

## **Parameter Identification for the Gumel-Mickens HIV Transmission Model under Missing Observable Data**

**Mercy Ngungu**

*Department of Mathematics and Statistics,  
Faculty of Science, Tshwane University of Technology,  
175 Nelson Mandela Drive, Private Bag 680,  
Pretoria 0001, Republic of South Africa.*

**Cliff Richard Kikawa<sup>1</sup>**

*Department of Mathematics and Statistics,  
Faculty of Science, Tshwane University of Technology.*

**Andrew Chikondi Mkolesia and Michael Yu Shatalov**

*Department of Mathematics and Statistics,  
Faculty of Science, Tshwane University of Technology,  
175 Nelson Mandela Drive, Private Bag 680,  
Pretoria 0001, Republic of South Africa.*

### **Abstract**

In this paper, based on the model proposed by Gumel, Moghadas and Mickens, which monitors the impact of live attenuated HIV vaccines, a new method for restoring the unknown data is proposed. Missing data in both epidemiological and ecological surveys and experiments can be due to a number of causes, like, unprovided domains of required data, complete refusal by respondents etc. Differential methods are used to linearize the non-linear equations for the model as the parameter space of the original models is increased. Least squares methods are employed to estimate the unknown model coefficients. Results show that the proposed method is able to restore the missing data with an acceptable accuracy. The efficiency of the proposed method is verified by comparing the restored and observed data using the absolute percentage error. The proposed method can be used in practical research

---

<sup>1</sup>Corresponding author.

and studies to provide a clue on which data to analytically restore in case of missing information.

**AMS subject classification:** 97M10; 97N20; 93C41; 93C10; 93C15.

**Keywords:** Epidemiology, Ecology, Infectious diseases, Parameter estimation, Transmission dynamics.

## 1. Introduction

Communicable diseases such as Human immunodeficiency virus (HIV), mumps, measles, rubella (German measles), smallpox, or malaria, are a fact of modern day life. The dynamics of transmission of infections is currently well known for majority of these diseases. On the overall, the mentioned diseases are transmitted by viral agents and confer immunity against reinfection. Other diseases such as tuberculosis, meningitis, and gonorrhoea are transmitted by bacteria and confer no immunity against reinfection [1].

Viruses are the most common and harmful microorganisms that cause severe diseases to human and other species. Influenza commonly known as flu which probably no person can easily run away from, is caused by viruses [2]. HIV is another deadly virus that spreads mainly through sexual contacts and causes AIDS [3].

Bacteria, (single-celled) organisms are well known microbes that cause a number of diseases. Even though, most of them are harmless and some are even beneficial to human beings. They are of importance when producing cheese, yogurt and medicines. They also have a great deal in synthesizing food particles in human intestine to produce energy [4].

### 1.1. Mathematical models

For many years, mathematical models have been employed to explore the transmission dynamics of infectious diseases. They have become important tools in analyzing the diseases spread and control of infectious diseases. Mathematical modeling is a mathematical process for studying phenomenon of infectious disease transmission. It can assist in building a relation that helps to predict and control further occurrences of the epidemic [5].

Much as it is believed that Daniel Bernoulli in 1760 formulated the first model for monitoring of smallpox outbreak and confirmed that vaccination against the disease can help improve life expectancy, deterministic epidemiology modeling seems to have started in the 20th century [6].

## 2. Problem

It is known that diseases modeling can contribute to the design and analysis of epidemiological surveys, propose important data that has to be collected, identify trends, make

predictions, and estimate the uncertainties in forecasts. However, obtaining complete information in a proposed HIV/AIDS survey is not always the case, as missing cases usually result. These may be due to poor questionnaire design usually too long, ambiguous, personal and embarrassing questions, missing domains for guided questions e.t.c. In that case, information regarding some variables is missing and subsequently not known. The problem of this research is to propose a mathematical procedure for estimating parameters of an HIV model with information concerning some variables not known.

### 3. The Gumel-Mickens Model

Gumel et al. [7] considered a five-dimension model that was presented in [8].

$$\frac{dX}{dt} = p_1 - \mu_1 X - \alpha_v Y_v \frac{X}{N} - \alpha_w (Y_w + Y_{vw}) \frac{X}{N}, \quad (1)$$

$$\frac{dY_v}{dt} = p_2 + \alpha_v Y_v \frac{X}{N} - \gamma (Y_w + Y_{vw}) \frac{Y_v}{N} - \mu_2 Y_v, \quad (2)$$

$$\frac{dY_w}{dt} = \alpha_w (Y_w + Y_{vw}) \frac{X}{N} - \mu_3 Y_w, \quad (3)$$

$$\frac{dY_{vw}}{dt} = \gamma (Y_w + Y_{vw}) \frac{Y_v}{N} - \mu_4 Y_{vw}, \quad \text{where} \quad (4)$$

$N = X + Y_v + Y_w + Y_{vw}$  is the total (sexually active) population size.

The model monitors four populations namely:  $X$  is the HIV susceptibles,  $Y_v$  is the population uninfected by wild type, but infected by the vaccine strain,  $Y_w$  is the unvaccinated population infected by wild type, and  $Y_{vw}$  is the population infected by both vaccine and wild strain. The coefficients  $p_1 = (1 - p)\pi$ ,  $p_2 = p\pi$ , where  $p$  is the proportion of the susceptibles vaccinated,  $\pi$  is the recruitment rate of susceptibles;  $\mu$  is the natural cessation of sexual activity,  $\alpha_1 = c\beta_r$ , where  $c$  is the number of sexual partners,  $\beta_r$  is the rate of transmission of HIV vaccine strain,  $\alpha_w = c\beta_w$ , where  $\beta_w$  is the rate of transmission of HIV wild strain,  $\alpha_w = c\beta_w$ , where  $\beta_w$  is rate of transmission of HIV wild strain,  $\gamma = (1 - \psi)c\beta_w$ , where  $\psi$  is the degree of protection against wild type,  $\mu_2 = \mu_1 + d_1$  where  $d_1$  is death rate due to infection with vaccine strain.  $\mu_3 = \mu_1 + d_2$ , where  $d_2$  is the death rate due to infection with wild strain,  $\mu_4 = \mu_1 + d_3$ , where  $d_3$  is death rate due to infection of both vaccine and wild strains.

Work done by [9] showed that all parameters of the system of Equations (1-4) can be estimated when incomplete data is available about  $N = N(t)$ ,  $X = X(t)$  and  $Y_v = Y_v(t)$ . In this case it is possible to distinguish between  $Y_w = Y_w(t)$  and  $Y_{vw} = Y_{vw}(t)$  populations. Hence, the complete problem of identification and prediction of healthy and infected populations can be solved.

In this study, it is shown that the analogous problem of parameter determination and population identification can be solved if incomplete information is available on other

different groups of the population, namely, in the case when  $N = N(t)$ ,  $Y_w = Y_w(t)$  and  $Y_{vw} = Y_{vw}(t)$  is known.

## 4. Method of Solution

### 4.1. Algorithm for the Differentiation approach

Assuming that  $Z = Z(t) = Y_{vw}(t) + Y_w(t)$ , where the new variable  $Z(t)$  represents the sum of both populations infected by vaccine and wild strain  $Y_{vw}$  and infected by vaccine only  $Y_v$ . Hence,  $X(t) + Y_v(t) = N(t) - Z(t)$  are known. Then  $X(t)$  and  $Y_v(t)$  are expressed from Equations (3) and (4) as follows:

$$X = X(t) = \frac{\dot{Y}_w + \mu_3 Y_w N}{\alpha_w} \frac{N}{Z}, \quad (5)$$

$$Y_v = Y_v(t) = \frac{\dot{Y}_{vw} + \mu_4 Y_{vw} N}{\gamma} \frac{N}{Z}. \quad (6)$$

From Equations (5) and (6) it can be deduced that,

$$a_1 \dot{Y}_w + a_2 Y_w + a_3 \dot{Y}_{vw} + a_4 Y_{vw} - (N - Z) \frac{Z}{N} = 0, \quad (7)$$

where,  $a_1 = \frac{1}{\alpha}$ ,  $a_2 = \frac{\mu_3}{\alpha}$ ,  $a_3 = \frac{1}{\gamma}$  and  $a_4 = \frac{\mu_4}{\gamma}$  are new unknown coefficients.

Integrating Equation (7) with respect to  $\tau \in [0, t_k]$  for  $k = 0, 1, 2, \dots, n$ ,  $t_n = T$  and  $t_k \in [0, T]$ . An over determined system of linear algebraic equations is obtained when the number of intervals,  $n > 4$ .

$$a_1 \Delta Y_{w,k} + a_2 I_{1,k} + a_3 \Delta Y_{vw,k} + a_4 I_{2,k} - I_{3,k} = 0, \quad (8)$$

where  $\Delta Y_{w1,k} = Y_w(t_k) - Y_{w,0}$ ;  $\Delta Y_{vw1,k} = Y_{vw}(t_k) - Y_{vw,0}$ ;  $I_{1,k} = \int_0^{t_k} Y_w(\tau) d\tau$ ;  $I_{2,k} =$

$$\int_0^{t_k} Y_{vw}(\tau) d\tau;$$

$$I_{3,k} = \int_0^{t_k} \frac{[N(\tau) - Z(\tau)Z(\tau)]}{N(t)} d\tau.$$

In order to estimate the coefficients  $a_1, a_2, a_3, a_4$  the least-squares method [10], the following goal function subjected to minimization;

$$G_1(\eta) = \frac{1}{2} \sum_{k=1}^n \{\Phi - I_{3,k}\}^2 \rightarrow \min, \quad (9)$$

where  $\eta = (a_1, a_2, a_3, a_4)$ ,  $\Phi = a_1 \Delta Y_{w1,k} + a_2 I_{1,k} + a_3 \Delta Y_{vw,k} + a_4 I_{2,k}$ .

The system of linear algebraic equations to define  $\eta$  is as follows:

$$\sum_{j=1}^4 b_{ij} a_j = \alpha_i, \quad (10)$$

for  $(i = 1, \dots, 4)$  and where

$$\begin{aligned}
 b_{11} &= \sum_{k=1}^n (\Delta Y_{w,k})^2; \quad b_{12} = b_{21} = \sum_{k=1}^n (\Delta Y_{w,k} I_{1,k}); \\
 b_{13} &= b_{31} \sum_{k=1}^n (\Delta Y_{w,k} \Delta Y_{vw,k}); \quad b_{14} = b_{41} \sum_{k=1}^n (\Delta Y_{w,k} I_{2,k}); \\
 b_{22} &= \sum_{k=1}^n (I_{1,K})^2; \quad b_{23} = b_{32} = \sum_{k=1}^n (I_{1,k} \Delta Y_{vw,k}); \\
 b_{33} &= \sum_{k=1}^n (\Delta Y_{vw,k})^2; \quad b_{34} = b_{43} = \sum_{k=1}^n (\Delta Y_{vw,k} I_{2,k}); \\
 b_{44} &= \sum_{k=1}^n (I_{2,k})^2; \quad d_1 = \sum_{k=1}^n (\Delta Y_{w,k} I_{3,K}); \quad d_2 = \sum_{k=1}^n (\Delta I_{1,k} I_{3,K}); \\
 d_3 &= \sum_{k=1}^n (\Delta Y_{vw,k} I_{3,K}); \quad d_4 = \sum_{k=1}^n (I_{2,k} I_{3,K}).
 \end{aligned}$$

On finding the values of  $a_1, a_2, a_3$  and  $a_4$ , parameters  $\alpha_w, \gamma, \mu_3$  and  $\mu_4$  are then computed as:

$$\alpha_w = a_1^{-1}; \quad \mu_3 = a_1^{-1} a_2; \quad \gamma = a_3^{-1}; \quad \mu_4 = a_3^{-1} a_4 \tag{11}$$

Now the unknown population  $X_k$  and  $Y_{v,k}$  can be approximately calculated from the Model Equations (5) and (6):

$$X_k \approx \frac{\dot{Y}_{w,k} + \mu_3 Y_{w,k} N_k}{\alpha_w Z_k}, \tag{12}$$

$$Y_{v,k} \approx \frac{\dot{Y}_{vw,k} + \mu_4 \tilde{Y}_{vw,k} N_k}{\gamma Z_k}. \tag{13}$$

Approximate calculations in this case imply that the corresponding derivatives are computed using any available finite difference schemes.

In order to compute the other unknown parameters  $p_1, p_2, \mu_1, \mu_2$  and  $\alpha_v$ , Equations (1) and (2) are used. After integration, from  $\tau = 0$  to  $\tau = t_4, t_k \in [0, T], k = 1, 2, \dots, n$ . Then the following overdetermined equations are obtained

$$P_1 t_k + \mu_1 I_{4K} + \alpha_v I_{5k} - I_{6k} = 0, \tag{14}$$

$$P_2 t_k + \mu_2 I_{6K} + \alpha_v (-I_{5k}) - I_{8k} = 0. \tag{15}$$

Formulating the second goal function;

$$G_2(\xi) = \frac{1}{2} \sum_{k=1}^n \{ P_1 t_k + \mu_1 I_{4k} + \alpha_v I_{5k} - I_{6k} + \Gamma \}^2 \rightarrow \min, \tag{16}$$

where  $\xi = (P_1, P_2, \mu_1, \mu_2, \alpha_v)$  and  $\Gamma = P_2 t_k \mu_2 I_{7k} + \alpha_v (-I_{5k}) - I_{8k}$ , and minimizing the formulated goal function,

$$\frac{\partial G_2}{\partial P_1} = 0; \quad \frac{\partial G_2}{\partial \mu_1} = 0; \quad \frac{\partial G_2}{\partial \alpha_v} = 0; \quad \frac{\partial G_2}{\partial P_2} = 0; \quad \text{and} \quad \frac{\partial G_2}{\partial \mu_2} = 0,$$

a linear system of equations for finding  $P_{1,2}$ ,  $\mu_{1,2}$  and  $\alpha_v$  is obtained;

$$\sum_{m=1}^5 C_{lm} Z_m = f_l, \quad (17)$$

for  $l = 1, \dots, 5$  and  $Z_1 = P_1$ ,  $Z_2 = \mu_1$ ,

$$Z_3 = \alpha_v, Z_4 = P_2, Z_5 = \mu_2, C_{11} = \sum_{k=1}^n (t_k^2), C_{12} = C_{21} = \sum_{k=1}^n (t_k I_{4k}),$$

$$C_{13} = C_{31} = \sum_{k=1}^n (t_k I_{5k}), C_{14} = C_{41} = C_{15} = C_{51} = 0, C_{22} = \sum_{k=1}^n (I_{4k})^2,$$

$$C_{23} = C_{32} = \sum_{k=1}^n (I_{4k} I_{5k}), C_{24} = C_{42} = C_{25} = C_{52} = 0, C_{33} = 2 \sum_{k=1}^n (I_{5k})^2,$$

$$C_{43} = C_{34} = - \sum_{k=1}^n (t_k I_{5k}), C_{35} = C_{53} = - \sum_{k=1}^n (I_{5k} I_{7k}), C_{44} = C_{11},$$

$$C_{45} = C_{54} = \sum_{k=1}^n (t_k I_{7k}), C_{55} = \sum_{k=1}^n (I_{7k})^2, f_1 = \sum_{k=1}^n (t_k I_{6k}),$$

$$f_2 = \sum_{k=1}^n (t_{4k} I_{6k}), f_3 = \sum_{k=1}^n I_{5k} [(I_{6k} - I_{8k})],$$

$$f_4 = \sum_{k=1}^n (t_k I_{8k}), f_5 = \sum_{k=1}^n (I_{7k} I_{8k}).$$

Having obtained the parameters  $P_{1,2}$ ,  $\mu_{1,2}$ ,  $\alpha_v$ ,  $\alpha_w$  and  $\gamma$ , it is then possible to obtain the initial parameters

$$I = P_1 + P_2, P = \frac{P_2}{P_1 + P_2}, \psi = 1 - \frac{\gamma}{\alpha_w}, \mu = \mu_1, d_1 = \mu_2 - \mu_1, d_2 = \mu_3 - \mu_1,$$

$$d_3 = \mu_4 - \mu_1, \frac{\beta_v}{\beta_w} = \frac{\alpha_v}{\alpha_w}.$$

It can be concluded that knowledge of all parameters and initial conditions makes it possible to predict both quantitatively and qualitatively the dynamics of all population groups in the habitats;

$$X = X(t), Y_v = Y_v(t), Y_w = Y_w(t) \text{ and } Y_{vw} = Y_{vw}(t) \text{ for } t > T.$$

### 5. Numerical Simulation

On the first stage, we assume that all parameters and initial conditions of the four populations are known and hence the Model (1)–(4), can be solved numerically. The parameter identification for the physical models is based on the results obtained by [7] under the knowledge of all required information;  $P_i = 2000, P = 0.8, C = 5.0, \beta_v = 0.5, \beta_w = 0.45, \psi = 0.6, \mu = 0.031, d_1 = 0.3, d_2 = 0.25, d_3 = 0.2$  Using the information given above, previous parameters are now recomputed into new parameters using the combinations:  $P_1 = (1 - P)P_i, \mu_1 = \mu, \mu_2 = \mu + d_2, \mu_4 = \mu + d_3, P_2 = P P_i, \alpha_v = C\beta_v, \alpha_w = C\beta_w, g = (1 - \psi)\alpha_w$  ( $g$  for gamma). In this study the combinations give what is assumed as experimental or observed data. The right hand side of the Cauchy system ( $u_0 = x, u_1 = Y_v, u_2 = Y_w, u_0 = Y_{vw}$ ) :

$$D(t, u) = \begin{bmatrix} P_1 - [\alpha_v M_1] - [\alpha_w M_2] - \mu_1 u_0 \\ P_2 + [\alpha_v M_1] - [g M_3] - \mu_2 u_1 \\ \alpha_w M_4 - \mu_3 u_1 \\ g M_5 - \mu_4 u_3 \end{bmatrix},$$

where

$$M_1 = \frac{u_0 u_1}{u_0 + u_1 + u_2 + u_3}, M_2 = \frac{u_0(u_1 + u_3)}{u_0 + u_1 + u_2 + u_3},$$

$$M_3 = \frac{u_1(u_2 + u_3)}{u_0 + u_1 + u_2 + u_3}, M_4 = \frac{u_0(u_2 + u_3)}{u_0 + u_1 + u_2 + u_3}, \text{ and}$$

$$M_5 = \frac{u_1(u_2 + u_3)}{u_0 + u_1 + u_2 + u_3}.$$

The initial conditions for the linear minimization are obtained from [7]:

$$u = (80.10^3 \ 2.10^3 \ 8.10^3 \ 8.10^3)^T$$

In this work, the time interval ( $T$ ) is shorter than that employed in [7], in that respect it is different.  $NN$  is the number of interval onto the given time interval.

$$T = 5, NN = 25, i = 0, \dots, NN.$$

In order to obtain the solution of the system of Equations (1)–(4) numerically, Adam’s method in Mathcad® software is employed using the parameter values and initial conditions from [7] as stated earleir.

On the second stage, we forget about the values of the parameters and assume that only  $N$ -total population,  $Y_w$ -population that is infected by virus  $W, Y_{vw}$ -population infected by virus  $V$  are known. At the same time we have knowledge of the sum  $N - Y_w - Y_{vw} = N - Z = X + Y_v = ZZ$ , which is the sum of healthy and virus,  $V$ -infected population.

From

$$N_i = (U^{(1)})_i + (U^{(2)})_i + (U^{(3)})_i + (U^{(4)})_i,$$

$$Z_i = (U^{(3)})_i + (U^{(4)})_i,$$

$$ZZ_i = N_i - Z_i,$$

where  $X = U^{(1)}$ ,  $Y_v = U^{(2)}$ ,  $Y_w = U^{(3)}$  and  $Y_{vw} = U^{(4)}$  are assumed to be the solution vectors obtained from the system of Equations (1-4) and these solutions are taken to be the experimental or observed information regarding the state variables  $Y_w$  and  $Y_{vw}$  and the model parameters to be unknown.

Table 1: Experimental data for the state variables,  $N$ ,  $Y_w$  and  $Y_{vw}$

Case	N	YW	YVW
0	$9.800 \times 10^4$	$8.000 \times 10^4$	$8.000 \times 10^3$
1	$9.679 \times 10^4$	$1.397 \times 10^4$	$7.725 \times 10^3$
2	$9.520 \times 10^4$	$2.086 \times 10^4$	$7.544 \times 10^3$
3	$9.320 \times 10^4$	$2.823 \times 10^4$	$7.493 \times 10^3$
4	$9.078 \times 10^4$	$3.541 \times 10^4$	$7.601 \times 10^3$
5	$8.798 \times 10^4$	$4.170 \times 10^4$	$7.869 \times 10^3$
6	$8.489 \times 10^4$	$4.655 \times 10^4$	$8.269 \times 10^3$
7	$8.159 \times 10^4$	$4.977 \times 10^4$	$8.745 \times 10^3$
8	$7.820 \times 10^4$	$5.143 \times 10^4$	$9.236 \times 10^3$
9	$7.479 \times 10^4$	$5.181 \times 10^4$	$9.689 \times 10^3$
10	$7.143 \times 10^4$	$5.121 \times 10^4$	$1.007 \times 10^3$
11	$6.816 \times 10^4$	$4.995 \times 10^4$	$1.037 \times 10^3$
12	$6.501 \times 10^4$	$4.826 \times 10^4$	$1.058 \times 10^3$
13	$6.200 \times 10^4$	$4.633 \times 10^4$	$1.070 \times 10^3$
14	$5.913 \times 10^4$	$4.428 \times 10^4$	$1.076 \times 10^3$
15	...	...	...

Table 1, shows the values of the populations that are assumed known, as earlier stated in the second estimation stage of model [1-4], by interpolation and subsequent differentiating using the computer algebra system (CAS) Mathcad<sup>®</sup>. Knowing the information given in Table 1, we then use the proposed algorithm in Section 4 to estimate model parameter values that will be employed to restore the unknown information.

Table 2 shows results of the experimental and the estimated data using the proposed method. The absolute percentage error is used to assess the percentage error between the experimental (observed) and the estimated values.

It is evident from Table 2 that the observed and the estimated data are close and for majority of the parameters ideally the same when comparison is based on the absolute percentage error. It should be noted that the parameters coefficients in Table 2 assist in the identification of the individual populations in the sexually active population  $N$ . In this study, however, only parameters that can be used to identify the required populations  $X_k$  and  $Y_{v,k}$  are estimated and shown in Table 2. Nevertheless, other parameter coefficients  $P_1, P_2, \mu_1, \mu_2$  and  $\alpha_v$  can be computed in the same way as indicated in Equation (16). Recall that,  $(U^{(1)})_i$  and  $(U^{(2)})_i$  are the estimated solution vectors of the supposedly unknown populations  $X_k$  and  $Y_{v,k}$ , obtained from solving the model [1-2],



Table 2: Parameter comparison for the observed and estimated values using the absolute % error

Parameter	Observed	Estimated	Abs % Error
$\mu_3$	0.281	0.282	0.356
$\mu_4$	0.231	0.231	0.000
$\alpha_w$	2.250	2.250	0.000
$\gamma$	0.900	0.900	0.000

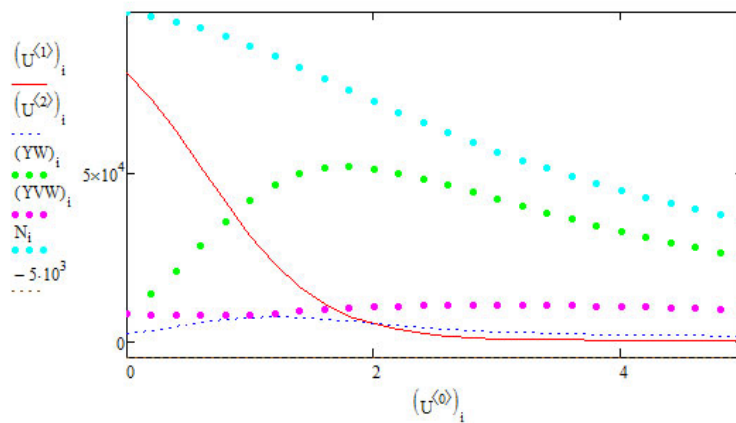


Figure 1: Profiles of  $U^{(1)}$ ,  $U^{(2)}$ ,  $Y_w$ ,  $Y_{vw}$  and  $N$  with parameters and initial values

using Adams estimation method as earlier indicated. In Figure 1, the profiles of all populations that comprise the sexually active population  $N_i$  inclusive are indicated. Further still, from Figure 1,  $(Y_w)_i$ ,  $(Y_{vw})_i$  and  $N_i$  are the solution vectors for the known populations  $Y_w$ ,  $Y_{vw}$  and  $N$  respectively as estimated using the computer Algebraic system in Mathcad software and considered as the observed data for this study, Table 1.

The known populations can further be approximated using:

$$\left. \begin{aligned} dY_w(t) &= \frac{dY_w(t)}{dt} \\ dY_{vw}(t) &= \frac{dY_{vw}(t)}{dt} \end{aligned} \right\}, \tag{18}$$

at  $t = U^{(0)}$ . Further analysis is performed on the generated populations  $dY_w(t)$  and  $dY_{vw}(t)$  at the initial values for  $t = U^{(0)}$ . In this case the trends for the respective populations are ascertained in the absence of the total sexually active population  $N$ . These are also generated using the non-standard method Adams in Mathcad Figure 2.

The generated populations, Equation (18) are then used in Equations (12) and (13) to restore the unknown populations  $X_k$  and  $Y_v$ . Graphs of the restored and observed

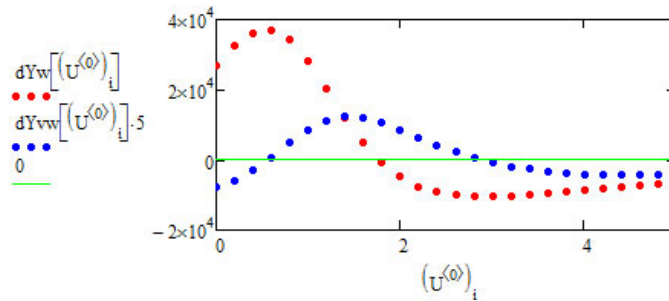


Figure 2: Profiles of  $Y_w$  and  $Y_{vw}$  with initial values  $(U^{(0)})_i$

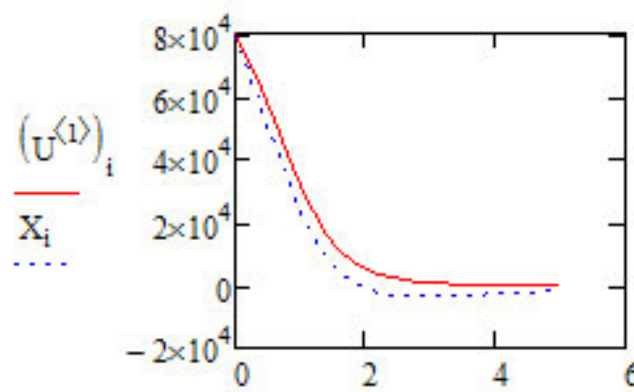


Figure 3: Profiles of observed  $(U^{(1)})_i$  and restored  $X_i$  populations over time,  $t_i$

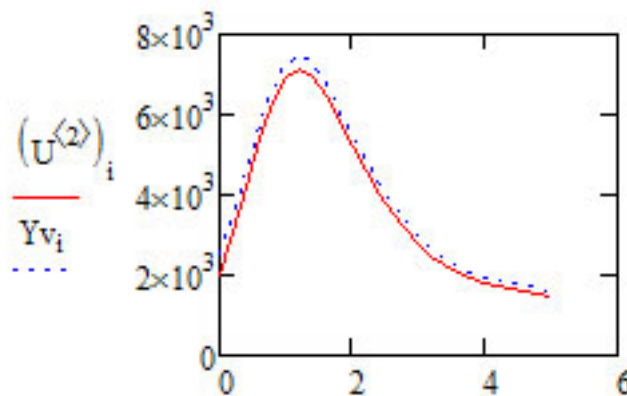


Figure 4: Profiles of observed  $(U^{(2)})_i$  and restored  $Y_{v_i}$  populations over time,  $t_i$

populations are plotted to study their profiles, Figures 3 and 4. It should be noted that  $X_k$  is the same as  $X_i$ , but the latter simply implies estimated in this case. Also  $Y_{v_i}$  implies  $Y_v$  but the former indicates estimated in this study.

From the profiles of Figures 3 and 4, it is observed that the experimental or observed

data *solidline* and restored data *dashedline* from the proposed method, are close and have the same pattern. This is an indication that in the presence of incomplete information, the proposed method can be used to restore the missing information.

## 6. Conclusion

In this study, a four-dimensional HIV transmission model proposed by [7], which monitors the impact of live attenuated HIV vaccines, was used to determine the performance of the proposed method which is capable of restoring missing information in epidemiological studies. Results indicate that the proposed approach is able to analytically estimate the missing data Table 2. In other words, when data in a survey is incomplete due to factors related to those mentioned in Section [2], the method can be employed to analytically restore it and this is shown in the profiles of Figures 3 and 4.

## Acknowledgment

The authors would like to thank Mr Alun Peck for sparing his time to convert the graphics to the required format. The Department of Mathematics and Statistics, Tshwane University of Technology for allowing the researchers use its resources.

## References

- [1] F. Brauer, *Compartmental Models in Epidemiology*. Springer, 2008, ch. Two.
- [2] M. Loeb, L. Russell, and L. Moss, "Effect of influenza vaccination of children on infection rates in hutterite communities: a randomized trial," *JAMA*, vol. 303, pp. 943–950, 2010.
- [3] M. Levine, M. Tapia, V. Hill, and O. Sow, "M.m. levine, m.tapia, a.v. hill, and s.o. sow, 2014. how the current west african ebola epidemic is altering views on the need for vaccines and is galvanizing a global effort to field test leading candidates, journal of infectious diseases, advance access." *Journal of Infectious Diseases*, *Advance access*, 2014.
- [4] M. Wexler, "Bacteroides: the good, the bad, and the nitty-gritty," *Clinical Microbiology Reviews*, vol. 20, p. 593–621, 2007.
- [5] A. Rahman, "Study of infectious diseases by mathematical models: Predictions and controls," Ph.D. dissertation, The University of Western Ontario, 2016.
- [6] W. Hethcote, "The mathematics of infectious diseases," *SIAM REVIEW Vol. 42, No. 4, pp. 599–653*, vol. 42, no. 4, p. 599–653, 2000.
- [7] A. Gumel, S. Moghadas, and R. Mickens, "Effect of a preventive vaccine on the dynamics of hiv transmission," *Communications in Nonlinear Science and Numerical Simulation*, vol. 9, p. 649–659, 2004.

- [8] S. Blower, K. Koelle, D. Kirschner, and J. Mills, “Live attenuated hiv vaccines: predicting the trade-off between efficacy and safety,” *Proc Nat Acad Sci*, vol. 98, no. 6, p. 3618–23, 2001.
- [9] P. Mathye, I. Fedotov, and M. Shatalov, “Identification of hiv dynamic system in the case of incomplete experimental data,” *BIOMATH*, vol. 4, pp. 1–7, 2016.
- [10] C. Kikawa, M. Shatalov, and P. Kloppers, “Estimation of the mean and variance of a univariate normal distribution using least-squares via the differential and integral techniques,” *International Journal of Probability and Statistics*, vol. 4, no. 2, pp. 37–41, 2015.