Applied Mathematical Sciences, Vol. 11, 2017, no. 58, 2855 - 2875
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2017.711320

Mathematical Modelling of Liver Cancer in Western Kenya

Otedo Amos[1], Estambale Benson and Omolo Ongati

Jaramogi Oginga Odinga University of Science and Technology
P.O. Box 210-40601, Bondo, Kenya
[1] Corresponding author

Simbiri Kenneth

Temple University, Philadelphia, Pennsylvania, USA

Abstract

Liver cancer, also known as hepatocellular carcinoma (HCC) is a primary cancer of the liver and the fifth cause of mortality world-wide. It is a global public health problem which is poorly addressed in the developing countries. Data on prevalence and incidence is scanty leading to inability to predict the burden of HCC in the developing world and this leads to poor policy framework for management and control of HCC. More-over, management and control of HCC is poorly addressed in Kenya. Most subjects with HCC in the developing world present late to the hospital leading to high mortality. The objectives of this study were to develop a predictive mathematical model to predict the proportion of subjects who would develop HCC over 5 years in western Kenya and to conduct a sensitivity analysis of the model developed to ascertain effectiveness in prediction. Liver cancer is the only cancer with both infectious and non-infectious causes. The design of the study was a hybrid mathematical model developed integrating both I.P.M (incidence, prevalence, mortality) and S.I.R (susceptible, infected, recovered) models. Ordinary differential equations were generated and solved using Matlab software to predict burden of HCC. MatLab software was also used to generate graphs to predict the number of subjects who will develop HCC over time. Parameters used were generated from empirical data from the study and secondary sources. The study was approved by the ethics committee at Jaramogi Oginga Odinga teaching and referral hospital. The study site was Kisumu

county and referral hospital and was conducted between June 2015 and June 2016. Out of 331 (231 males and 100 females) subjects screened, 257 (178 males and 79 females) subjects were included and 74 (50 males and 24 females) were excluded (no liver cancer). Ordinary differential equations were developed which included temporal parameters in the model which were the susceptible population, HCC incidence rates, death rates and birth rates. A schematic hybrid model modified from I.P.M + S.I.R model was developed. S.I.R + I.P.M model was used because HCC cases invariably die but some cases of HCC in early stages survive for some time. In conclusion, this study has established the applicability of the hybrid mathematical model from SIR and IPM for liver cancer burden. Liver cancer burden in western Kenya would increase over time unless risk factors are controlled. The models are sensitive and effective in predicting the burden of liver cancer in the community. This study provides health care workers and stake holders with information to enable generation of policy for management and prevention of HCC.

Keywords: Liver cancer, SIR and IPM model, prediction of burden, Western Kenya

1. Introduction

Liver cancer, also known as hepatocellular carcinoma (HCC), is a malignant cancer of the liver which is the second largest organ of the body. It is a global public health problem. It is the 5th common cancer worldwide and the 3rd and 4th important cause of cancer related deaths in men and women world-wide respectively [14]. It is a public health problem which is often under diagnosed or diagnosed late. HCC is a "silent killer," because signs and symptoms usually don't appear until the disease is at an advanced stage. Data is lacking to predict the number of subjects who will develop HCC in the community, making planning to be difficult. Estimating the burden of liver cancer in the community is important in understanding adequate resource allocation, mobilization and prevention strategies. Disease models describing the relationship between incidence, prevalence and mortality (I.P.M) are used to detect data problems or supplement missing data [18, 19]].

Estimates of disease-specific incidence, prevalence and mortality (I.P.M) are essential and have been used in burden of disease calculations. It's noteworthy that empirical data, however, are often difficult to obtain or are of questionable validity. To remedy some of these data problems, disease models have been developed that describe the relationship between the epidemiological parameters by exploiting the causal structure of a disease [18, 19]. The incidence, prevalence, mortality (I.P.M) models formalize the relationship between the three parameters, using the fact that incidence has to precede prevalence and that cause-specific mortality can only follow disease. IPM models have been used frequently both to supplement missing data, study burden of disease and the agreement between different epidemiological data [3, 11, 14, 15, 18, 19]. Hybrid SIR + IPM model was

used because most HCC cases invariably die but, some HCC cases in early Barcelona Clinic Liver clinic (BCLC) stages 0, 1 and 2 survive for some time.
Most prediction approaches use appropriate statistical models by analyzing the available data to predict cancer incidences [2, 16, 22, 25]. Pancreas and liver cancers (HCC) are projected to surpass breast, prostate, and colorectal cancers and are likely to become the second and third leading causes of cancer-related death by 2030, respectively in the US [14]. HCC is an ignored non communicable disease (NCD) in the developing world due to changing epidemiology secondary to unknown causes and HIV, poor policy frame work for prevention, inadequate specialized human resource and insufficient surveillance and research. In Africa, there are only three population based cancer registries in three countries namely; Gambia, Uganda and Zimbabwe [23]**.** The Kenya cancer registry based at Kenya medical research institute (KEMRI) is not cited in literature because it is poorly maintained and there is no population cancer registry in western Kenya. The population based cancer registry is important in evaluating the burden of the disease [26]. Projection of burden of liver cancer will enable the government to allocate enough resources to manage HCC and enact policies for prevention. Health policy and planning depend on quantitative data of disease epidemiology.

2. Materials and Methods

The study was approved by the ethics committe of the IRB (Institutional Review board) of Jaramogi Oginga Odinga Teaching and Referral Hospital, Kisumu. The study site was Kisumu county and referral hospital liver clinic and medical wards and the study utilized data from subjects with liver cancer who were referred from the peripheral hospitals in western Kenya. Informed, signed consent was obtained from each subject, (< 18 years old signed by parent (s) or guardian). Using data collected between June 2015 and June 2016 which included 331 (231 males and 100 females) subjects screened, 257 (179 males and 78 females) liver cancer cases diagnosed using both imaging (liver ultrasound and alfa-fetoprotein levels) were included.
A standard questionnaire was administered for each study subject and it included: patients baseline information:- demographics-bio-data [age (years), sex, (M/F], symptoms and signs of HCC. Each study subject underwent a physical clinical examination. Under aseptic technique, blood was drawn from the cubital fossa and used to determine alpha-fetoprotein levels, liver function tests (ALT, AST and INR), markers of HBV, HCV and HIV and urine for aflatoxins. Liver ultrasound was done for all the subjects. The liver cancer cases were classified by BCLC system [7]. BCLC (Barcelona Clinic Liver Clinic) is a staging system of liver cancer.
The study design was a hybrid S.I.R + I.P.M model. From the data generated, statistical analysis was conducted to determine the prevalence of risk factors of HCC and the parameters generated were included in this model. Other parameters used were generated from published data and other secondary sources. These included: - crude birth rates, crude death rates, prevalence of liver cancer, survival

and recovery rates from liver cancer. These are represented in the diagram in figure 1.

3. Theoretical framework:- Model Description and Analysis

3.1 I.P.M and S.I.R Models

In modeling of disease states, a basic procedure in modeling of disease states is to use a compartmental model where a population is divided into groups. The S.I.R model divides the groups in to susceptible, infected and recovered and is useful in epidemiology and prediction of disease [4, 24]. Projected liver cancer burden in this community was done using a modified SIR + IPM model. The model describes a population in two states; as being diseased or susceptible. The model then allows for calculation of burden of HCC disease.

Ordinary differential equations (ODE) of the infectious risk factors were then generated. Morbidity and mortality indicators of liver cancer were obtained from literature and from the study and these included, crude birth rate, crude death rate, mortality due to HCC, mortality due to other diseases, survival and recovery rates from HCC, incidence, prevalence, susceptible population, infected population and recovered subjects from HCC. These are represented in diagram in figure 1.

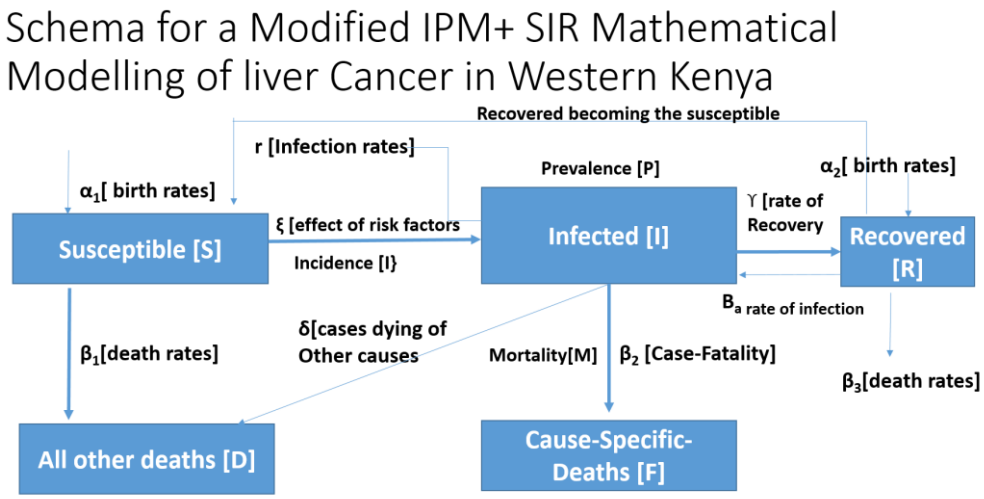

Figure 1.0. A schema for modified model combining S-I-R and I.P.M model concepts

In this model, the population is broken into three groups: $S(t)$, the number of susceptible people, $I(t)$, the number of infected people, and $R(t)$, the number removed through death or recovery. In this scenario, time (t) is the number of years since the epidemic was discovered in western Kenya with a population of 11,347,638 people [16].

This, keeping in assumptions that some of the susceptible people in the model may die because of other causes other than the liver cancer, we also isolate the cause-specific deaths of the infected individuals.

The deterministic model is:

- $\frac{dS}{dt} = S(\alpha-\beta) - \mu I$ (1)
- $\frac{d\,I}{dt} = \mu I - I\,(r + \beta)$ (2)
- $\frac{dR}{dt} = rI + R\,(\alpha + \beta)$ (3)

Where S, I and R are the numbers of susceptible, infected and recovered patients from liver cancer. The parameters r and μ respectively control the rate of infection and rate of recovery. The deterministic solution was calculated for the initial conditions of 1 infected and **257** susceptible using the ode45 integrator in MatLab

The parameters were represented by the symbols included which are:-

, α, μ, S, I, R, I, P, M, D

Description of Parameters
S(t)- Susceptible Individuals at time t
I(t) Infected Individuals at time t
R(t) Recovered Individuals at time t
D(t) Dead Individuals due to all causes at time t
F(t) HCC specific deaths at time t
α(t)- number of births at time t
β_1 deaths rates amongst the susceptible population
β_2 deaths rates amongst the infected population
β_3 deaths rates amongst those who have recovered
δ death rates amongst infected due to non-HCC causes
ϒ rate of recoveries

Analysis was done by utilizing first order ordinary differential equations (ODE) in MatLab which were generated to model liver cancer in western Kenya. These deterministic models were then compared with discrete stochastic models in R that rely upon bio-statistical techniques to model liver cancer and it's predicted burden.

3.2 Parameterisation
While models can provide us with a deeper understanding of the attack and control of non-communicable disease, to be applied more specifically to liver cancer, there is need to parameterise our models to match the observed behaviour

of the disease. Therefore, with the known values for S, I and R including their rates, the future disease occurrences in the population was predicted. In this case, the following in the model was applied.
I(1)=0.43; % As from the statistical analysis
R(1)=0.31; % This changes for every HCC, BCLC class i.e 0.0305 or between 0.06 to 0.07
S(1)=0.059; % As from the statistical analysis
D(1)=0.69; % This also changes depending on the survival rates (which depends on the HCC BCLC class).

Table 1. List of baseline parameters

Symbol	**Description**	**Value**	**More information**
t	Time	Years	Number of years
S(t)	Susceptible Individuals at time t	0.059%	Empirical statistical analysis
I(t)	Infected Individuals at time t	0.43%	This changes for every HCC, BCLC class i.e 0.0305 or between 0.06 to 0.07
R(t)	Recovered Individuals at time t	0.31%	Empirical statistical analysis
D(t)	Proportion of dead Individuals at time t	0.69%	This also changes depending on the survival rates (which depends on the HCC BCLC class)
A	Proportion of susceptible	α = [0.0005].	Proportion of susceptible to cancer
P	Total Population of western Kenya 2016	11,347,638	The population of western Kenya [S(t) + I(t) + R(t)] [20]]
I_n	Total number of Infected cases	257	Empirical statistical analysis
R_n	Total number of infected people recovered	0	A very low value to due unlikely of full recovery
μ	The rate of recovery	0.093	Very low value
R	The rate of infection [incidence]	0.257 per 10,000	Calculated from empirical data
(βs)	Rate of infection from infected population [I]	0.43	

Table 1. (Continued): List of baseline parameters

(βa)	Rate of infection from the recovered population [R]	0.18	Obtained from secondary sources [6]
ξ	The average rate of infection or having HCC (liver cancer) after exposure to risk factors	59%	Effect of risk factors on HCC determined by statistical analysis

TABLE 2: Referral proportions per county of the 257 HCC subjects

County	**HCC Numbers (%)**	**County**	**HCC Numbers (%)**
Kisumu	43 (16.7)	Siaya	40 (15.6)
Vihiga	11 (4.3)	Kakamega	25 (9.7)
Busia	20 (7.8)	Homabay	41 (15.9)
Migori	50 (19.5)	Kisii	10 (3.9)
Nyamira	3 (1.1)	Kericho	5 (1.9)
Nandi	9 (3.5)		

3.3 Empirical Framework: Model Design and Implementation of the integrated S.I.R + I.P.M model

The average rate of infection or having HCC (liver cancer) after exposure to main risk factors is (1.59) 59%. Let the total population be described as one, susceptible fraction of the population (S(t)), infected fraction of the population (I(t)). We know β1= 0.257 i.e. 257 per 10000 individuals is the rate of infection. In the general case the change in the infected population is a product of I and S: We assume initial conditions with one infected individual in Western Kenya where the population is about 11,347,638 in a given year i.e. 365 days.

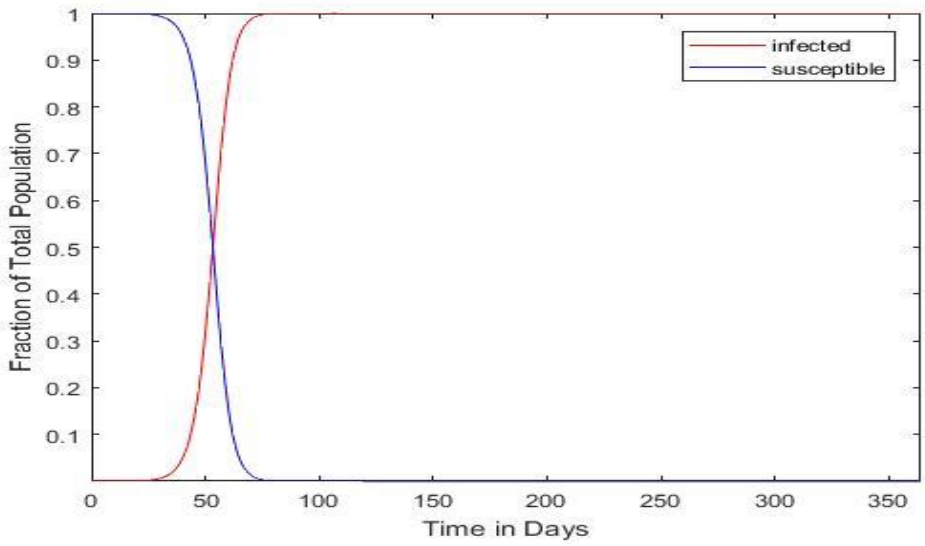

Fig 2. Susceptible and infected fractions of the population over time.

Figure 2 shows that after the first 75 days of exposure in each year to the different risk factors, the entire population moves from the healthy to the infected population. The rate at which the infected population grew was exponential.

3.4 Model Design and Implementation including Recovery Rate.

Let the total population be defined as one, infected fraction of the population (I(t)), infected recovered fraction of the population (R(t)), infected-dead fraction of the population (D(t)), susceptible fraction of the population (S(t)), the rate of infection from the I population (βs) is 0.43, the rate of infection from the R population (βa) is 0.18. The parameter γ is called the removal or recovery rate. For most diseases, the period of infection can be estimated relatively precisely from epidemiological data.

Note that we know that S + I + R = 1, hence knowing S and I will allow us to calculate R.
These equations have the initial conditions S (0) > 0, I(0) > 0 and R(0) = 0.
The MatLab code was used to solve the system of differential equations for a 3 - year duration period of liver cancer progression in the population in western Kenya.

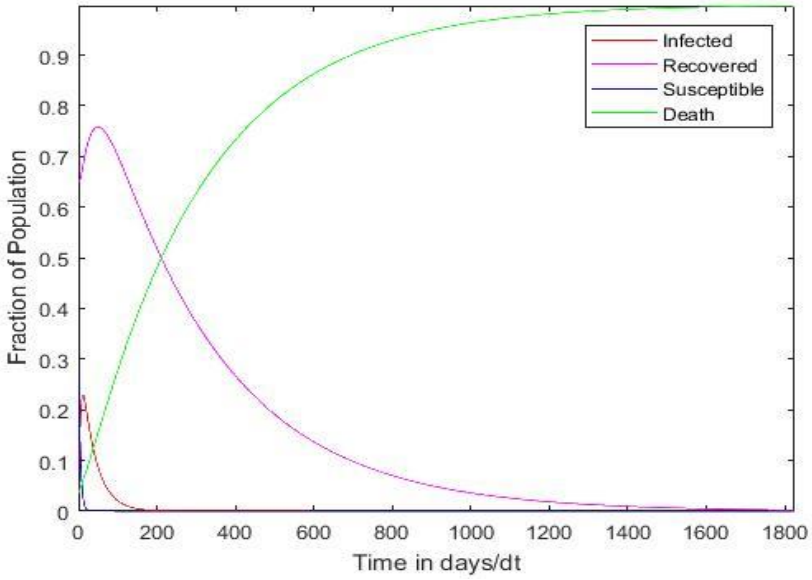

Figure 3: General graphical representation of the model

Figure 3 shows that there is an exponential increase of liver cancer cases over time if there is no control

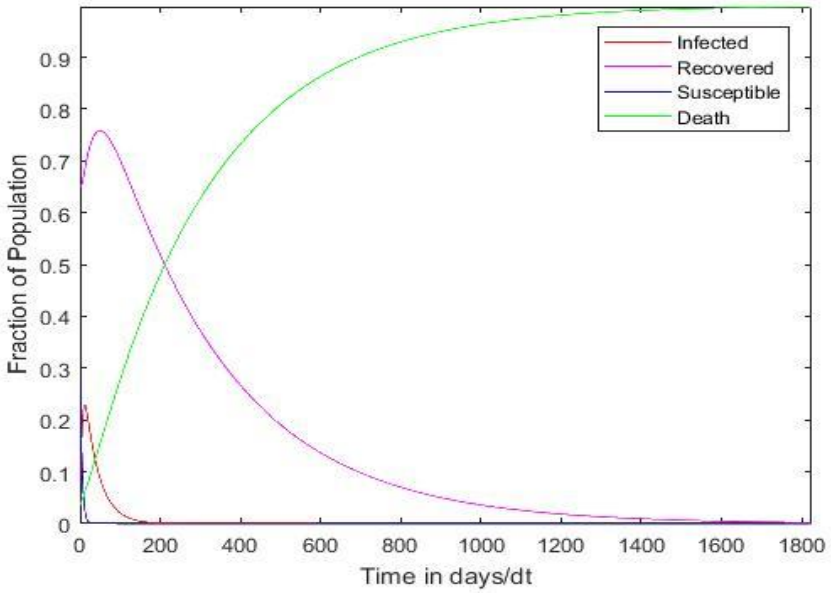

Figure 4: Graphical model display for 5-year localized liver cancer.

(Recovery rate =0.31, Death rate=0.69) [Localized liver cancer is BCLC stages 0, 1 and 2]

The Epidemic graph in Figure 4 depicts the fact that 90% of the population was infected during epidemic. This was because of the high infection rate (β). While the rate of infection in the population was 0.31, the rate of recovery was as low as 0.093. It is very important to implement control measures to reduce infection rate in the population while increasing the recovery rate.

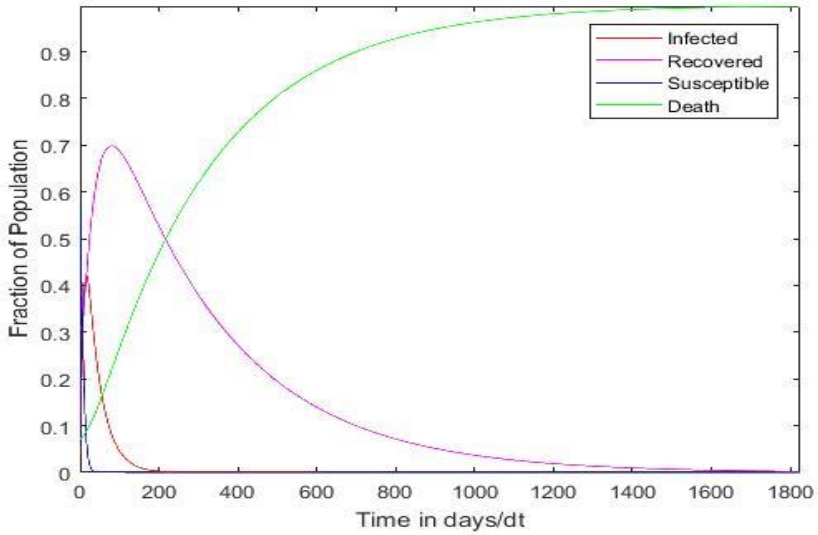

Figure 5: Graphical model display for 5-year Regional Liver Cancer (Recovery rate =0.305, Death rate=0.695) [Regional liver cancer is BCLC stage 3].

Epidemic graph in figure 5 shows epidemic population arises because of more infected people leaving the infective group as compared to the epidemic model.

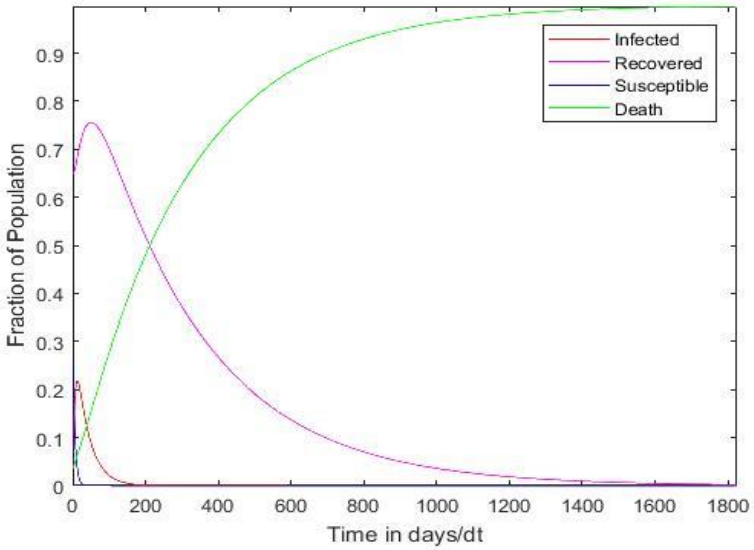

Figure 6: Graphical model display for 5-year distant liver cancer (Recovery rate is between 0.6 and 0.7, Death rate is between 0.4 and 0.3) [Distant HCC is BCLC stage 4]

Epidemic graph in Figure 6, shows that the rate of susceptibility is fading out. This may be attributed to the fact that most people are becoming aware of the cancer prevention measures and are implementing them or there is a high mortality rate from HCC due to the advanced disease and metastases (spread of liver cancer).

3.5 Forecasting future liver cancer burden.

The above models therefore accurately predicts the burden of liver cancer among the population from one year to the next. Two principles are at play. The first principle is that

i.) The disease is spread because of exposure to the risk factors (susceptibility). This principle means that percentage of the population infected will generally increase over time (if the exposure rate is not reduced) at a rate proportional to the amount currently susceptible.
Let's call x_n the proportion of the population that is exposed in year n, where $0 \leq x_n \leq 1$. This first principle suggests that $x_{n+1} = I \cdot x_n$, where I is the disease infection rate with unknown value.
In other words, if I = 3 and 2% of the population is exposed this year, then according to this principle, 6% of the population will be infected next year: $[x_{n+1}] = 3 \cdot 0.02 = 0.06$.

ii.) The second principle in understanding the outbreak is that people who are exposed are more likely to be infected into the next year. This principle suggests that a low infection rate in one year will generally become a higher infection rate in the following year as a considerable proportion of the infected people do not recover but stay long with the disease. Ignoring the first principle, the second principle suggests that $x_{n+1} = R \cdot (1 - x_n)$, where R is a disease recovery rate. If R = 2 and 75% of the population is exposed this year, then according to this principle, 50% of the remaining population will be infected next year: $x_{n+1} = 2 \cdot (1 - 0.75) = 0.50$.
Accounting for both principles at play, the infection rate in year n and number of susceptible, we can accurately predict the infection rate in year $n + 1$ according to: $x_{n+1} = T \cdot x_n \cdot (1 - x_n)$ where T is a factor that combines I and R, the yearly rates of disease Infection and Recovery.

3.5.1 Limitations

There are two unknown parameters about this disease; the first unknown is what proportion of the population is currently susceptible. Let's call this value x_1. The second unknown is the value of T. The values of x_1 and T were therefore determined.

3.6 Summary

The total population of Western Kenya is 11,347,638 individuals [20].

Then S(t) + I(t) + R(t) = 11,347,638. This is the original population of western Kenya. Since every person is exactly one of susceptible, infected, or recovered/died. This does not depend on t.

Since infections emerge from lifestyle malpractices among susceptible, it's believed that the number of susceptible will decrease over time at a rate proportional to the number of susceptible and proportional to the type of lifestyle they are leaving, with a constant of proportionality – α (α =proportion of susceptible), where $\alpha > 0$ [0.0005].

Therefore; $dS/dt = -\alpha S(t)I(t)$.

Naturally, people are "removed" i.e. die at a rate proportional to the number of people infected, with a constant of proportionality μ (μ =death rate [0.12] hence we have;

$dR/dt = \mu I(t)$.

Assuming the total number of people being counted is always the original population of the western Kenya community, differentiating the equation gives;

$dS/dt + dI/dt + dR/dt = 0$.

And so; $dI/dt = \alpha S(t)I(t) - \mu I(t)$.

With the solution of these differential equations, sequences can be used to approximate the solutions, as in Euler's method.

Let $S_n = S(n)$, $I_n = I(n)$, and $R_n = R(n)$, that is, the subscript of the sequence denotes the number of years that have passed.

We can approximate dS/dt by setting the expression we found for it equal to $\Delta S/\Delta t$.

Using this approximation, along with $\Delta S = S_{n+1} - S_n$, and a step-size of $\Delta t = 1$ year, to find a recursive formula for Sn+1 in terms of S_n and I_n.

Therefore; $S_{n+1} = S_n - aS_nI_n$.

Using $dI/dt \approx \Delta I\ \Delta t$, $\Delta I = I_{n+1} - I_n$, and a step-size of $\Delta t = 1$ year, to find a recursive formula for In+1 in terms of S_n and I_n.

Therefore; $In+1 = I_n + \alpha S_nI_n - bI_n$

Using $dR/dt \approx \Delta R\ \Delta t$, $\Delta R = R_{n+1} - R_n$, and a step-size of $\Delta t = 1$ year, to find a recursive formula for Rn+1 in terms of R_n and I_n.

We find; $R_{n+1} = R_n + bI_n$

We therefore estimate that if no interventions are put into place, a = 0.0005 and b = 0.12 then the following table shows the progression of the liver cancer in western Kenya.

Table 3. Progression of liver cancer burden over 3 years.

n(year)	S_n	I_n	R_n	Susceptibility Rate	Infection Rate	Recovery Rate
0	11347638	257	0	0.000617	0.00226	0.0000
1	11347381	10232	32	0.002265	0.09017	0.0003
2	11337149	9302	17	0.090171	0.08205	0.0001
3	11327847	6135	98	0.082049	0.05416	0.0009

Table 3 shows that as the recovery rate changes due to an intervention in the population to manage liver cancer to improve the well-being of the HCC patients, the burden of HCC decreases.

4.0 Determination of the Effectiveness of the Model

This uses techniques from Markov chains, applied to epidemic modelling that may allow for more detailed analysis of epidemics.

i) Markov Chains for Stochastic Epidemic Models

All but the simplest epidemic models are nonlinear, preventing direct analytic computation of the time dependent probability distribution [27]. However, numerical solutions can be computed for systems of moderate size (100's or 1000's individuals).

A continuous time Markov chain is a discrete set of states, known as the state space, and a set of rules for transitions among the states. The Markov property means transitions between states occur randomly and depend only on the previous state.

Let X_t be a random variable that takes a value on the state space and represents the state of the Markov chain at time t then,

$P(X(t_n) = j \mid X(t_{n-1} = i_{n-1}, ..., X(t_1) = i_1) = P(X(t_n) = j \mid X(t_{n-1} = i_{n-1})$

for all sequences, $\{i_n\}$ in the state space and all **times** $0 \leq t_1 < t_1 < ... < t_n$**.**

Deterministic model-These are models in which outcomes are precisely determined through known relationships among states and events, without any room for random variation. In such models, a given input will always produce the same output, such as in a known chemical reaction or disease state. Therefore,

The deterministic model is:

$$\frac{dS}{dt} = a * I - r * S * I$$

$$\frac{dI}{dt} = r * S * I - a * I$$

A stochastic model is a tool for estimating probability distributions of potential outcomes by allowing for random variation in one or more inputs over time. The random variation is usually based on fluctuations observed in historical data for a selected period using standard time-series techniques [4].

The stochastic model is:

$$S + I^{[f(r)]} = 2 * I$$

$$I^{[f(r)]} = S$$

ii) Conditional Probability

One adjustment that needs to be made is to recognize that the stochastic simulation will give a significant probability of approximately 26%, of there being

no liver cancer. Of course, the premise is that an outbreak did occur. To correct for this, we look not at the final simulated probability distribution but at the final simulated conditional probability distribution, that is, the distribution of possible numbers of infected given that there is at least one infected. One computes the conditional probability thus:
P (k infected / 1 or more people are infected) = P (k infected)/P (1 or more people are infected)
where the probability that one or more people are infected is *1-P (no susceptibility occurs)*

4.1 Disease Progression using stochastic Model

The disease progression model simulate the transition from exposed to infected, to liver cancer, and also the transition from liver cancer to recovered in the model. The full structure of this module function, progress, is demonstrated in figure 7. Instead, both progression events are individual-level stochastic processes, in which there is a constant hazard of transition. The times spent in each disease compartment therefore follow a geometric distribution. The code here is for the transition between exposed to infected, to liver cancer states. For each susceptible person for transition, the event is modelled as a stochastic process following a Bernoulli distribution with the rate parameter, i.e. rate in R. Model parameterization was appropriately done.

4.2 Simulation

The model was simulated over 1000 time steps 10 times and is depicted in the figure 7 below.

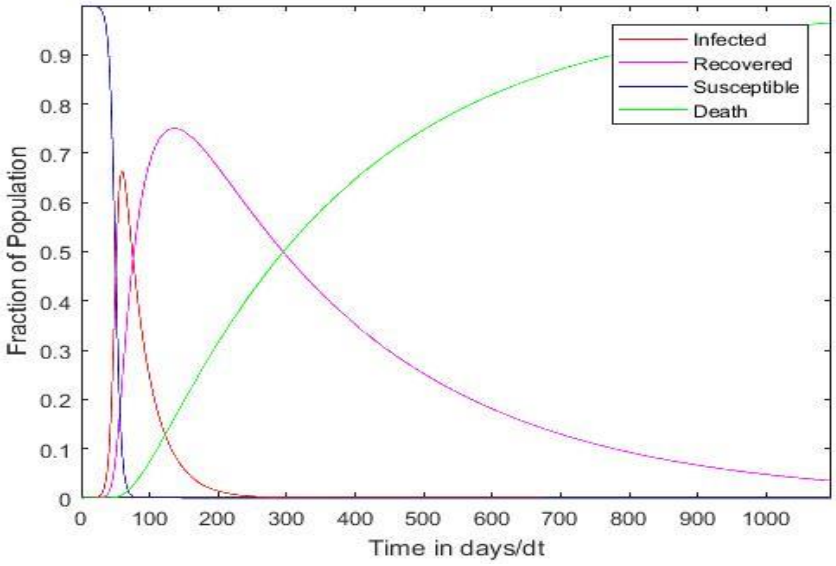

Figure 7. Simulation of the SIR model of HCC in western Kenya.

4.3 Model Comparison

The comparison shows that the ODE and stochastic models have complementary strengths. The stochastic models account for individual interaction and provide the entire probability distribution output. For N individuals, the state space in the stochastic model has N+1 elements, the state space in the stochastic SIR model requires (N+1) (N+2)/2 elements.

The two models, generally, to a greater extent generate the same trend and the same curve, which confirms that the ODE model is good enough for the progression of the liver cancer in such a case in the community.

4.4 Model Comparison against the predictive stochastic model

4.4.1 Comparing the accuracies of the two Models

A statistical test was conducted to assess whether the model had a better accuracy than a more complex stochastic model using a 10-by-10 repeated cross-validation *t* test. Predicted parameters were loaded as a data set then a cost matrix was created. Two ECOC templates were then created: one that uses linear SVM binary only and one that uses SVM binary equipped with the RBF kernel. C1(model 1) and C2 (model 2) are ECOC template objects. C1 is prepared for linear SVM. C2 is prepared for SVM with an RBF kernel. The null hypothesis was then tested that the model (C1) is at most as accurate as the more complex stochastic model (C2) in terms of classification costs. The 10-by-10 repeated cross-validation test was then conducted. Request to return *p*-values and misclassification costs was done.

The following output was achieved using Matlab code.

Parameter	Count	Percentage (%)
Susceptible rate	4	33.33
Infection rate	4	33.33
Recovery rate	4	33.33

4.4.2 Explanation

The empirical distribution of the classes is uniform, and the classification cost is slightly imbalanced. The *p*-value is slightly greater than 0.10, which indicates to retain the null hypothesis that the model is at most as accurate as the more complex stochastic model. This result is consistent for any significance level (Alpha) that is at most 0.10. e_1 and e_2 are 10-by-10 matrices containing misclassification costs. Row *r* corresponds to run *r* of the repeated cross validation. Column *k* corresponds to test-set fold *k* within a cross-validation run. Matlab code were used.

4.4.3 Comparing the Predictive Accuracies Between the two Models

The two models were tested to determine whether they have equal predictive accuracies. The top predictors were identified in terms of their importance.

Explanation
In this case testckfold treats classification models as templates, and so it ignores all fitted parameters in C. That is, testckfold cross validates C using only the specified options and the predictor data to estimate the out-of-fold classification losses.
h = 0 indicates to not reject the null hypothesis that the two models have equal predictive accuracies. This result favours the hybrid S.I.R + I.P.M model developed, though the results can vary and this shows that the models are sensitive in accurately predicting burden of liver cancer.

5. Discussion

In this study, a hybrid S.I.R-I.P.M model was developed for estimating liver cancer burden because there was only a small liver cancer data set to explore. The number of subjects who will develop HCC over time and at different recovery rates of the disease was estimated. It is worth noting that liver cancer starts from a segment of the liver and gradually spreads over time. The extent and severity of damage to the liver and effects on it's functions depends on how severely the lobes of the liver are affected. Ordinarily, without intervention, the liver cancer will progress and the patient will die. Majority of liver cancer cases die because of lack of intervention and they also present late to the hospital for medical care.
The model demonstrates that the future estimated burden of liver cancer, in table 3, shows an increase of cases over time if there is no intervention. Such an increase might be convincing as liver cancer incidence in developing countries is expected to rise principally due to lack of multiple intervention processes like lack of awareness in the public sector, lack of HBV vaccination for adults and long life patients with HIV have while on HAART [13, 14].

Quantitative descriptions of disease epidemiology, such as incidence, prevalence and mortality, by age and sex, are essential inputs for burden of disease studies and cost effectiveness analyses of interventions. Such studies serve as an important source of information for policy-making, planning, and research prioritization in health care. Estimation of disease burden due to liver cancer is therefore feasible by modeling methods.

Making future prediction of liver cancer incidence in western Kenya is difficult due to the lack of population based liver cancer data and registry. In this study, we have used the data of liver cancer from subjects who were referred to the tertiary health facility from lower levels of health care service delivery for specialized care to estimate the future burden of disease in the next 1, 2 and 3 years respectively. It demonstrates that there is an exponential increase in burden of liver cancer. The projected rates indicate that liver cancer in western Kenya will be increasing over time unless it is controlled. Indeed, liver cancer is referred to as a recalcitrant cancer, which is defined as those cancers that have 5-year relative survival rates below 50%. HCC progression in this study was done over three years

because most patients with HCC die within 3-6 months of diagnosis [8, 17] especially where there is no intervention like Kenya. Adequate attention therefore needs to be given to liver cancer to reduce morbidity and mortality. This will need the attention of policy makers at the national and county governments to enact appropriate policies to manage HCC. I.P.M models have been used frequently both to supplement missing data and to study the agreement between different epidemiological data [1, 5, 9, 12].

Suffice it to mention that, while the cancer registry in KEMRI is not well managed and kept, the health sector strategy plan 2015-2021 too has very minimal emphasis on management and prevention and liver cancer [10].

In other contexts, mathematical models have been used to predict the outcome in cancer treatment. This was shown by researchers to be more accurate than doctors when predicting cancer treatment outcomes [21]. At present, prediction models are not used as widely as they could be by doctors [21]. Some studies have however suggested that some models lack clinical credibility while others have not yet been tested. For doctors to have confidence in models, they need to be available and easy to use. Many doctors still think that seeing a patient gives them information that cannot be captured in a model and a study conducted showed that it is very unlikely that a doctor can outperform a model [21]**.**

This study demonstrates the accurate predictive value of mathematical modeling to understand the burden of liver cancer cases in western Kenya. More focus needs to be put to have a sound scientific strategy and framework to manage liver cancer to improve the efforts from the research community in identifying and preparing for liver cancer management which claim more lives because patients come late for health care services. With the increasing number of cases of HCC, there is need to have a concerted effort from all stakeholders—scientists, researchers, policy makers, clinicians and the public to enable prevention, early diagnosis and management of liver cancer. The government will need to be engaged to avail treatment for both symptomatic and specific care of liver cancer.

The sensitivity analyses of the models also show that the models can effectively be used to predict the number of liver cancer cases over time in a community. The prediction is helpful to understand and assess variation and trends of the disease over the years. This will impact positively on policy generation for management and prevention of HCC.

Due to the limitation of the study design, this estimate may not exactly predict the way the population based future estimates would have. However, it provides useful information about the possible exponential increase in burden of liver cancer particularly in western Kenya. This estimation will help in planning for and allocating resources for liver cancer treatment and control in the region.

In conclusion, this study establishes the applicability and effectiveness of the hybrid S.I.R- I.P.M model for predicting burden of liver cancer. Liver cancer rates in western Kenya will increase over time unless controlled and this can inform management and prevention strategies.

6. Recommendations

Population based liver cancer data is lacking and is important in planning for managegemnt of liver cancer in the community. It is therefore important to:

i.) Generate a policy to support research in liver cancer in the community

ii.) Generate a policy to enable availability of treatment options for different stages of liver cancer

iii.) To build capacity of the health care workers to early detect HCC and appropriately manage it, and

iv.) Create awareness on the importance and value of using models in predicting the burden of liver cancer.

Acknowledgements. Sincere acknowledgements to: National Commission for Science, Technology and innovation (NACOSTI) -Kenya for funding this study. Ethics review board of Jaramogi Oginga Odinga Teaching and referral hospital (JOOTRH); Most sincerely, to the patients who consented and participated in the study. Medical superintendant Kisumu county and referral hospital for permitting the studies to be carried out. Dr. Kouko J.O., Dr. Otieno W., and Dr. Reddi P. all of whom are consultant radiologists, for the radiology work. Laboratory in-charges at cell-path and Kisumu county hospital for biochemistry, immunology and haematology results. The clinical officers and nurses in the medical wards, medical outpatient clinic and liver clinic Kisumu county and referral hospital who assisted in counselling, specimen collection and management of patients. Colletta for secretarial work and insight.

Conflict of Interest: There is no conflict of interests. The investigator is a recipient of research funds from National Council of Science, Innovation and Technology and has no conflict of interest (NACOSTI).

Authors Contributions

Otedo Amos designed the study, carried out the research and analysed and drafted the manuscript. Estambale Benson, Omolo Ongati and Simbiri Kenneth participated in the study design and technical input and mathematical modelling advice by Dr. Vincent Were. All authors read the manuscript.

References

[1] P. Albertsen, The use of models in the estimation of disease epidemiology, *J. Urology*, **170** (2003), 697-699.

[2] R.M. Anderson and R.M. May, *Infectious Disease of Humans: Dynamics and Control*, Oxford: Oxford University Press, 1991.

[3] J.J. Barendregt, C.A. Baan and L. Bonneux, An indirect estimate of the incidence of non-insulin-dependent diabetes mellitus, *Epidemiology*, **11** (2000), 274-279. https://doi.org/10.1097/00001648-200005000-00008

[4] Chris Bauch and David Earn, Interepidemic Intervals in Forced and Unforced SEIR models, *Dynamical Systems and Their Applications in Biology*, Ed. Shigui Ruan, Gail S. Wolkowicz and Jianhong Wu. New York: American Mathematical Society, 2003, 33-43. https://doi.org/10.1090/fic/036/03

[5] F. Berauer, P. Van Den Driessche, Some directions for mathematical epidemiology, *Fields Inst. Commun.*, **36** (2003), 95-112.

[6] CIA, World factbook, Major Infectious Diseases, Centre for the study of intelligence, Washington DC, (2011).

[7] Umberto Cillo, Alessandro Vitale, Francesco Grigoletto, Fabio Farinati, Alberto Brolese, Giacomo Zanus, Daniele Neri, Patrizia Boccagni, Nela Srsen, Francesco D'Amico, Francesco Antonio Ciarleglio, Alessio Bridda, Davide Francesco D'Amico, Prospective validation of the Barcelona Clinic Liver Cancer staging system, *J. Hepatol.*, **44** (2006), 723–731. https://doi.org/10.1016/j.jhep.2005.12.015

[8] Y. Fong, Ruth L. Sun, William Jarnagin, Leslie H. Blumgart, An Analysis of 412 cases of hepatocellular carcinoma at a western centre, *Annals of Surgery*, **229** (1999), no. 6, 790-800. https://doi.org/10.1097/00000658-199906000-00005

[9] T. Johnson and B. McQuarrie, *Mathematical Modeling of Disease: Susceptible-Infected-Recovered (SIR) Model*, In University of Minnesota, Morris, Math 4901 Senior Seminar, 2009.

[10] Kenya National Cancer Control Strategy (2011-2016), Ministry of public health and sanitation and Ministry of medical services.

[11] M.E. Kruijshaar, J.J. Barendgret and N. Hoeymans, The use of models in the estimation of disease epidemiology, *Bulletin of world Health Organization*, **11** (2002), 274-279.

[12] M.E. Kruijshaar, J.J. Barendregt and L.V. Van de Poll-Franse, Estimating the prevalence of breast cancer using a disease model: data problems and trends, *Population Health Metrics*, **1** (2003), 1-5. https://doi.org/10.1186/1478-7954-1-5

[13] M.H. Lee, H.I. Yang, J. Liu, R. Batria-Utermann, C.L. Jen, U.H. Iloeje, S.N. Lu, S.L. You, L.Y. Wang and C.J. Chen, Prediction models of long term cirrhosis and hepatocellular carcinoma risk in chronic hepatitis B patients: Risk scores integrating host and virus profiles, *Hepatology*, **58** (2013), 546-554. https://doi.org/10.1002/hep.26385

[14] Lola Rahib, D. Benjamin Smith, Rhonda Aizenberg, B. Allison Rosenzweig, M. Julie Fleshman and M. Lynn Matrisian, Projecting Cancer Incidence and Deaths to 2030, The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States, *Cancer Research*, **74** (2014), no. 11, 2913-2921. https://doi.org/10.1158/0008-5472.can-14-0155

[15] C.D. Mathers, E.T. Vos, C.E. Stevenson and S.T. Begg, The Australian Burden of Disease Study, measuring the loss of health from diseases, injuries and risk factors, *Med. J. Aust.,* **172** (2000), 592-596.

[16] S. Mukerji, *Dynamics of Population and Family Welfare*, (Eds. Srinivasan K and Pathak KB), New Delhi: Himalaya Publishing House, 1989.

[17] A. Musibi, H. Saidi, E. Njuguna, I. Weru, A. Amayo, N. A. Othieno-Abinya and I. Odongo, *National Guidelines for Cancer Management Kenya*, August 2013, 61-65.

[18] C.J.L Murray and A.D. Lopez, *The Global Burden of Disease, a Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020*, Cambridge: Harvard University Press 1996.

[19] C. J. Murray and A. D. Lopez, Quantifying disability: data, methods and results, *Bull. World Health Organ.*, **72** *(*1994), 481-494.

[20] NASCOP KAIS 2013, Kenya AIDS indicator survey 2012, Preliminary report, Nairobi 2013.

[21] Oberije, Mathematical models out-perform doctors in predicting cancer patients' responses to treatment. European society for Radiotherapy and oncology (ESTRO) conference; 19 April, 2013.

[22] D.M. Parkin and L.M. Fernández, Use of Statistics to Assess the Global Burden of Breast Cancer, *Breast J.*, **12** (2006) no. s1, S70-80. https://doi.org/10.1111/j.1075-122x.2006.00205.x

[23] Pierrre Haulnut and Peter Boyle, Curbing liver cancer epidemic in Africa, *Lancet*, **371** (2008), 367-368. https://doi.org/10.1016/s0140-6736(08)60181-6

[24] A. John Rhodes and S. Elizabeth Allman, *Mathematical Models in Biology, An Introduction*, New York, Cambridge UP, 2003. https://doi.org/10.1017/cbo9780511790911

[25] B. Sathian, C.R. Bhatt, S. Jayadevan, J. Ninan, N.S. Baboo and G. Sandeep, Prediction of Cancer Cases for a Hospital in Nepal, A Statistical Modelling, *Asian Pacific J. Cancer Prev.,* **11** (2010), 441-445.

[26] Shanmugaratnam, Cancer registration: Principles and methods, IARC scientific publication, no. 95, Jensen OM et al., (Eds) Chap 1, 1991.

[27] B. I. Strehler and A.S. Mildvan, General theory of mortality and aging, A stochastic model relates observation on aging, physiologic decline, mortality and radiation, *Sci. See Saiensu,* **132** 31-DEC-60.

Received: November 15, 2017; Published: November 28, 2017