International Journal of TROPICAL DISEASE & Health

Volume 44, Issue 18, Page 1-14, 2023; Article no.IJTDH.105450
ISSN: 2278–1005, NLM ID: 101632866

An Analysis of the Hidden Markov Model for Surveilling the Transmission of Lassa Fever Epidemic Disease in Nigeria during Dry Season

Nkemnole E. B. [a*] and Oyewole J. O. [a]

[a] Department of Statistics, University of Lagos, Nigeria.

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/IJTDH/2023/v44i181473

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/105450

Original Research Article

Received: 12/07/2023
Accepted: 14/09/2023
Published: 25/09/2023

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 illnesses and death. Finding some mathematical patterns that explain the mechanisms of Lassa fever transmission, as well as a thorough understanding of the biological contributing to affecting the disease, are necessary in putting in place a surveillance system with a view to preventing further spread of the disease. In this study, we developed a Hidden Markov Model (HMM) approach to surveil the transmission of Lassa fever virus infections in Nigeria. The HMM was developed using the susceptible Infection recovered (SIR) model to formulate the transition matrix and data from past outbreaks of the disease to compute the observations. Our results showed that the dry season as the peak period for Lassa fever and recorded the lowest numbers during the rainy season. The transition matrix showed a 98% chance of transitioning to the infected

**Corresponding author: Email: enkemnole@unilag.edu.ng;*

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 epidemic outbreaks.

Keywords: Lassa fever; hidden Markov model; infectious diseases; transition probability matrix; emission matrix; Viterbi algorithm.

1. INTRODUCTION

Lassa fever is a viral illness that is 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 [1, 2, 3]. These rats are abundant in rural areas of parts of some West African countries, including Nigeria, Sierra Leone, Liberia, and Guinea [4]. The low standards of living and poor sanitations are the available reasons this virus is dominantly common in these rural communities.

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

In Nigeria, Lassa fever is endemic and is typically detected during the dry season (December–April each year) 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. 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 in Nigeria. Preventive measures such as improved household sanitation, proper storage of foodstuffs, environmental sanitation, and keeping cats pets could reduce the ailment. Animal products should be meticulously cooked properly, early detection and proper treatment of the virus, seclusion of infected persons are thought to be some of the actions that could reduce the risk of person-to-person transmission of the disease in health care facilities [4].

Dissemination of protective guidelines for proper case management, and infection prevention and control (IPC); enhanced surveillance activities in Lassa fever-affected areas to increase detection of cases; provision of special treatment health-centres facilities for clinical management in affected areas; increased laboratory capacity to ensure timely processing and diagnosis of samples; at risk communications individuals' and community engagement activities through television, radio, print, social media, and other strategies are strongly endorsed to lessen the fatality rate [5].

A 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 all over the world. The model considered here, can assist among the various existing measures against the disease for being more effective in predicting its occurrence. Though Lassa fever appears in WHO's lists of prioritizing diseases for research and development in emergency contexts, only a few data 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 endemics of diseases [6, 7].

A multiple-patch model was developed by Onah and Collins, [6] to examine the effects of socioeconomic class on Lassa fever (LF), and showed a sensitivity analysis and a numerical illustration of the effect of parameter models in the spread of disease and incidence. Their results revealed that humans' socioeconomic status has a significant influence on the dynamics of viral LF transmission. The study

recommends that human socioeconomic classes should 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*., [8]. In regard to eradication of LF in rural areas, the authors used a mathematical model to experiment numerous control approaches in rural Upper Guinea to determine the length and frequency of control strategies as to when they should be performed. The mathematical model suggests that the best strategy for eradicating LF is continuous control or rodent vaccination against the viral disease. A spatial analysis of Lassa fever data from human cases and infected rodents from 1965 to 2007 was performed in Fichet-Calvet and Rogers [9], to describe the LF risk maps in Western African Region. 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 and showed that rainfall has strong effect in defining high risk areas, while temperature has a lesser effect in high-risk areas endemic for the disease. Additionally, the risk maps revealed that the most unsafe region is located between Guinea and Cameroon within the West African Region.

By using a mathematical model [10] 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*. [11] used mechanistic modelling that takes into account quarantine, isolation, and hospitalization processes of Lassa fever victims during epidemics in Nigeria from 2016–2019. Particularly, data showed some similarities in the transmission dynamics driving three major Lassa fever outbreaks, from 2016–2019 in Nigeria, as were outlined by their study. Ndenda *et al*. [12] also used fractional-order dynamic modelling to study the effects of environmental viral load of Lassa, interpersonal contact, and infected rodents on the transmission dynamics. They discovered that with multiple interventions and control measures, such as environmental sanitation, the methods could significantly help in eradicating infections. In another study, Marien *et al*. [8] applied a mathematical model to study the impact of rodents’ control to fight Lassa fever, and the team showed that rodent vaccination is an approach that could eliminate Lassa virus in the disease endemic area. Abdulhamid *et al*. [13] also used a deterministic mathematical model to study Lassa fever dynamics and the study revealed that the existence of backward bifurcation in the model makes the control of Lassa fever more difficult to achieve. Zhao *et al*. [14] presented another mathematical model to studying the effects of rainfall on Lassa fever epidemics in Nigeria, by quantifying the association between reproduction number and rainfall for several locations in Nigeria, Results from their study shows clear evidence of rainfall impacts on LF epidemics in Nigeria

Innocent and Omo [15] developed a mathematical model for investigating the dynamics of the LF disease, and made recommendation to the effect that avoiding contact with species that carried the viruses and introducing vaccines against it for humans would be the most effective method of control. In the other hand, Akinpelu and Akinwande [16] developed a mathematical model for sensitivity analysis of Lassa fever, where the model was divided into five compartments of susceptible (S), latent (L), infected (I), isolated (I), and recovered (R). By using the next-generation method, what they obtained showed that disease-free equilibrium was locally and globally asymptotically stable. James *et al*., [17] analyzed stability for Lassa fever and recommended quarantines and making strategies for permanent immunity was an alternative method of LF control. Bakare *et al*. [6] study worked on the transmission dynamics of the disease and derived a nonlinear ordinary differential equation model by introducing the seasonal parameters. The results showed a basis for planning and designing cost-effective strategies for interventions in eradicating Lassa fever.

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

Adigun, *et al*. (2019) examined the environmental elements that contribute to the transmission of various infectious diseases. The study calculated the diseases' patterns of transition, testing the Markovian property on how stationary the process is over the study period and concluded that the past history of infectious diseases would have an impact on the present state through the current condition. In the study, the team recommended that more interventions need 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 would 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 showed that over the long term, 8.4% of the population were at risk for COVID-19 infections and 26.4% others would become infected; 61.2% of those who became infected would recover, while 4% of those who become infected would likely die.

Nkemnole and Osunkeye (2016) investigated the prevalence of endemic diseases in a population and their duration of resistance using a Stochastic model, similar to the Markov's chain, which has a continuous time and discrete state space and necessitated that the Monte Carlo simulation would produce the desired results in disease control. 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, and revealed a significant correlation between gender and the duration of endemic disease persistence in a particular population. Nkemnole and Udoh (2022) had built a transition, as well as emission probabilities of Covid-19 cases and created a hidden Markov model of prediction. 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, the model predicted a recovery rate of 25% chance, and moving from an infected state to death had a 50% chance; and the chance of still being re-infected after recovery, there would be a 25% chance.

These studies have made significant progress in the dynamics and control of Lassa fever infections. Nevertheless, finding a reliable mathematical pattern that explains 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 disease epidemic is a toll order. The Nigeria government should find it necessary to put in place a surveillance system aimed at preventing further disease spread. As a result, it is important to examine the trends of the seasonality of Lassa fever with a view to employing the Hidden Markov Model to accurately predict the occurrence of Lassa fever in Nigeria.

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

3. METHODS

3.1 Epidemiology of Lassa Fever

Epidemiology is the study of how diseases and their causes are distributed throughout human 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 outcomes, 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 and predicting the transmission of Lassa Fever Virus infections.

3.2 Epidemic Surveillance

Epidemic surveillance is a process of collecting, analyzing, and disseminating data to detect and respond to disease outbreaks or other public health events (WHO, 2002). A surveillance system can be classified into two main types: passive and active.

3.3 Passive Surveillance

systems rely on the reporting of health events by healthcare providers, laboratories, or individuals. Such systems are useful for detecting disease outbreaks after they have started, but they have limitations in detecting outbreaks in real-time. In the other hand, active surveillance systems, seek out cases of a disease or health events by actively searching for cases within a community in question. This type of surveillance is typically more resource-intensive, but however, can detect outbreaks of diseases quite earlier and provide more detailed information about the impending outbreaks. There are also hybrid systems that combine both the elements of

passive and active surveillance, taking advantage of the strengths of both types (Petersen & Fullerton, 2018).

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

3.4 Susceptible-Infected-Recovered (SIR) Model

The SIR model is a classical mathematical model used in epidemiology to study the spread of infectious diseases (Hethcote, 2000) in a community/population. It represents the three distinct states 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 further.

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 β is the transmission rate, and γ is the recovery rate. The solutions obtained from these equations gives the evolution of the number of individuals in each state over time (William & McKendrick, 1927).

The formulation of the SIR model in this study for Lassa fever transmission between states was established. The different 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 parameters of each model.

3.5 The 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, Λ_h, populates the susceptible human population through birth or immigration and from recovered subpopulation due to γ 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
α_1	Infection rate of rodents
α_2	The force of infection
α_3	Infection rate of human interact
τ_c	The connection of humans with drugs
τ_{nc}	The rate at which humans do not have a relationship with drugs
r_c	Rate of awareness
γ	Loss of immunity
Λ_h	Birthrate of humans
μ_h	Deathrate of humans
δ	Rate of mortality of an infectious class
Λ_r	The birthrate of rodents
μ_r	The deathrate of rodents
$\sigma_i . i = 1,2,3,4,5$	Randomness of each sub-population

It was opined that some of the models' characteristics ought to be validated for accuracy, whether or not would represent the transmission of Lassa fever; few of those characteristics are given below.

3.6 Transition Probabilities from the SIR Model

Given the earlier defined SIR model, the transition probabilities from one hidden state to another in 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 state (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 β and γ are transmission rand recovery rate respectively.

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

3.7 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.

3.8 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. $\forall t \geq 0, s \in 0, \ldots, t$

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

3.9 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 in the model. The parameter estimation is to find the values that best describe the underlying processes that generated the data. The parameter estimates obtained from this process are then used to make predictions the future disease 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 (π) are calculated based on the number of hidden states.
ii) Compute the forward probability matrix α and the backward probability matrix β 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:

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

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.

One has to 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.

3.9 Estimating the Stages of the LF Virus with HMM

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

3.10 Transition and Emission Matrix

A is the transition probability matrix of 3×3 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.

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

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

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

$$\begin{array}{c} \\ \mathbf{S} \\ \mathbf{I} \\ \mathbf{R} \end{array} \begin{array}{c} \begin{array}{ccc} \mathbf{DC} & \mathbf{S} & \mathbf{IN} \end{array} \\ \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \end{array}$$

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

The initial probability vector $\pi = \{\pi_1, \pi_2, \pi_3\}$ represents the probability of starting the process from the susceptible, infected or recovered state of Lassa fever. The hidden Markov model θ 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:

i) **Finding the hidden sequence:** The transition matrix A, emission matrix B and a sequence of given observations $Q = (q_1, q_2, q_3, \ldots, q_T)$ 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

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

ii) **Estimating the parameters:** The transition matrix A and emission matrix which are then estimated, and 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

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

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

3.11 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 infections 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:

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

$$= \prod_{i=1}^{T} P(q_t|s_t) \cdot \prod_{i=1}^{T} (s_t|s_{t-1}) \quad (26)$$

$$= P(q_1|s_1) \cdot P(q_2|s_2) \cdot P(q_3|s_3) \cdots P(q_T|s_T) \cdot P(s_1|s_0) \cdot P(s_2|s_1) \cdot P(s_3|s_2) \cdots P(s_T|s_{T-1})$$
$$= B(s_1, q_1) \cdot B(s_2, q_2) \cdots B(s_T, q_T) \cdot B(s_0, s_1) \cdot B(s_2, s_1) \cdots B(s_{T-1}, s_T)$$
$$= \prod_{i=1}^{T} B(s_t, q_t) \cdot \prod_{i=1}^{T} A(s_{t-1}, s_t) \quad (27)$$

The sequence S that maximizes the equation above is:

$$S^* = arg\max_{S} P(O, S|A, B) \quad (28)$$

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

i) Initialize: For each hidden state s of the state of Lassa fever at each time t:
$g[1, s] \leftarrow B(s, o_1) \cdot A(s_0, s)$

ii) For $t = 2\ to\ T$: For each hidden state s,
$g[t, s] \leftarrow \max_{s'} g[t-1, s'] \cdot A(s', s) B(s, o_t)$

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

iii) Follow $h[t, s]$ to find $s_T^*, s_{T-1}^*, \dots, s_1^*$. Starting at $t = T$

$$s_T^* \leftarrow \underset{s}{\operatorname{argmax}}\, g[T, s]$$
$$s_t^* \leftarrow h[t+1, s_{t+1}^*] \quad \text{for } t = T-1, T-2, \dots, 1$$

3.12 Estimating the Parameters of the Model

The current study used the Baum-Welch algorithm (Expectation-Maximization procedure) to obtain the estimated values of the transition matrix A and emission matrix B.

Initially A and B were initialized at random, then their values were repeatedly updated up to 1000 iterations. The Expectation step and the Maximization step are the first two steps in each updating iteration.

3.13 Expectation Step

Since the transition matrix A and emission matrix B are known, we then 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:

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

$$= \frac{P(s_t = s|Q,A,B)}{P(Q|A,B)}$$

$$= \frac{\alpha[t,s]\cdot\beta[t,s]}{\sum_{s'}\alpha[t,s']\cdot\beta[t,s']}$$

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

$$= \frac{P(s_{t-1}=s', s_t=s|Q,A,B)}{P(Q|A,B)}$$

$$= \frac{\alpha[t-1,s]\cdot A(s',s)\cdot\beta[t,s]}{\sum_{s'}\alpha[t,s']\cdot\beta[t,s']} \quad (31)$$

3.14 Maximization Step

With $\gamma[t,s]$ and $\xi[t,s',s]$ obtained from the expectation steps, we then 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:

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

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

4. RESULTS

4.1 Data

The dataset for this study were obtained from the weekly Lassa fever outbreaks report on the National Centre for Disease Control and prevention (NCDC) website for the period between the 7th January 2021 and 29th January 2023, spanning 108 weeks. The fields for the dataset were: 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.

Table 2 reveals that there was an average of 130 suspected cases of LF in Nigeria between 2021 and 2023, with a minimum of 29 and a maximum of 560 reported. 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 during the study period; 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 them were either confirmed or likely cases. This could mean that identifying all cases of the disease or making an accurate diagnosis would be difficult.

The times series representation of the data is given :

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

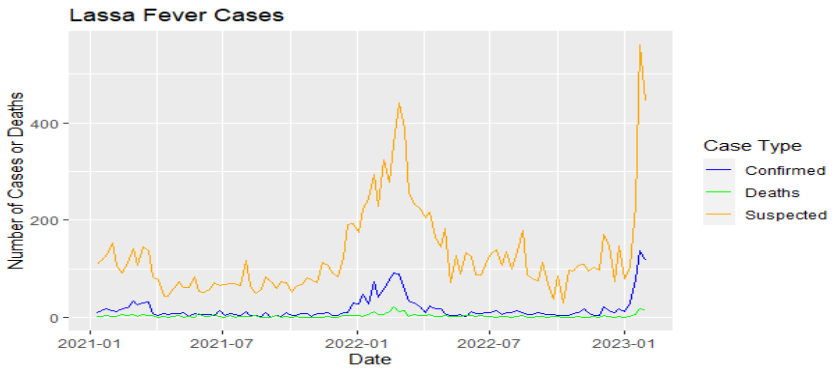

Fig. 1. Time Series plot of weekly incidence of LF cases in Nigeria between 2021 and 2023

Table 2. Descriptive statistics of lassa fever cases in Nigeria between 2021 and 2023

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
3rd quart	145.25	17.25	3.00
Max	560.00	137.00	21.00

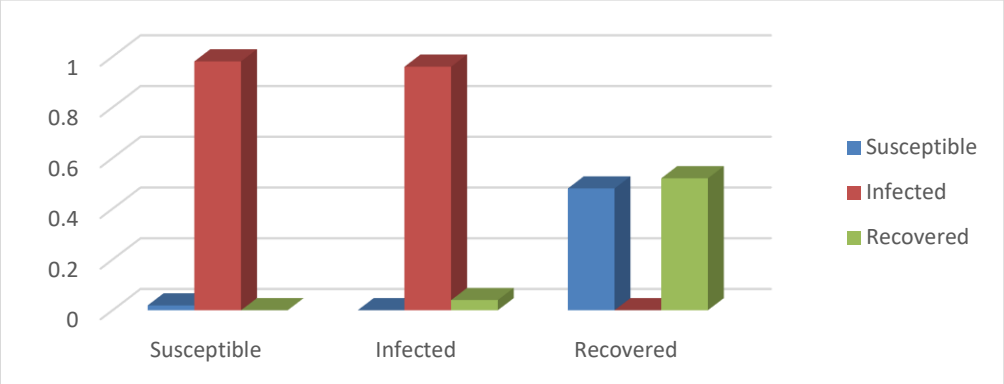

Fig. 2. Graphical representation of transition probability matrix

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 could be 0.

The transmission matrix obtained from the solution to the transition of the SIR model is as given in Table 4:

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 in Table 5 as:

Table 3. Initial state probabilities

Susceptible	Infected	Recovered
0.924	0.076	0.000

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

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

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

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

The stable state probability defined b y π × A is recorded as $[0.017 \quad 0.979 \quad 0.0034]$. 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"

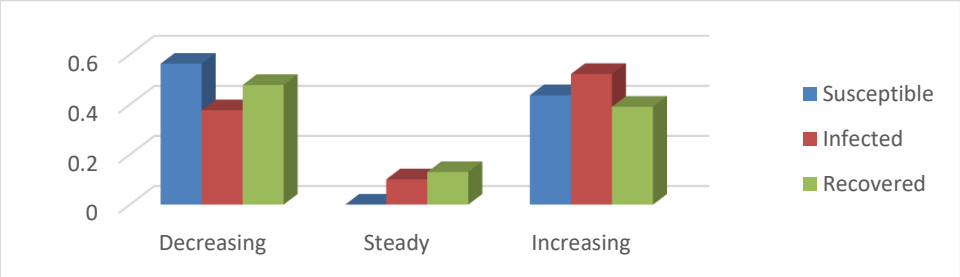

Fig. 3. Graphical Representation of Emission Probability Matrix

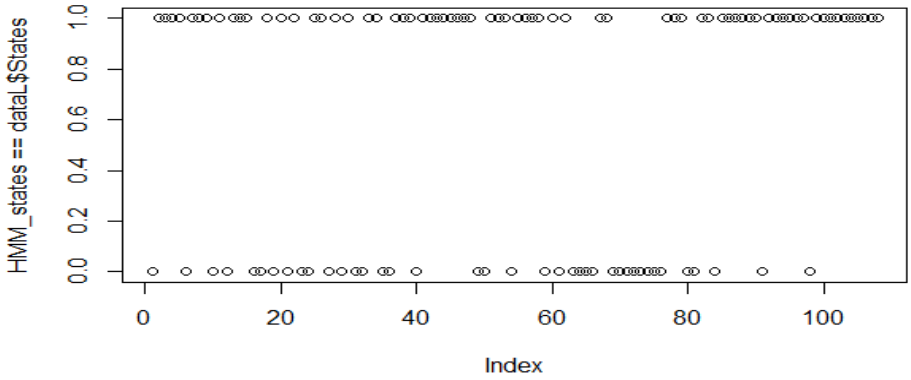

Fig. 4. Comparison of actual hidden sequence to predicted hidden sequence

To obtain the proportion of predicted hidden sequence that tallies with the actual data, we use:

Sum(HMM_statesdataL$States)/length(dataL$States) [1] 0.6422018

By comparing the predicted hidden sequence to the actual data, we recorded a 0.648 value, which means that 64.8% of the hidden states was 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 were crucial for comprehending the HMM's behavior and making predictions based on the symbols that had 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 could 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 Lassa Fever infections

	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 Table 6, the estimated probability of staying in the susceptible state would be very high at 91.5%, while the estimated chances of transitioning from the susceptible state to the infected state also would be very low at 0.85% of the chances. The estimated probability of staying infected state would relatively be more than half at 51.5% of the chances.

The probability of observing the emission-infected individuals, 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.s

5. DISCUSSION AND CONCLUSION

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

The HMM was then applied to the observations from the Lassa fever dataset, which was 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 viral Lassa fever given the observation of the confirmed cases of the disease. The result showed that the hidden sequence was responsible of 65% of the observed states. The stable state probability was estimated at 97.9% probability of transitioning to the infected state, and 1.7% chance of transitioning to the susceptible state [18].

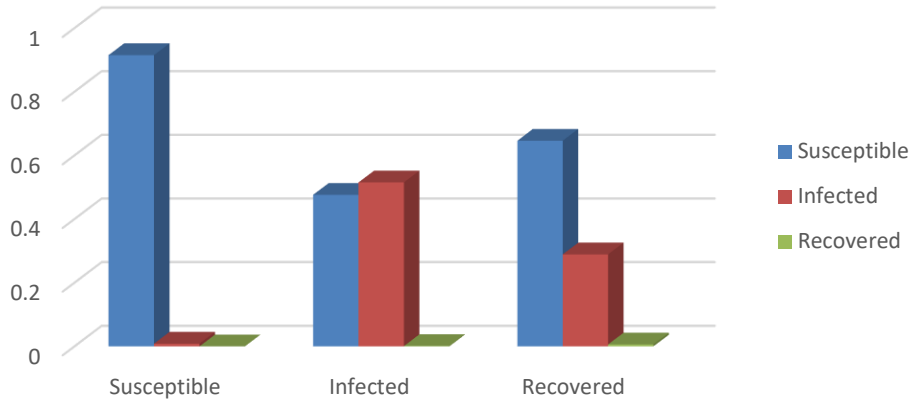

Fig. 5. Graphical Representation of Estimated Transition Probabilities

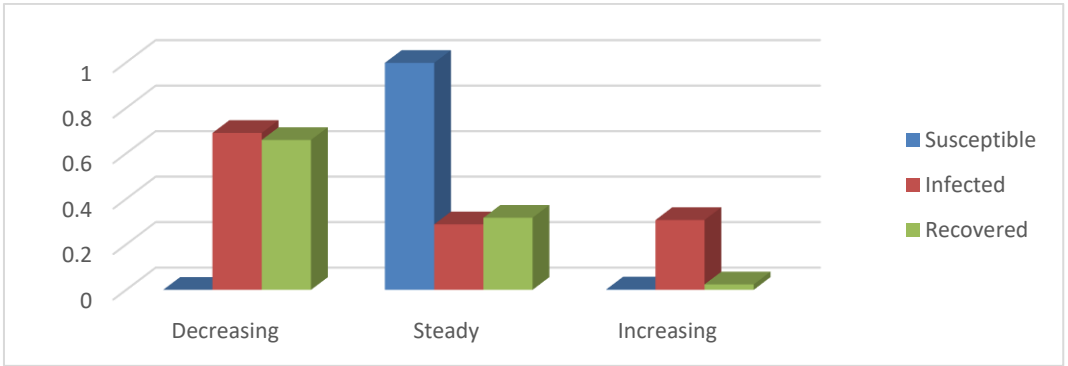

Fig. 6. Graphical Representation of the Estimated Emission Probabilities

Table 7. Estimated emission probabilities

	Decreasing	Steady	Increasing
Susceptible	0.000	0.998	0.002
Infected	0.690	0.288	0.307
Recovered	0.659	0.317	0.024

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

Conclusively, this research work has demonstrated 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 prediction of disease transmission process, showing the most probable state of infection and thereby proving to be a valuable tool for public health officials to track and respond to outbreaks of the disease in question.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Ibrahim MA; Denes AA 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. 2021;60:103310.
2. Hamblion EL, Raftery P, Wendl A, Dweh E, Williams GS, George RN, Soro L, Katawera V, Clement P, Gasasira AN. et al. The challenges of detecting and responding to a Lassa fever outbreak in an Ebola-affected setting. Int. J. Infect. Dis. 2018;66:65–73.
3. Davies J, Lokuge K, Glass K. Routine and pulse vaccination for Lassa virus could reduce high levels of endemic disease: A mathematical modelling study. Vaccine. 2019;37:3451–3456.
4. World Health Organization, Lassa Fever-Nigeria; 2022
   Available:https://www.who.int/emergencies/disease-outbreaknews/item/lassa-fever---nigeria
   accessed on: 2023
5. World Health Organization. Lassa Fever Fact Sheets; 2022.
   Available:https://www.who.int/news-room/fact-sheets/
6. Onah IS, Collins OC. Dynamical system analysis of a Lassa fever model with varying socioeconomic classes. J. Appl. Math. 2020;1-12.
7. Bakare E, Are E, Abolarin O, Osanyinlusi S., Ngwu B, Ubaka ON. Mathematical modelling and analysis of transmission dynamics of Lassa fever. J. Appl. Math.2020:18.
   Available:https://doi.org/10.1155/2020/6131708
   Article ID 6131708
8. Marien J, Borremans B, Kourouma F, Baforday J, Rieger T, Günther S, Magassouba NF, Leirs H, Fichet-Calvet E. Evaluation of rodent control to fight Lassa fever based on field data and mathematical modelling. Emerg. Microbes Infect. 2019;8(1):640–649.
9. Fichet-Calvet E, Rogers DJ. Risk maps of Lassa fever in west Africa. PLoS Negl. Trop Dis. 2009;3(3):e388.
   DOI:10.1371/journal.pntd.0000388
10. Ojo MM, Gbadamosi B, Benson TO, Adebimpe O, Georgina AL. Modeling the dynamics of Lassa fever in Nigeria. J. Egypt. Math. Soc. 2021;29:1–9.
11. Musa SS, Zhao S, Gao D, Lin Q, Chowell G, He D. Mechanistic modelling of the large-scale Lassa fever epidemics in Nigeria from 2016 to 2019. J. Theor. Biol. 2020;493:110209.
12. Ndenda JP, Njagarah JB, Shaw S. 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.
13. Abdulhamid A, Hussaini N, Musa SS, He D. Mathematical analysis of Lassa fever epidemic with effects of environmental transmission. Results in Physics Journal. 2022;35:105-335.
14. Zhao S, Musa SS Fu H, He D, Qin J. Large-scale Lassa fever outbreaks in

Nigeria: Quantifying the association between disease reproduction number and local rainfall. Epidemiol. Infect. 2020;148:1-12.
15. Innocent EB, Omo E. Lassa fever and its Control Measures. Sci. Res. 2017;7:75–79.
16. Akinpelu FO, Akinwande R. Mathematical Model for Lassa fever and Sensitivity Analysis. J. Sci. Eng. Res. 2018;5:1–9.
17. James TO, Akinyemi ST, Oluwade B. Stability Analysis of Lassa fever with Quarantine and Permanent Immunity. Int. Inst. Acad. Res. Dev. 2015;1:71–81.
18. Collins OC. Kevin JD. Using Data of a Lassa Fever Epidemic in Nigeria. A Mathematical Model Is Shown to Capture the Dynamics and Point to Possible Control Methods. MDPI, Mathematics 2023;11(5):1181.
19. Nkemnole EB, Udoh N. Modeling Covid 19 as Hidden Markov Chain. Review and Research: Journal of Statistics. 2022; 11(2):27-35.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/105450