Transactions of the Royal Society of Tropical Medicine and Hygiene (2006) **100**, 811–816

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevierhealth.com/journals/trst

Distribution of survival times of deliberate *Plasmodium falciparum* infections in tertiary syphilis patients

W. Sama[a,*], K. Dietz[b], T. Smith[a]

[a] *Department of Public Health and Epidemiology, Swiss Tropical Institute, Socinstrasse 57, Postfach CH-4002, Basel, Switzerland*
[b] *Institut für Medizinische Biometrie, Westbahnhofstrasse 55, D-72070 Tübingen, Germany*

Received 12 July 2005; received in revised form 3 November 2005; accepted 3 November 2005
Available online 31 January 2006

KEYWORDS
Plasmodium falciparum;
Survival time;
Distribution;
Quantile–quantile plots;
Neurosyphilis;
Malariatherapy

Summary Survival time data of *Plasmodium falciparum* infections from deliberate infection of human subjects with *P. falciparum* between 1940 and 1963 as a treatment for neurosyphilis in the USA (Georgia) have been used to test the fits of five commonly used parametric distributions for survival times using quantile–quantile plots. Our results suggest that the best fit is obtained from the Gompertz or Weibull distributions. This result has important implications for mathematical modelling of malaria, which has for the past century exclusively assumed that the duration of malaria infections has an exponential distribution. It is desirable to know the correct distribution because its shape profoundly influences the length of monitoring needed in an intervention programme for eliminating or reducing malaria.

1. Introduction

Mathematical modelling of malaria has flourished since the days of Ross (1911), who was the first to model the dynamics of malaria transmission. Ross assumed that as far as infection with malaria is concerned, an individual host could be in only one of two states: either susceptible or infected, and that susceptibles have some constant probability per unit time (the infection rate, λ) of becoming infected, and infected individuals have some constant probability per unit time (the recovery rate or the clearance rate, μ) of recovering.

The events of infection and recovery were assumed to be Poisson processes, in that they occur randomly in time within the population. Hence, it follows from the well known property of the Poisson distribution (Cox and Miller, 1965) that the intervals of time between successive arrivals (and also between arrival and departure) follow an exponential distribution with constant rate λ (and μ, respectively). It is likely that recovery rates are dependent on a number of time-related factors, for instance the age of the infection and the development of host immunity. Nevertheless, this assumption of a constant recovery rate is one of the more prevalent assumptions in epidemic theory as a whole (for example, see: Anderson, 1982; Anderson and May, 1991; Bailey, 1957, 1975, 1982). There is extensive literature on mathematical models of malaria with some fundamental departures

* Corresponding author. Tel.: +41 61 284 82 82; fax: +41 61 284 81 05.
E-mail address: wilson.sama@unibas.ch (W. Sama).

0035-9203/$ – see front matter
doi:10.1016/j.trstmh.2005.11.001

from the view of malaria epidemiology as conceived by Ross (see, for instance, work by Dietz (1988) and the references therein). However, even in models where it is assumed that the duration of an infection depends on the age of the host (an assumption that attempts to incorporate acquired immunity), it is still assumed that the duration at a given age is exponentially distributed (Smith and Vounatsou, 2003).

Clements and Paterson (1981) used a Gompertz distribution for the duration of a malaria vector's life time, i.e. the duration of the infection in the vector, but we are not aware of any published models that have used alternatives to the exponential distribution for the duration of malaria infection in humans, or where this assumption has been tested. There are two major difficulties in collecting empirical data from malaria-endemic areas for use in testing this assumption. When there is an obligation to treat all the infections discovered, this precludes monitoring and limits the possibilities for estimating the duration in non-immune individuals. Individuals in endemic areas are subjected to repeated reinfection and the common diagnostic tool (light microscopy) in the field is not able to differentiate between infections derived at different points in time. PCR-based methods have now demonstrated that individuals indeed harbour multiple infections and such methods are being used to classify malaria infections into different types (Felger et al., 1993, 1999). However, given that even PCR is not 100% sensitive, the continual appearance and disappearance of specific types in consecutive blood samples in longitudinal studies makes it difficult to characterise such types as new or persisting infections.

The duration of infection is most easily estimated from the rate of clearance of parasites from the blood following a single infection. Owing to difficulties in obtaining such data in the field, some studies (Eichner et al., 2001; Molineaux et al., 2001; Paget-McNicol et al., 2002; Recker et al., 2004) have used data (malariatherapy data) from *Plasmodium falciparum* infections deliberately induced as a treatment for neurosyphilis in the mid 20th century in the USA, at a time when there were no antibiotics for the treatment of neurosyphilis (Collins and Jeffery, 1999; Jeffery and Eyles, 1955). Although these data were collected from non-immune individuals in a non-endemic area, they represent an excellent source of data for testing the assumption of an exponential distribution for the survival time.

We studied the fits of five alternative distributions commonly used for survival data as an approximation to the lifetime of malaria infections within the host using malariatherapy data and address the question of the applicability of the results to endemic areas. In the following, we use the word hazard to mean the clearance rate (or recovery rate) of infections.

2. Methods

2.1. Data

Malariatherapy data were collected in the USA (Milledgeville Hospital, Georgia and National Institute of Health Laboratories, Columbia, South Carolina) during 1940—1963, at a time when malariatherapy was a recommended treatment for neurosyphilis (Collins and Jeffery, 1999). Different strains of *P. falciparum* were inoculated with either sporozoites (generally through mosquito bite) or infected blood. Microscopic blood examinations were performed almost daily.

Of a total of 334 patients in our database, 157 were from Georgia and 177 were from South Carolina. The average duration of infections was 135.2 days (standard error (SE) 8.8 days) in Georgia patients and 75.4 days (SE 4.2 days) in South Carolina patients, irrespective of whether they received treatment or the time at which treatment was given. Ninety-nine of 157 Georgia patients received treatment, with a total of 540 days when treatment was given, whilst 116 of 177 South Carolina patients received treatment, with a total of 1030 days when treatment was given. This suggests that infections in patients from the Georgia hospital persisted for longer than those in South Carolina and this appears to reflect more treatment in the latter hospital. Hence, it is more likely that the Georgia infections were more similar to untreated natural infections in a typical malaria-endemic setting in Africa. The period for monitoring after the last positive slide also varied, hence so does the confidence that an infection was spontaneously cleared. For this analysis, we consider only patients from the Georgia hospital who did not receive any antimalarial treatment on their last day of positivity of asexual parasitaemia and were followed up for a qualifying period of at least 60 days after their last positive slide. This consists of 54 patients; 29 of them received the Santee-Cooper strain, 23 received the El Limon strain and 2 received the McLendon strain. No substantial differences in the mean (arithmetic) duration of infections were observed by varying the qualifying period. Among the 54 patients, 23 of them received some subcurative treatment before their last positive slide.

2.2. Distributional assumptions

We consider five different distributions commonly used to model survival data: exponential, log-normal, gamma, Weibull and Gompertz. We focus mainly on the fit of these distributions as an approximation to the lifetime of malaria infections within the host using data from malariatherapy patients. For more on the basic properties of the above distributions, for instance hazard rates, quantiles, see the following publications (Evans et al., 2000; Johnson et al., 1995; Klein and Moeschberger, 1997; Wilk and Gnanadesikan, 1968).

A powerful tool for exploring distributional fit to data is by using the graphical technique referred to as the quantile—quantile plot or probability plot (Chambers et al., 1983). The basic idea behind this plot is the following. Suppose that y_1 to y_n are the observed data and that $y_{(1)}$ to $y_{(n)}$ are the values of the data sorted from smallest to largest, so that y_i is the p_i empirical quantile for $p_i = (i - 0.5)/n$. (The $y_{(i)}$ are commonly called the order statistics.) Also suppose $F(y)$ is the cumulative distribution function of the theoretical distribution in question. Now the p quantile of F, where $0 < p < 1$, is a number that we will call $Q_t(p)$, which satisfies $F(Q_t(p)) = p$. In the theoretical quantile—quantile plot, $Q_e(p_i)$ (the p_i empirical quantile, which is equivalent to $y_{(i)}$) is plotted against $Q_t(p_i)$. If the theoretical distribution is a close approximation to the empirical distribution, then the quantiles of the data will closely match the theoretical quantiles