LONG-TERM SURVIVORS IN AN EPIDEMIC WITHOUT RECOVERY

Giuseppe Schinaia

1. A COMPREHENSIVE SI/I EPIDEMIC MODEL

In recent years major advances in the clinical and pharmaceutical research have produced treatment schemes and drugs able to put a halt or, at least, to significantly slow down to the local progression of some infections. In other cases, infectious agents have developed resistance to traditional treatments and a selection of drug-resistant infectives can no longer be removed from those participating in the spreading of the infection. Epidemic modeling of infectious diseases like HIV/AIDS, some forms of hepatitis, herpes, tuberculosis and others have thus to be re-thought of so as to include, among their aspects, the existence of long-term survivors or long-survivors: a minority of infectives, whose disease progression evolves along qualitatively different lines from the others. A clear example of their detection and of the problems involved is in Buchbinder *et al.* (1994).

On the one side, their presence poses serious and unexplored questions at epidemiological level while, on the other, it makes the modeling process of the spreading of infections all the more complex, since it must also account for the contribution of a number of infectives with a peculiarly long period of infectiousness.

This paper introduces a few problems related to presence of long-term survivors in a typical susceptible-infective (SI) epidemic scheme, presenting possible solutions and implications on epidemiological indicators. The dynamics of susceptibles, not being the object of this study, will only be represented in its simplest, general form as $S(t+dt) = \beta_0 S(t) I(t) dt$, where the force-of-infection coefficient β_0 may include expressions of any complexity, without altering the validity of the long-term survivor modeling approach.

The usual way to include long-term survivors in a classical SI compartmental scheme of an epidemic is to insert a new compartment of infectives controlled by specific parameters and state variables (time of stay, contribution to the force of infection) and run it parallel to the main course of the epidemic, as in figure 1. Standard Markov modeling techniques assign numbers to the compartments involved and produce simulation results.

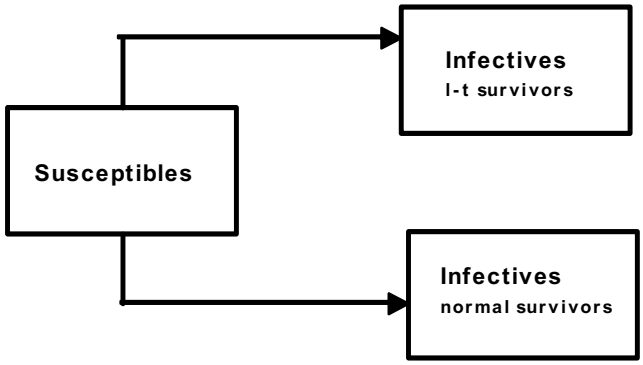

Figure 1 – Schematic representation of an elementary SI epidemic compartmental model, approximating a Markov model with separate compartments for long-term survivors and normal survivors.

In general, a Markov modeling approach (or its deterministic approximation) is acceptable in a short- or medium-term simulation with a homogeneous survival of infectives, where the time span prevents the approximations of all the dependencies to negatively affect the simulation results (Schinaia, 1997). The presence of long-term survivors, broadening the time interval involved, makes Markov model forecasts somehow unreliable, because of the lack of flexibility of the interactions among all the model entries. In particular, a simple modification of a Markov model, like the one proposed in the last section of this paper, by simply using a parametric representation of long-term survivors adds such features to the model structure that make the nature of the variables involved no longer compatible with the definition of Markov property.

Although the natural setting of theoretical epidemic models is of a stochastic nature and all the results in this paper refer to random variables, nevertheless the mathematical expressions will be explicitly given in terms of the more transparent dynamics of approximating deterministic systems. Moreover, the approach here presented to epidemic modeling with long-term survivors is in the form of mixtured models (Farewell, 1982 and Ghitany *et al.*, 1994), where different survival experiences are treated with different distributional assumptions. An extension of this approach to nonparametric models is mathematically more complex, since it may involve isotonic regression techniques and it is still the object of research-in-progress. A discussion of the comparison of the parametric vs. nonparametric approach is in Cantor and Schuster (1992).

2. LONG-TERM SURVIVOR MODELS

The presence of long-term survivors in a survival study can be naturally acknowledged by direct inspection of the infective survivor curve *H*: the right-hand tail of the curve in figure 2 clearly shows the existence of a fraction of right-censored failure times exceeding the observation interval.

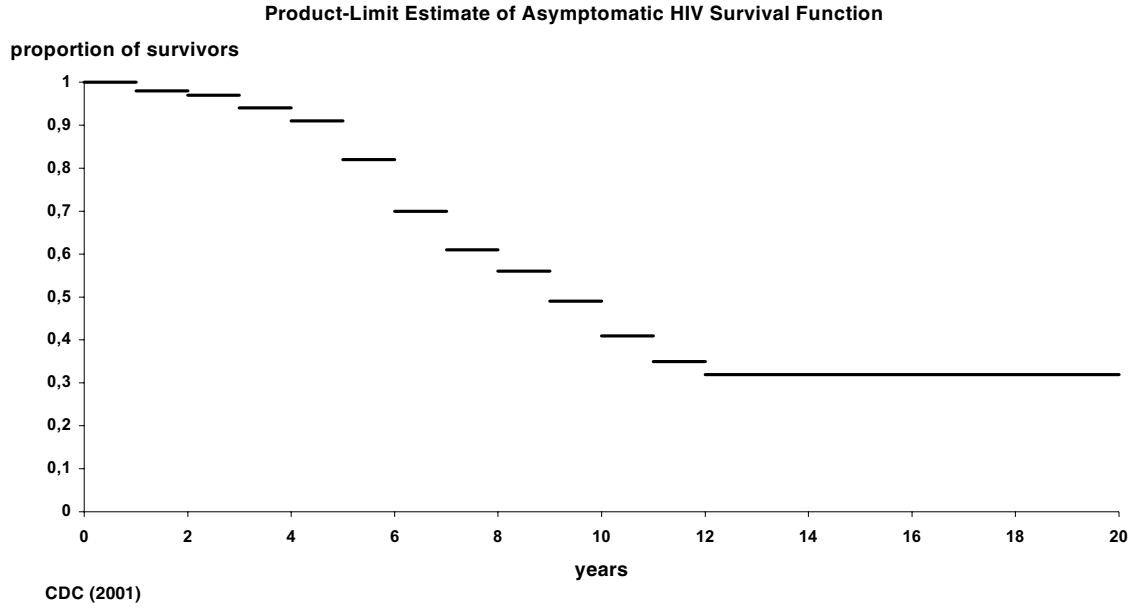

Figure 2 – An example of survivor curve with a positive right tail, showing the presence of long-term survivors.

Whether these 'tail' cases should be regarded as long-term survivors obviously depends on the duration of the study and on the proportion of cases detected; however, in principles, if $H(\tau) > 0$, where τ is the right-hand point of the observational interval, then not all the individuals at risk of failure actually failed. This implies that the cumulative distribution function $F(t)$ of random failure time T is improper and can be locally factored as

$$\begin{aligned} F(t_i) &= \Pr(T \le t_i) = \Pr((T \le t_i) \cap [(\varphi_i = 1) \cup (\varphi_i = 0)]) = \\ &= \Pr(T \le t_i \mid \varphi_i = 1) \Pr(\varphi_i = 1) = (1-\alpha) F^*(t_i) \end{aligned} \tag{1}$$

where

$$\varphi_i = \begin{cases} 1 & i\text{-}th \text{ individual subject to failure} \\ 0 & i\text{-}th \text{ individual long-progressor} \end{cases}$$

and $\alpha = \Pr(\varphi_i = 0)$. The cumulative distribution function $F^*(\cdot)$ is proper (total unit mass on real positive axis); in global terms, the survivor function is now given by

$$H(t) = \alpha + (1-\alpha) H^*(t) \tag{2}$$

where $H^*(t)$ is the survivor function of individuals subject to failure.

The scheme presented above is due to the work of several authors in the past years (Boag, 1949; Maller and Zhou, 1992, 1995, 1996; Rossi and Schinaia 1993,

1994; and Sposto *et al.*, 1992). Their researches mainly contributes to include long-term survivors in the analysis of survival data; however, little is left to the study of the contribution of long-term survivors to the diffusion of infectious diseases: in fact, the usage of this survival scheme in a simple model of disease progression, such as the SI scheme, cannot lead any further than stating that all susceptibles will eventually be infected. The prevalence of infectives at time t is given by

$$\begin{aligned} I(t) &= \beta_0 \int_0^t S(x)I(t)H(t-x)dx = \\ &= \alpha\beta_0 \int_0^t S(x)I(x)dx + (1-\alpha)\beta_0 \int_0^t S(x)I(x)H^*(t-x)dx \end{aligned} \tag{3}$$

which monotonically increases with time t and is unbounded $\forall \beta_0$, unless $S(t) = 0 \ \ \forall t \geq \tau$ for some τ. Therefore, unless the amount of susceptibles is limited with respect to the whole reference population, the number of infectives will eventually include all the individuals in the population.

However, actual epidemics appear to evolve somehow more flexibly than the situation described by (2) and (3); in fact, it is often the case, for instance, that long-term survivors tend to lose their infectivity, as time goes on and this event opens a wide range of possible evolutions of the spread of the disease in the susceptible population.

To be more general, let us consider the case where long-term survivors reduce their contribution to the spread of the epidemic as time-of-infection increases and the local contribution can be formalized $\forall i, t$ as

$$\varphi_i(t) = \begin{cases} 1 & \text{i - } th \text{ individual subject to failure} \\ 0 & \text{i - } th \text{ individual long-progressor} \end{cases} \tag{4}$$

where $\Pr(\varphi_i(t) = 1) = \alpha(t),\ t \in \mathbb{R}^+$. Similarly to (1), the global cumulative distribution function of T is then given by $F(t) = \Pr(T \leq t) = (1-\alpha(t))F^*(t)$ and its corresponding survivor function is

$$H(t) = \alpha(t) + (1-\alpha(t))H^*(t) \tag{5}$$

where both $F^*(t)$ and $H^*(t)$ refer to individuals subject to failure.

Using (5), (3) can be re-written as

$$I(t) = \beta_0 \int_0^t S(x)I(t)H(t-x)dx = \tag{6}$$

$$= \beta_0 \left[\int_0^t S(x)I(x)\alpha(t-x)dx + \int_0^t S(x)I(x)H^*(t-x)dx - \int_0^t S(x)I(x)\alpha(t-x)H^*(t-x)dx \right]$$

thus including long-term survivors into an expression for the prevalence of infectives, with (5) as a special case when $\alpha(t) = \text{constant}$. Remembering that, using (4), $\alpha(t)$ is defined as a probability, further hypotheses can be reasonably added on the form of $\alpha(t)$ that make $F(\cdot)$ a proper cumulative distribution function (*i.e.* $\lim_{t\to\infty} F(t) = 1$):

i) $\alpha(t)$ non increasing

ii) $\lim_{t\to+\infty} \alpha(t) = 0$; $\lim_{t\to 0} \alpha(t) = 1$

iii) $\alpha(t)$ right-continuous

The development of the epidemic depends now on the form of all the terms included and will not *a priori* evolve towards a definite direction (endemic steady state, extinction, explosion); a closer analysis of this question is in the next section, where the basic reproduction number is considered.

3. EFFECTS OF LONG-TERM SURVIVORS ON EPIDEMIC SIMULATIONS

The use of long-term survivor model (6) or of others with similar characteristics can effectively improve the quality of the epidemic simulation by adding important features to the mathematical structures. Effects of these additions can be directly detected on the output of the simulation runs and a comparison of different hypotheses in terms of prevalence curves can be used to decide what type of modeling approach best fits given epidemiological conditions.

In this section, some theoretical aspects related to the presence of long-term survivors in epidemic models will be discussed and examined and some numerical examples will be used to show the effects of different distributional hypotheses on the curves of prevalence of infections.

3.1. *Characterization of infectives*

With the progression of scientific knowledge of the infectious disease, both at microscopic and macroscopic level, the characterization of the epidemic becomes more stringent and generates closer mathematical models and more precise estimates of the statistical quantities involved. Both these aspects involve an increasingly refined classification of susceptibles and infectives to incorporate different

characteristics of such populations into comprehensive modeling structures. Splitting infectives into normal- and long-term survivors operates a first, although important, refinement of SI models; however, more information on the level of particular biological, social and demographic variables can be effectively incorporated to produce more accurate forecasts of the spread of the epidemic. This is, in general, not trivial in compartmental Markov models, since they need to increase the number of compartments (states), with the undesirable effect of a growing number of parameters controlling the corresponding transitions.

The model here proposed is general enough to allow for a natural and simple use of covariates (cofactors); like any survival model, $H(t)$ can be easily extended to $H(t,\mathbf{x})$, including any multivariate vector $\mathbf{x}$ of specific characteristics of groups of infectives, by means of standard multivariate survival techniques, without substantially altering the model structure and the numerical computations. Moreover, the actual form of $H(t)$ in (5), considered as a mixture of two separate survival models, $H^*(t)$ and $\alpha(t)$, can be used to incorporate variable infectivity, since it, in fact, introduces a classification of individuals, where $\alpha(t)$ can be regarded as a survivor function, with the end of infectivity period as the failure-defining event.

3.2. *Basic reproduction number*

One of the most effective indicators of the future development of an epidemic spread is the basic reproduction number R_0 (Heesterbeek and Dietz, 1996), representing the number of new infections generated by a single infected individual. It provides information on the severity of an epidemic, by establishing whether it will potentially invade the whole population of susceptibles and settle in the environment; in this paragraph, the general expression for R_0 is derived, when (6) is assumed and direct calculations in a simple case are explicitly given as an example.

The general form of R_0 is given by

R_0 = {# susceptibles at time 0}×{force of infection}×{estimated mean of infection time}

Let us now assume that hypotheses i)-iii) of section 2. hold for $\alpha(t)$, then R_0 can be expressed as

$$R_0 = S(0)\beta_0 \int_0^\infty H(t)dt$$

where, using the mixtured model (6), we have

$$R_0 = S(0)\beta_0 \left[\int_0^\infty \alpha(t)dt + \int_0^\infty H^*(t)dt - \int_0^\infty \alpha(t)H^*(t)dt \right] \tag{7}$$

If we refer to the interpretation of $\alpha(t)$ in terms of variable infectivity, the integrals in (7) can be viewed, respectively, as the estimated mean time of infectivity, the estimated mean time of survival and the estimated mean time of infectivity while surviving.

Direct computations of the additional severity to the epidemic spread due to long-term survivors obviously depend on the form of all the terms involved, but as an example, let us consider a simple case, where base survival $H^*(t) = e^{-\lambda t}$ for some λ and infectivity lasting $\varkappa$, i.e.

$$\alpha(t) = \begin{cases} 1 & t \le \varkappa \\ 0 & t > \varkappa \end{cases}$$

When no long-term survivors are present, it is known that $R_0^* = \lambda^{-1}\beta_0 S(0)$, while (7), with a first order approximation of $e^{-\lambda t}$, becomes

$$R_0 \cong \left[1 + \frac{\lambda^2 \varkappa^2}{2}\right] R_0^*$$

thus showing that the order of magnitude of the contribution of long-term survivors to the diffusion of the infection is no smaller than the square of the period during which long-term survivors are infective.

3.3. *Simulation examples*

A few simulation run are here presented to show the influence of long-term survivor model (6) on the form of the curve of prevalent infectives. Parameters to the survival curves are taken from Schinaia (2000) and are summarized in table 1; however, the simulation curves do not necessarily refer to any actual situation, since they only mean to offer a visual interpretation of the model presented in the previous section. As a matter of fact, parameters controlling the epidemic obviously need to be estimated from specific data that include observations on long-term survivors, when mixture model (6) is used to produce forecasts to epidemic spread.

TABLE 1

Summary of base survival models used in the simulation examples

Base Model	Survival Function	Parameter Values	Median Base Survival
Exponential	$e^{-\lambda t}$	$\lambda = \frac{1}{9}$	6.24 years
Weibull	$\exp\left\{-\left(\frac{t}{e^b}\right)^{\frac{1}{s}}\right\}$	$b = 2.299$ $s = 0.778$	6.22 years
Lognormal	$\frac{1}{2}\left[1 - F_{N(0;1)}\left(\frac{\ln t - \ln M}{\sigma\sqrt{2}}\right)\right]$	$M = e^{2.208}$ $\sigma = 0.683$	6.22 years

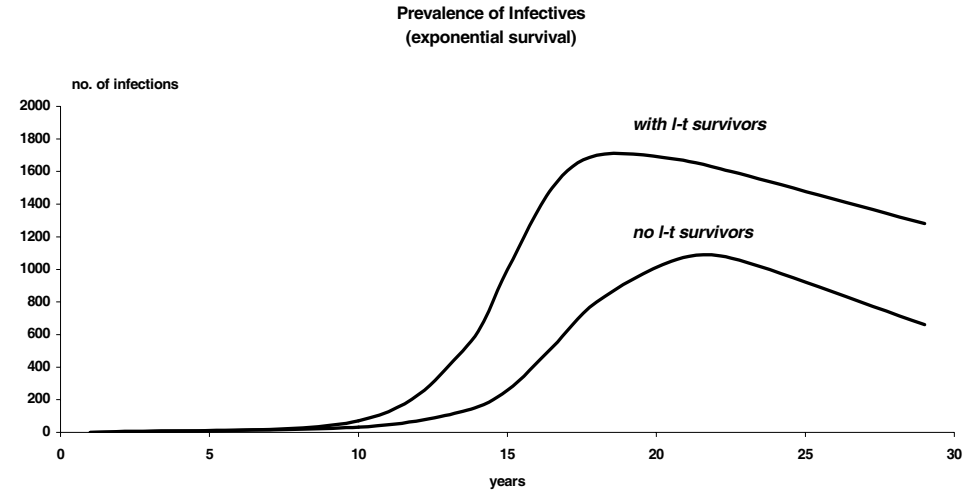

Figure 3a

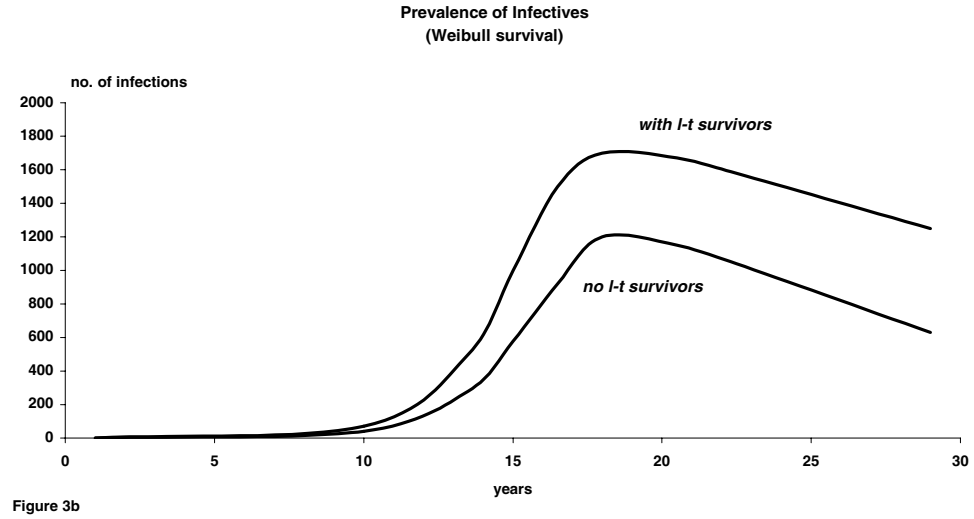

Figure 3b

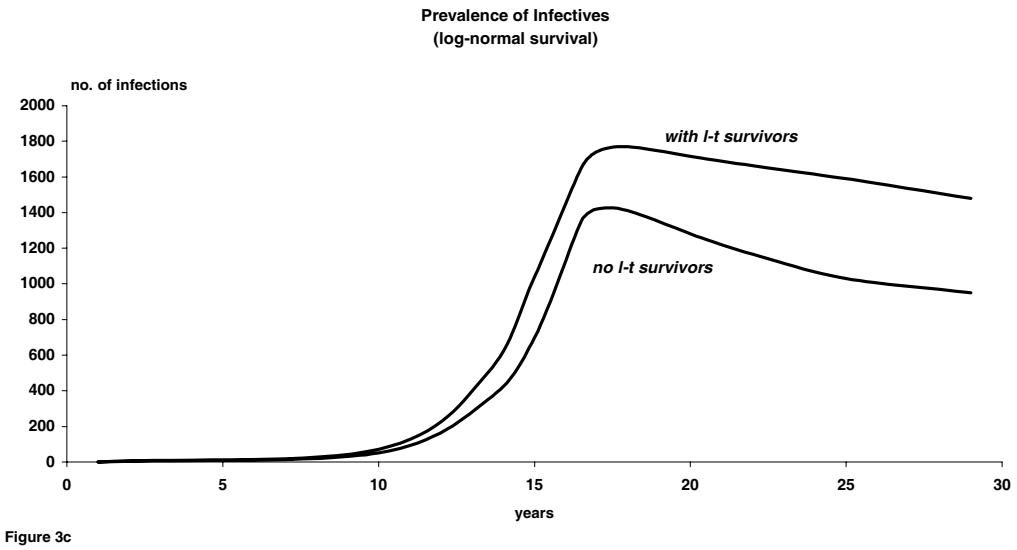

Figure 3c

Figures 3a, 3b, 3c – Prevalence curves of infectives in simple SI models, with and without long-term survivors. Exponential, Weibull and log-normal survivor functions are used for the base survival, while long-term survivor probability $\alpha(t)$ is exponential with mean 15 in all the three cases.

In figures 3a to 3c three different models are used for $H^*(t)$, exponential, Weibull and log-normal respectively, while $\alpha(t) = \exp\{-t/15\}$ is used in all the three long-term survivor curves. Note that, while $H^*(t) = e^{-\lambda t}$ in the exponential case in figure 3a produces the approximation of a Markov model, with the addition of long-term survivors as in (6), it becomes $H(t) = \exp\left\{-\frac{t}{15}\right\} + e^{-\lambda t} - \exp\left\{-t\left(\lambda + \frac{1}{15}\right)\right\}$, which is no longer approximating a Markov process.

Facoltà di Economia GIUSEPPE SCHINAIA
Università di Roma "La Sapienza"

BIBLIOGRAPHIC REFERENCES

J.W. BOAG, (1949), *Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy*, "Journal of the Royal Statistical Society", B11, pp. 15-44.

S.P. BUCHBINDER, M.H. KATZ, N.A. HESSOL, P.M. O'MALLEY, S.D. HOLMBERG, (1994), *Long-term HIV-1 Infection without Immunologic Progression*, "AIDS" 8, pp. 1123-1128.

A.B. CANTOR, J.J. SHUSTER, (1992), *Parametric versus Non-parametric Methods for Estimating Cure Rates Based on Censored Survival Data*, "Statistics in Medicine", 11, pp. 931-937.

V.T. FAREWELL, (1982), *The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors*, "Biometrics", 38, pp. 1041-1046.

M.E. GHITANY, M.A. MALLER, S. ZHOU, (1994), *Exponential Mixture Models with Long-Term Survivors and Covariates*, "Journal of Multivariate Analysis", 49, pp. 218-241.

J.A.P. HEESTERBEEK, K. DIETZ, (1996), *The Concept of R_0 in Epidemic Theory*, "Statistica Neerlandica", 50 1, pp. 89-110.

M.A. MALLER, S. ZHOU, (1992), *Estimating the Proportion of Immunes in a Censored Sample*, "Biometrika", 79, pp. 731-739.

M.A. MALLER, S. ZHOU, (1995), *Testing for the Presence for Immune or Cured Individuals in Censored Survival Data*, "Biometrics", 51, pp. 1197-1205.

M.A. MALLER, S. ZHOU, (1996), *Survival Analysis with Long Term Survivors*, Wiley, New York.

C. ROSSI, G. SCHINAIA, (1993), *Classification Methods and Expert Systems from a Bayesian Viewpoint*, in C.N.R.-S.I.S. "Proceedings of the International Workshop on Probabilistic Methods in Expert Systems", Roma, 1993, pp. 159-180.

C. ROSSI, G. SCHINAIA, (1994), *Dealing with Uncertainties in Automatic Procedures to Support Decision Making*, in "Proceedings of Workshop on Uncertainty in Probabilistic Expert Systems", Trest, pp. 211-226.

G. SCHINAIA, (1997), *Some Remarks on Time-Dependency in Epidemic Models*, in "Annali 1997 del Dipartimento di Studi Geoeconomici, Statistici, Storici per l'Analisi Regionale", Patron, Roma.

G. SCHINAIA, (2000), *Modelling the HIV/AIDS Epidemic via Survivor Functions*, "European Journal of Epidemiololy", 16, 6, pp. 573-579.

R. SPOSTO, H.N. SATHER, S.A. BAKER, (1992), *A Comparison of Tests of the Difference in the Proportion of Patients who are Cured*, "Biometrics", 48, pp. 87-99.

RIASSUNTO

Lungo-sopravviventi in un epidemia senza guarigioni

L'analisi dell'epidemia di una malattia senza possibilità di guarigione spesso implica la presenza di lungo-sopravviventi nel processo di modellizzazione. Tali individui sono caratterizzati da un tempo di permanenza particolarmente lungo nella condizione di infezione senza alcun segno di progressione verso stadi successivi della malattia. Un approccio parametrico alla modellizzazione della loro presenza in un'epidemia coinvolge una suddivisione degli stadi di progressione di tipo non-markoviano. Nel presente lavoro si propone un approccio che coinvolge la funzione di sopravvivenza e si presentano alcune osservazioni sugli indicatori epidemiologici, assieme ad esempi applicativi di tipo numerico e grafico.

SUMMARY

Long-term survivors in an epidemic without recovery

The analysis of epidemics without recovery often implies the inclusion of long-term survivors in the modeling process. Such individuals are characterized by a peculiarly long sojourn in a state of infection without any sign of progression towards the subsequent evolutionary state of the disease. A parametric modeling of their presence in an epidemic involves non-Markov staging of the infection process and of its development. In the present work, an approach involving survivor function is proposed, along with some remarks on epidemiological indicators and some numerical and graphical examples of its application.