tool for public health officials to track and respond to outbreaks of the disease.

## References

Abdulhamid, A.; Hussaini, N.; Musa, S.S.; He, D. (2022) Mathematical analysis of Lassa fever epidemic with effects of environmental transmission. *Results Phys.* 35, 105335.

Akinpelu, F.O.; Akinwande, R. (2018). Mathematical Model for Lassa fever and Sensitivity Analysis. *J. Sci. Eng. Res.* 5, 1–9.

Bakare, E., Are, E., Abolarin, O., Osanyinlusi, S., Ngwu, B., Ubaka, O.N.: Mathematical modelling and analysis of transmission dynamics of Lassa fever. *J. Appl. Math.* 2020 (2020)

Fichet-Calvet, E., Rogers, D.J. (2009). Risk maps of Lassa fever in west Africa. *PLoS Negl. Trop Dis.* 3(3), 388.

Hamblion, E.L.; Raftery, P.; Wendl, A.; Dweh, E.; Williams, G.S.; George, R.N.; Soro, L.; Katawera, V.; Clement, P.; Gasasira, A.N.; et al. (2018) The challenges of detecting and responding to a Lassa fever outbreak in an Ebola-affected setting. *Int. J. Infect. Dis.* 66, 65–73.

Ibrahim, M.A.; Denes, A. A (2021) Mathematical model for Lassa fever transmission dynamics in a seasonal environment with a view to the 2017–20 epidemic in Nigeria. *Nonlinear Anal. Real World Appl.* 60, 103310.

Innocent, E.B.; Omo, E. Lassa fever and its Control Measures. *Sci. Res.* 2017, 7, 75–79.

James, T.O.; Akinyemi, S.T.; Oluwade, B. Stability Analysis of Lassa fever with Quarantine and Permanent Immunity. *Int. Inst. Acad. Res. Dev.* 2015, 1, 71–81.

Marien, J.; Borremans, B.; Kourouma, F.; Baforday, J.; Rieger, T.; Günther, S.; Magassouba, N.F.; Leirs, H.; Fichet-Calvet, E. (2019). Evaluation of rodent control to fight Lassa fever based on field data and mathematical modelling. *Emerg. Microbes Infect.* 8 (1), 640–649.

Musa, S.S.; Zhao, S.; Gao, D.; Lin, Q.; Chowell, G.; He, D. (2020). Mechanistic modelling of the large-scale Lassa fever epidemics in Nigeria from 2016 to 2019. *J. Theor. Biol.* 493, 110209.

Ndenda, J.P.; Njagarah, J.B.; Shaw, S. (2022) Influence of environmental viral load, interpersonal contact and infected rodents on Lassa fever transmission dynamics: Perspectives from fractional-order dynamic modelling. *AIMS Math.* 2022, 7, 8975–9002.

Nkemnole E.B. and Udoh, N. (2022). Modeling Covid 19 as Hidden Markov Chain

Ojo, M.M.; Gbadamosi, B.; Benson, T.O.; Adebimpe, O.; Georgina, A.L. (2021) Modeling the dynamics of Lassa fever in Nigeria. *J. Egypt. Math. Soc.* 2021, 29, 1–9.

Onah, I.S., Collins, O.C. (2020). Dynamical system analysis of a Lassa fever model with varying socioeconomic classes. *J. Appl. Math.* 1-12.

World Health Organization. Lassa Fever Fact Sheets. 2022. Available online: <https://www.who.int/news-room/fact-sheets/>

World Health Organization, Lassa Fever-Nigeria. 2022. Available online: <https://www.who.int/emergencies/disease-outbreaknews/item/lassa-fever---nigeria> (accessed on 15 January 2023)

Zhao, S.; Musa, S.S.; Fu, H.; He, D.; Qin, J. (2020) Large-scale Lassa fever outbreaks in Nigeria: Quantifying the association between disease reproduction number and local rainfall. *Epidemiol. Infect.* 148, 1-12.