Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 7017–7022, June 1999
Microbiology

A practical method for simultaneously determining the effective burst sizes and cycle times of viruses

(reproductive statistics/age-specific fertility/inverse problem/therapeutics/HIV)

JOHN L. SPOUGE*† AND SCOTT P. LAYNE‡

*National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894; and ‡Department of Epidemiology, School of Public Health, University of California, Los Angeles, CA 90095

Communicated by Stirling A. Colgate, Los Alamos National Laboratory, Los Alamos, NM, April 14, 1999 (received for review December 11, 1998)

ABSTRACT **We describe combined analytic and experimental methods for determining reproductive statistics from time-series data. Our computational methods derive four fundamental measures from laboratory experiments: (*i*) average number of viral daughters; (*ii*) mean viral cycle time; (*iii*) standard deviation of the viral cycling time; and (*iv*) viral doubling time. Taken together, these four reproductive statistics characterize "age-specific fertility," a quantity that provides complete information on the reproduction of the average viral particle. In this paper, we emphasize applications relating to HIV and experiments for assessing cellular tropism, viral phenotypes, antiviral drugs, humoral immunity, and cytotoxic cellular immunity. Nevertheless, our method is quite flexible and applicable to the evaluation of drugs against bacterial, fungal, and parasitic infections, antineoplastic agents against cancer cells, and perturbations involving pest and wildlife releases in ecosystems.**

Biology is the study of reproduction, mutation, and selection (1). Accordingly, many biological studies focus on quantifying one of these three processes. One problem, however, is that current methods for studying reproduction in a population often require an investigator to associate parents with their offspring. In a few populations (e.g., a human population), a reproductive census easily satisfies the requirement for parent–offspring associations. In many other populations (e.g., a viral population), however, a census is effectively impossible, so the requirement for parent–offspring associations cannot be met. Accordingly, we show in this paper how to bypass the requirement for parent–offspring associations and present new analytic and computational methods for calculating reproductive statistics.

The methods require adherence to three basic conditions: (*i*) starting with a population that is initially synchronized; (*ii*) monitoring the growth of the entire population over time; and (*iii*) distinguishing the initial generation over subsequent generations that follow. Such conditions are achievable in the laboratory when dealing with pathogens such as viruses, bacteria, fungi, parasites, or tumor cells. They are even achievable with more complicated ecological systems involving fish and wildlife. Because of its worldwide impact, we will discuss experiments pertaining to HIV, but minor variations on the same procedures will also work for many other problems in biology and medicine.

Many key questions in HIV research can be cast in terms of reproduction. How does viral reproduction influence pathogenicity, virulence, or person-to-person transmissibility? How does viral reproduction pertain to cell types, culture conditions, coreceptors, or definitions of tropism? How is viral reproduction altered by antiviral drugs, humoral immunity, and cellular immunity? With current methodologies, researchers are able to measure doubling times for viral assays (2), but not much else. Without making unwarranted or unproven assumptions about burst reproduction, researchers still cannot determine fundamental parameters such as the average number of daughters per mother virion (3). Out of necessity, scientific investigations have come to rely on assay methods that generate ambiguous and incomplete data (e.g., designating HIV strains as "fast-high" or "slow-low" phenotypes) and as a result, many important efforts cannot measure up to their full potential (4). Thus, in many situations, what virologists actually need are better methods for quantifying viral reproduction.

Each round of viral replication begins with "mother" virions giving rise to "daughters," which leads to a growing population of viral particles. For retroviruses such as HIV, the ratio of noninfectious to infectious physical particles ranges from 10^5:1 to 10^7:1 (2). Based on available technology, such findings illustrate that the overwhelming majority of HIV particles are somehow defective, and so reliable measures of HIV reproduction cannot be based solely on counting physical particles, with quantitative HIV-RNA PCR assays, for instance. Because we are concerned with reproduction and because only virions that successfully infect can actually reproduce, we can dependably overlook the presence of noninfectious particles and potentially infectious virions that fail to reproduce. We therefore define the terms mother and daughter to refer the small population of HIV particles that successfully infect their cellular targets. Note that similar concepts can be applied to the reproduction of any organism.

The organization of this paper is as follows. In *Methods*, we explain the experimental and analytical procedures for determining the four reproductive statistics described in the *Abstract*. In *Results*, we compare previous approaches to our methodology and examine how experimental noise gives rise to errors in the reproductive statistics. In the *Discussion*, we consider how reproductive statistics can be used for assessing cellular tropisms, viral phenotypes, antiviral drugs, humoral immunity, and cytotoxic cellular immunity. Readers focused on conducting laboratory experiments can gloss over the mathematical subsections without losing grasp of the essential points.

METHODS

Previous methods for determining viral reproductive statistics have made the assumption that each mother produces all of her daughters at once, in a burst distribution. This assumption often breaks down under experimental conditions (5–8). On the other hand, our methods for determining viral reproduc-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: Gen, viral generation; m, average number of daughters; τ, doubling time; μ, mean doubling time; σ, SD of cycling time.

†To whom reprint requests should be addressed. e-mail: spouge@nih.gov.

tive statistics make no *a priori* assumptions on the underlying distribution.

Experimental Procedures. Fig. 1 outlines steps for determining HIV reproductive statistics from lymphocytes, monocytes, macrophages, or dendritic cells. One-arm protocols are useful for assessing cellular tropisms and viral phenotypes, which reflects HIV's ability to infect and multiply in certain cell types. Two-arm protocols (i.e., control and experiment) are useful for assessing antiviral drugs, humoral immunity, and cytotoxic cellular immunity. At the outset, synchronized viral infections are begun in a relatively small number of labeled cells (designated as generation zero, Gen_0) which are then combined with a much larger number of unlabeled cells (collectively designated as generation plus, Gen_+). Subsequently, time-series samples are collected from the combined cultures for at least one complete generation time, until the initial generation (Gen_0) has produced almost all of its offspring. To assure reproducibility, culture conditions must be held stable over the period of entire experiment, which could span several days (Fig. 1).

Certain strains of HIV cause cell membranes to self-adhere and then fuse into multicellular syncytia (4). When performing experiments with wild-type isolates, caution is therefore advised because syncytium formation can potentially interfere with flowing and scoring of individual Gen_0 and Gen_+ cells. If the fraction of infected to noninfected cells remains small in the experiments, however, syncytial interference will likely be negligible. It is also important to keep in mind that *in vitro* determinations of age-specific fertility may not accurately reflect *in vivo* reproductive statistics. In order to increase the odds for making relevant comparisons, experimental conditions should be made as physiologic as possible, and experimental sensitivities to changes in the conditions should always be examined.

Basic Quantities. A reproductive census determines when mothers have daughters and how many daughters they produce (Table 1). The census results can be presented as a histogram that plots the number of births (y axis) against the mother's age at childbirth (x axis). This birth histogram is equivalent to the age-specific fertility curve that we now describe for viruses.

Let us consider an arbitrary viral population. Next, select any reference point in the viral life cycle that a virus must pass through before it replicates (e.g., release from a cell, attachment to another cell, the start of replication, etc.). To be specific, let us select viral attachment as the reference point. Now, choose any successfully infecting virus from the population and let t be its age, with $t = 0$ being its attachment to a cell. Define $i(t)dt$ as the average number of successfully infecting daughter virions that stem from this mother virion and go on to attach to cells in the short time interval from t to $t + dt$. We call $i(t)$ the age-specific fertility curve for the population.

Age-specific fertility describes the reproduction of an "average" virion. The average number of viral daughters is therefore

$$m = \int_0^\infty i(\chi)d\chi. \qquad [1]$$

Dividing $i(t)$ by m gives a probability density: $p(t) = m^{-1}i(t)$, where $\int_0^\infty p(x)dx = 1$. The probability distribution has mean

$$\mu = \int_0^\infty xp(x)dx. \qquad [2]$$

Eq. **2** gives the mean cycle time between the attachment of a mother virus and the attachments of its daughters. The probability distribution has a standard deviation σ:

Table 1. Reproductive statistics and the doubling time

	Reproductive statistics			
Family name	m	μ	σ	τ
Big	8	3	0	1
Small	4	2	0	1
Wide	4	3	1.9	1

Examples demonstrating how the mean number of daughter m, mean cycle time μ, and standard deviation of cycle time σ interact to influence the doubling time, τ. All four population-based measures (m, μ, σ, and τ) characterize the age-specific fertility curve.

Age-specific Fertility. In the Big family, each mother virion produces eight daughters at age 3 days, in a burst. In the Small family, each mother virion produces four daughters at age 2 days, also in a burst. By contrast, in the Wide family, each mother virion produces two early daughters at age 1.1 days and another 2 late daughters at age 4.9 days, in a doubled burst.

Distinctions. The Smalls and Wides produce half as many daughters as the Bigs ($m = 4$ vs. $m = 8$). The Smalls have a shorter mean cycle time than the Bigs and Wides ($\mu = 2$ vs. $\mu = 3$). The Bigs and Smalls produce their daughters all at once in a burst ($\sigma = 0$), whereas the Wides produce early and late daughters and so have a variation in their cycle time ($\sigma = 1.9$).

Outcomes. All three families have exactly the same fertility, as indicated by the same doubling time ($\tau = 1$). This occurs despite the fact that the Bigs produce twice as many daughters as the Smalls or the Wides.

Explanations. Table 1 clearly demonstrates that early daughters are a major force that drives viral fertility. In general, any spread in the cycle time that permits the production of more early daughters will lead to shorter doubling times. σ therefore has a profound impact on viral fertility. Also notice that the Bigs and Wides have identical values for the mean cycle time ($\mu = 3$) and the doubling time ($\tau = 1$), yet they show very different values for the average number of daughters ($m = 4$ vs. $m = 8$). It is obvious that by themselves, the mean cycle time μ and the doubling time τ do not determine the average number of daughters m.

$$\sigma^2 = \int_0^\infty x^2p(x)dx - \mu^2. \qquad [3]$$

If $\sigma = 0$, viral reproduction occurs in a burst distribution. A key improvement of our methods over previous ones is that we make no *a priori* assumptions that $\sigma = 0$ (3).

Viral infection results in the intracellular manufacture of proteins and nucleic acids, which can serve as markers of infection. As illustrated in Fig. 1, laboratory experiments can follow the growth of a viral population by measuring the manufactured markers. Accordingly, let the amount of marker at time t be $M(t)$.

On inoculation into cell cultures, most viruses soon settle into a pattern of exponential growth, $M(t) \approx Ce^{\rho t}$ for some parameters C and ρ. The parameter ρ quantifies the population fertility. The age-specific fertility $i(t)$ determines ρ through a characteristic fertility equation (9, p. 230, Problem 2),

$$\int_0^\infty e^{-\rho x}i(x)d\chi = 1. \qquad [4]$$

The population doubling time during exponentially growth is $\tau = \rho^{-1}\ln 2$, where ln denotes a natural logarithm, to the base $e \cong 2.718\ldots$. The age-specific fertility $i(t)$ therefore determines the four fundamental reproductive statistics (m, μ, σ, and τ) that summarize the reproduction of the average virus. (The biologically inclined reader may wish to skip to the *Results* at this point.)

Analysis Based on Burst Reproduction. Clearly, burst reproduction is a biologically unrealistic assumption, and the *Results* section shows that when analyzing real data, it leads to unrealistic conclusions. In this paragraph only, we assume that

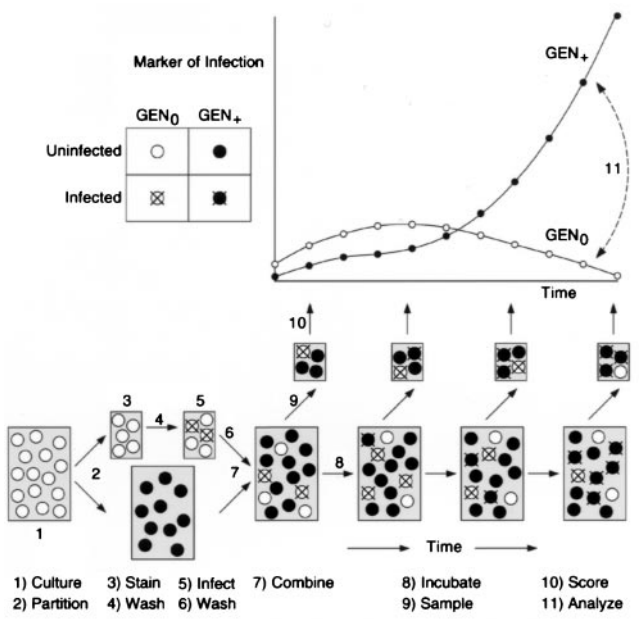

FIG. 1. Time-series experiment for determining HIV age-specific fertility. The figure shows two groups of cells, Gen_0 (fluorescent) and Gen_+ (nonfluorescent), along with the various steps for preparing and using these cells. The figure also shows stylized plots of the time-series data from Gen_0 (○) and Gen_+ cells (●), which are derived from hypothetical flow-cytometry measurements. In general, experiments are performed under conditions that limit the number of doubly infected cells, because such events fail to reflect *in vivo* conditions. Initially, such conditions require a relatively low multiplicity of infection for Gen_0 cells [multiplicity of infection (moi) < 0.1] and throughout the assay, a relatively low moi for the combined Gen_0 and Gen_+ cells (moi < 0.5). The step-by-step procedures are as follows. (*i*) Before assays, cell cultures are grown under conditions that ensure constant susceptibility to infection. (*ii*) Cell cultures are partitioned into two unequal portions, one for preparing cells for generation zero (Gen_0) and the other for preparing cells for generations one, two, three, etc. (designated collectively as Gen_+). (*iii*) Gen_0 cells are labeled with fluorescent dye and (*iv*) washed to remove excess dye from the surrounding medium. For such steps, several dyes are available that form stable associations with cytoskeletal proteins but, at low concentrations, do not appear to interfere with cell growth and metabolism (21). (*v*) Labeled Gen_0 cells are infected with cell-free HIV for a period of 1–2 hours and (*vi*) washed again to remove excess virions from the surrounding medium. After the second wash, cell-free virions must be absent, because their presence can corrupt numerical determinations of age-specific fertility. (*vii*) Labeled and infected Gen_0 cells are then combined with unlabeled and uninfected Gen_+ cells. In order to ensure that the growing infection propagates into Gen_+ cells and not Gen_0 cells, the ratio between Gen_0 and Gen_+ cells should be relatively large (e.g., >1:30). Such a large ratio also helps to keep the moi low throughout the assay. (*viii*) Cells are incubated under constant culture conditions (5% CO_2 and 37°C) with gentle mixing to prevent clumping. (*ix*) Over a period of several days, cells mixtures are sampled periodically (for example, every 3–6 hours), washed, and fixed with preservative to halt viral reproduction. To reduce perturbations from sampling procedures, the total volume of cells removed should be a fraction (<50%) of the culture's starting volume. (*x*) Multilaser flow cytometry is used to score for cell generation and markers of infection. Cells are stained with fluorescent mAbs against viral markers appearing very early within the cytoplasm (anti-p10, anti-p32, etc.) or somewhat later on the cell surface (anti-p24, anti-p41, etc.). Gen_0 and Gen_+ cells are distinguished by the presence or absence of fluorescent cytoskeletal labels, respectively. Cellular markers of infection can be scored by threshold and continuous counting procedures: as infected (+) and noninfected (−) according to cutoff immunofluoresence values; or as infected according to the actual amount of immunofluoresence detected. In principle, both counting procedures should yield comparable data and offer cross-checks of experimental consistency. At the start of the experiment, Gen_0 and Gen_+ cells will express relatively few markers of viral infection. Such sparse data will therefore demand the counting of more cells to collect reliable statistics and reduce experimental noise. (*xi*) Analysis of the data is carried out by mathematical procedures as described in *Methods*. The appearance of the marker in Gen_0 cells together with the age-specific fertility curve determines the appearance of the marker in Gen_+ cells. Thus, age-specific fertility can be viewed as a mapping between the marker on Gen_0 cells and Gen_+ cells (dashed line).

each mother virion produces an average of m_0 daughter virions in a single burst (3). We also assume that the daughters then go on to attach to their cellular targets at time μ later than the mother attached. These assumptions correspond to an age-specific fertility that is a burst distribution $i(t) = m_0\delta(t - \mu)$, where $\delta(t)$ is the Dirac delta function (10). The fertility equation then becomes $\int_0^\infty e^{-\rho x} i(x)dx = m_0 e^{-\rho\mu} = 1$ and yields a population fertility of $\rho = \mu^{-1} \ln m_0$. The population doubling time $\tau = \rho^{-1} \ln 2$ can be derived in accordance with previous results (3).

Analysis Based on Age-Specific Fertility. We now show that the age-specific fertility $i(t)$ can be extracted from time-series experiments measuring almost any type of manufactured viral marker. For viruses such as HIV, the marker could be reverse transcriptase, nucleic acids, or cell-associated proteins like p24, gp41, and gp120. Recall that $M(t)$ denotes the total amount of marker in a growing viral culture at time t, and define $M_0(t)$ to be the amount associated with the initial generation, Gen_0. The following renewal equation applies (9):

$$M(t) = M_0(t) + \int_0^t i(x)M(t - x)dx. \qquad [5]$$

We can now prove Eq. **5** as follows. Eq. **5** is linear in both $M(t)$ and $M_0(t)$, so both sides can be divided by the total number of viruses in Gen_0. The division in effect focuses the equation on a single progenitor virus in Gen_0, and after this reduction, Eq. **5** can be interpreted as follows. At time t, the marker $M(t)$ is associated either with the progenitor virus [corresponding to the term $M_0(t)$] or with a virus somewhere in Gen_+ (i.e., Gen_1, Gen_2, etc.), with Gen_+ being represented by the integral in Eq. **5**. Every virus in Gen_+, however, has a lineage that can be traced back to a specific viral daughter in Gen_1. In Gen_1, the average number of successfully infecting daughters attaching between x and $x + dx$ is given by $i(x)dx$. From time x to time $t - x$, each daughter and its progeny then give rise on average to the amount of marker $M(t - x)$. This concludes our proof.

Eq. **5** makes the physical assumption that every successfully infecting virion gives rise to the same amount of marker on average, regardless of its generation. This assumption is reasonable as long as experimental conditions remain constant over time and as long as factors such as mutation and selection do not appreciably perturb the age-specific fertility $i(t)$. The marker does not even have to be a conserved quantity. For example, it can serve as a carbon source in supporting cell metabolism or be incorporated into the cell genome over time. In fact, the marker can even rise or fall in any pattern. All such complexities do not matter as long as they remain constant across all generations, Gen_0, Gen_1, Gen_2, etc. The foregoing concludes our analysis of the forward problem.

Prepare for solving the inverse problem by defining

$$M_+(t) = M(t) - M_0(t) \qquad [6]$$

to be the amount of marker associated with Gen_+. Also, change the variable of integration in Eq. **5** from x to $t - x$, yielding the following equation

$$\int_0^t i(t - x)[M_0(x) + M_+(x)]d\chi = M_+(t). \qquad [7]$$

The Naive Method. Eq. **7** furnishes a mathematical formula for determining the age-specific fertility $i(t)$. Its continuous form, however, is poorly suited to experimental time-series data, which are inherently discrete. Let us now examine the effects of discretization and random experimental noise on the solution of Eq. **7**.

Typical laboratory experiments start at time $t_0 = 0$ and collect samples at a discrete number of time points t_0, t_1, t_2, . . . t_k (Fig. 1). These samples yield a series of marker values, designated as $M_0(t_0)$, $M_0(t_1)$, $M_0(t_2)$, . . . $M_0(t_k)$ for Gen_0 and $M_+(t_0)$, $M_+(t_1)$, $M_+(t_2)$, . . . $M_+(t_k)$ for Gen_+. These discrete marker values are associated with respective errors $\sigma_0(t_0)$, $\sigma_0(t_1)$, $\sigma_0(t_2)$, . . . $\sigma_0(t_k)$ and $\sigma_+(t_0)$, $\sigma_+(t_1)$, $\sigma_+(t_2)$, . . . $\sigma_+(t_k)$. The error estimates σ_0 and σ_+ will play a crucial role in our solutions.

In principle, there are many possible ways to estimate age-specific fertility from time-series data (e.g., parametric curve fitting). Yet after carefully considering the various possibilities, we elected to discretize the age-specific fertility $i(t)$ in accordance with the time series: $i(t_0), i(t_1), \ldots i(t_k)$. We chose this method of solution by discretization for its flexibility, primarily because it does not restrict the shape of the age-specific fertility curve.

The first step in our solution is to approximate the continuous integral in Eq. **7** with discrete values corresponding to the time points from the experimental data. Accordingly, let us define

$$y(t_j) = i(t_a - t_j)[M_0(t_j) + M_+(t_j)] = i(t_a - t_j)M(t_j). \quad \textbf{[8]}$$

If the Trapezoidal rule is applied to Eq. **7** at each of the experimental time points $t = t_a$ (where $a = 1, \ldots k$), we derive k equations:

$$\int_{t_0}^{t_a} y(x)dx \cong y(t_0)\,\tfrac{1}{2}(t_1 - t_0) + \sum_{j=1}^{a-1} y(t_j)\,\tfrac{1}{2}(t_{j+1} - t_{j-1}) + y(t_a)\,\tfrac{1}{2}(t_a - t_{a-1}). \quad \textbf{[9]}$$

Because mother virions do not produce daughters immediately, we have $i(t_0) = i(0) = 0$. For convenience, let us define

$$X_a = i(t_a)M(0)\,\tfrac{1}{2}(t_1 - t_0) + \sum_{j=1}^{a-1} i(t_a - t_j)M(t_j)\,\tfrac{1}{2}(t_{j+1} - t_{j-1}) - M_+(t_a). \quad \textbf{[10]}$$

In view of Eqs. **8** and **9**, we can approximate Eq. **7** by setting $X_a = 0$ for $a = 1, 2, \ldots k$. After the linear interpolation

$$i(t) \cong i(t_{i-1})\,\frac{t_i - t}{t_i - t_{i-1}} + i(t_i)\,\frac{t - t_{i-1}}{t_i - t_{i-1}} \quad \textbf{[11]}$$

for $t_{i-1} \le t \le t_i$ is applied to Eq. **10**, we can use the experimental time-series data to determine X_a ($a = 1, 2, \ldots k$) as a function of $i(t_a)$ (where $a = 1, 2, \ldots k$) in Eq. **10**.

The discretized set of simultaneous equations $X_a = 0$ for $a = 1, 2, \ldots k$ approximate the continuous convolution in Eq. **7**. Naively, one might expect that the solutions to these linear equations would approximate the unknowns $i(t_1)$, $i(t_2)$, . . . $i(t_k)$. Unfortunately, these naive solutions are unstable and oscillate wildly in response to any combination of discretization errors and random experimental noise (Fig. 2). Therefore, another method of solution must be sought.

Tikhonov Method. The quantity $X_a = 0$, if there is no random experimental noise and if there are no numerical errors due to the Trapezoidal rule and linear interpolation. Clearly, such conditions are highly improbable. Our immediate objective is therefore to calculate the random distribution of $\mathbf{X} = (X_1, X_2, X_3, \ldots X_k)$.

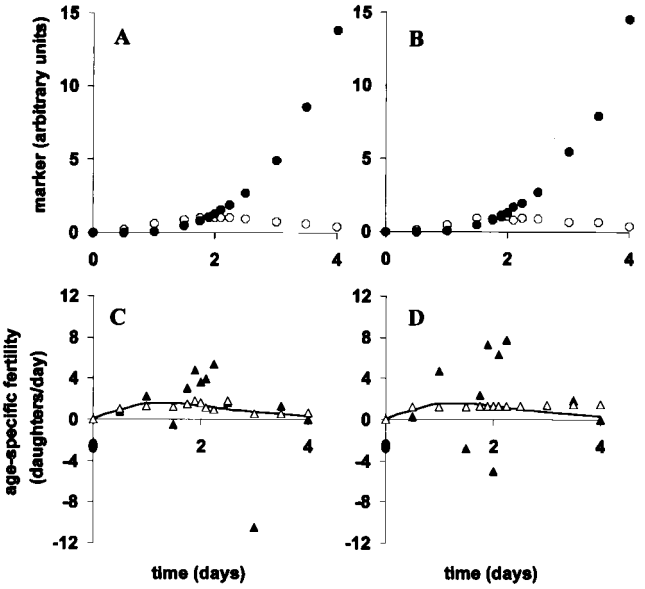

FIG. 2. A critical comparison of the naive and Tikhonov methods. The figure illustrates our numerical analysis of simulated experimental data that contain various levels of random noise. *A* and *C* are based on 1% random noise whereas *B* and *D* are based on 10% random noise. *A* and *B* plot cellular markers of infection that are scored from fluorescent Gen_0 (○) and nonfluorescent Gen_+ (●) cells. On the other hand, *C* and *D* plot our numerical analysis of the data from *A* and *B*, respectively, using the naive (▲) and Tikhonov methods (△). Note that the solid line in *C* and *D* represents the true age-specific fertility curve, which is the actual mapping between markers of infection in Gen_0 and Gen_+ cells (○ and ●). First, to generate noiseless data for Gen_+ cells, $M_+(t)$, Eq. **7** was used to combine the true age-specific fertility, $i(t)$, with the true time-dependent appearance of markers of infection in Gen_0 cells, $M_0(t)$. Next, 1% and 10% levels of random noise were added to generate the noise-containing data plotted in *A* and *B*, respectively, for Gen_0 and Gen_+ (○ and ●). Subsequently, we used the naive and Tikhonov methods to estimate age-specific fertility from these noise-containing data. In the presence of small levels of random noise (1%), the naive method, based on least-squares fitting is unexpectedly unstable and produces unusable solutions, whereas the Tikhonov method based on multidimensional minimization yields a good approximation to the true age-specific fertility curve. In the presence of moderate levels of random noise (10%), the naive method fails completely, whereas the Tikhonov method yields a somewhat looser approximation but still works (Table 2). In virological experiments utilizing flow cytometry, we expect that actual data will contain ~10% random noise. Hence, the Tikhonov method should perform well in laboratory applications.

Each coordinate of the vector $\mathbf{X}$ is a linear combination of $M_0(t_j)$ and $M_+(t_j)$ for $j = 0, 1, 2, \ldots k$. The quantities $M_0(t_j)$ and $M_+(t_j)$ contain random errors and can be viewed as random variables, with variances $\sigma_0^2(t_j) = \text{var}[M_0(t_j)]$ and $\sigma_+^2(t_j) = \text{var}[M_+(t_j)]$. Because experimental errors in $M_0(t_j)$ and $M_+(t_j)$ are independent, all of their covariances are zero. From Eq. **6**, it follows that the error in $M(t_j)$ is $\sigma^2(t_j) = \text{var}[M_0(t_j) + M_+(t_j)] = \sigma_0^2(t_j) + \sigma_+^2(t_j)$, and thus the covariance matrix $\mathbf{C}$ of the $1 \times k$ row vector $\mathbf{X}$ is a $k \times k$ matrix, with entries

$$C_{ab} = \text{cov}(X_a, X_b) = i(t_a)i(t_b)\sigma^2(0)\,\tfrac{1}{4}(t_1 - t_0)^2 + \sum_{j=1}^{c_*-1} i(t_a - t_j)i(t_b - t_j)\sigma^2(t_j)\,\tfrac{1}{4}(t_{j+1} - t_{j-1})^2$$

$$+ \delta_{ab}\sigma_+^2(t_a) - (1 - \delta_{ab})i(t_{c^*} - t_{c_*})\sigma_+^2(t_{c_*}) \times \tfrac{1}{2}(t_{c_*+1} - t_{c_*-1}), \quad [12]$$

where the Kronecker's delta $\delta_{ab} = 1$ if $a = b$, and $\delta_{ab} = 0$ otherwise. In Eq. **12**, the definitions $c_* = \min(a, b)$ and $c^* = \max(a, b)$ apply.

If the errors in $M_0(t_j)$ and $M_+(t_j)$ are normally distributed, the random variable $\chi^2(\mathbf{i}) = \mathbf{X}\mathbf{C}^{-1}\mathbf{X}^T$ is usually chi-square-distributed with k degrees of freedom (11), regardless of the values in the vector $\mathbf{i} = [i(t_1), i(t_2), \ldots i(t_k)]$ that it depends upon. Thus, the overall experimental error is about $\chi^2(\mathbf{i}) = k$, because k is the mean of the underlying chi-square distribution. Because $\mathbf{X} = 0$ implies $\chi^2(\mathbf{i}) = 0$, it should not be surprising that the naive solution failed.

In contrast to the naive method, which assumes $\chi^2(\mathbf{i}) = 0$, the Tikhonov method selects the smoothest curve $i(t_1), i(t_2), \ldots i(t_k)$ that is consistent with the overall experimental error $\chi^2(\mathbf{i}) = k$. The Tikhonov method is a good choice for solving our problem, because the age-specific fertility $i(t)$ is an average over the viral population and therefore must be a smooth function of the age t.

The Tikhonov method requires the use of a smoothing functional. The functional we have selected is

$$\Omega(\mathbf{i}) = 2\sum_{j=2}^{k-1}\left\{\frac{\ln[i(t_{j+1})/i(t_j)]}{t_{j+1} - t_j} - \frac{\ln[i(t_j)/i(t_{j-1})]}{t_j - t_{j-1}}\right\}^2 \times (t_{j+1} - t_{j-1})^2. \quad [13]$$

Eq. **13** approximates the integral $\int[(\ln i)'']^2 dt$ and therefore quantifies how much the age-specific fertility $i(t)$ oscillates. The particular form of Eq. **13** was chosen because at large times, we expect the age-specific fertility $i(t)$ to decline exponentially, giving $\ln i(t) = at + b$ and $(\ln i)'' = 0$. Other smoothing functionals $\Omega(\mathbf{i})$ were also examined, but Eq. **13** produced the most accurate estimates of the reproductive statistics m, μ, σ, and τ in repeated computer simulations.

The Tikhonov method is well recognized as an inversion method for many ill-posed problems like the one we address here (12). In our case, the Tikhonov method chooses values $i(t_1), i(t_2), \ldots i(t_k)$ that minimize $\Omega(\mathbf{i})$ subject to the constraint that the overall experimental error is $\chi^2(\mathbf{i}) = k$. The Tikhonov method can be implemented with a Lagrange multiplier [which is determined by the constraint $\chi^2(\mathbf{i}) = k$] and the Powell multivariable minimization method (12).

RESULTS

The main results of our paper are summarized in Fig. 2 as well as Table 1 and Table 2. The results show that our methodology is able to derive reproductive statistics for viruses such as HIV. In contrast to previous methodologies (3), our approach makes no *a priori* assumptions regarding burst reproduction, and its validity is reflected in the results that follow.

Table 1 demonstrates three key points regarding reproductive statistics. First, early daughters are a major driving force behind viral reproduction. The Small family produces only half as many daughters as the Big family but despite this difference, it has the same doubling time. Second, variations in the viral cycle time are equivalent to the production of early daughters. The Wide family produces only half as many daughters as the Big family but despite this difference, it also has the same doubling time. Third, the mean cycle time μ and the doubling time τ do not fix the average number of daughters m. The Wide family has the same mean cycle time and the same doubling time as the Big family but despite this similarity, it has only half as many daughters. In fact, with different numbers this disparity in the number of daughters could have been even more pronounced.

Table 2. Robustness of Tikhonov method to noise

Noise, %	Reproductive Statistics			
	m	μ	σ	τ
0	4.00	2.00	1.20	0.500
1	4.11	1.92	0.95	0.497
10	5.09	2.22	1.08	0.470

Reproductive statistics showing that the Tikhonov method is robust against noise in simulated experimental data. The table lists true and numerically derived values for m, μ, σ, and τ. The true values are derived from the solid lines in Fig. 2 *C* and *D*, which are precisely the same. The values for 1% and 10% noise are numerically derived from ▲ in Fig. 2 *C* and *D*, respectively. As expected, the errors in the numerically derived reproductive statistics increase as the experimental noise increases. A level of 10% random noise is a reasonable estimate for typical laboratory data based on flow cytometry. With such levels of noise, the table illustrates that our methods generate useful estimates of the true reproductive statistics. Based on many numerical simulations, we have found that τ is least sensitive to random noise, m and μ are immediate in sensitivity, and σ is most sensitive. These sensitivities are reflected in the results of Table 2.

Deriving Table Quantities. The solid line in Fig. 2 *C* and *D* represents the true age-specific fertility curve that was selected for this particular simulation. The four true reproductive statistics were derived by applying Eqs. **1–4** in the *Methods* to this solid line. The Tikhonov method generates discrete points for the age-specific fertility curve and, in order to derive the reproductive statistics from such points, Eqs. **1–4** were used in conjunction with the trapezoidal rule.

Table 2 shows that the errors in the estimated reproductive statistics generally become larger as the experimental noise increases. Even at 10% noise, which is a biologically realistic level of experimental error, the Tikhonov method still provides reasonable estimates of the reproductive statistics. The Tikhonov algorithm is therefore thoroughly practical for extracting reproductive statistics from time-series data.

DISCUSSION

Researchers have lacked effective tools for measuring reproduction. In this paper, we present practical methods for determining population-based reproductive statistics from time-series experimental data when it is not possible to make parent–offspring associations. The experiments require adherence to three basic conditions that are easily satisfied when dealing with viruses and unicellular and multicellular organisms (see Introduction). Our analysis sets forth a yardstick called age-specific fertility, which is especially powerful because we make no *a priori* assumptions on underlying reproductive statistics. In this particular discussion, we focus on laboratory experiments pertaining to HIV while recognizing that the spirit of the procedures in Fig. 1 will work for many other problems in medicine and biology.

Previous analytical methods for measuring reproduction have relied on time-series data exhibiting exponential growth or decay (2, 13). On the other hand, our new approach does not rely on exponential reproduction kinetics, enabling researchers to work with a broader spectrum of time-series data. Our methods in Fig. 1 will work with any data set containing enough scored samples from the initial generation (Gen_0) and subsequent ones (Gen_+).

Although the assumption of one synchronous crop of daughters ($\sigma = 0$) may help to simplify mathematical formulations (3), it can break down when determining reproductive statistics from simple data sets. As shown with the Big and Wide families (Table 1), the average cycle time μ and the population doubling time τ produce markedly different estimates for the average number of daughters m when $\sigma = 0$ and $\sigma > 0$. Such outcomes illustrate that a single burst distribution is far too constrained to work with real laboratory data, where various patterns of synchronous or asynchronous reproduction are going to be encountered. In contrast, the Tikhonov method

analyzes data reliably even in the presence of realistic levels of experimental noise (Fig. 2 and Table 2).

We now move on to practical matters. One perplexing problem that AIDS researchers face deals with assessing different viral isolates, antiviral drugs, and immune activities in a comparable manner. Our methods are an excellent way to make such experimental comparisons. For example, most antiviral drugs against HIV are small molecules that readily enter the intracellular compartment and interfere with early (reverse transcription) or late (protease cleavage) stages in viral manufacture. Monoclonal and polyclonal immunoglobulins and solubilized cellular receptors (such as soluble CD4 and chemokines) are large molecules that remain in the extracellular compartment and, among other actions, randomly block the attachment of virions as they diffuse from cell to cell. Cytotoxic T lymphocytes are enormous objects that migrate by a combination of cell adhesion and chemotaxis and suppress HIV reproduction by killing whole infected cells. It is conceivable that various antiviral drugs, humoral immune activities, and cellular immune activities could alter measures of age-specific fertility in patterns (involving m, μ, σ, and τ) that are somehow characteristic of underlying actions. Current assays, which are based on a variety of experimental approaches, are not directed at discerning such patterns. On the other hand, assays based on age-specific fertility generate a uniform set of reproductive statistics, which provide a basis for direct comparison.

Cellular Tropisms and Phenotypes. The value of m provides virologists and immunologists with a fundamental measure of tropism (14). Larger values of m reflect greater viral-cell tropisms (i.e., a greater number of successful daughter virions from a single infected cell) whereas smaller values reflect lesser tropisms (15). Another fundamental indicator of tropism is the population doubling time (τ), which has an inverse relationship to viral reproduction. Smaller values of τ reflect faster reproductive rates whereas larger values reflect slower rates. Of course, m and τ are merely two of the four measures that characterize the age-specific fertility curve. In many experimental situations, AIDS researchers will actually need to consider all four measures (i.e., m, μ, σ, and τ) for complete phenotypic characterizations.

Various strains of HIV appear to require the simultaneous expression of CD4 and chemokine receptors on the cell surface in order to initiate successful infection (16). AIDS researchers have thus undertaken extensive surveys of viral tropism according to cell-surface expression of CC- and CXC-chemokine receptors, with the expectation that tropism surveys would offer guidance to therapeutic and vaccine development efforts. Yet such surveys have achieved only partial success to date, in large part because researchers have lacked reliable methods for determining HIV reproductive statistics (17). We therefore suggest that reliable surveys of chemokine tropism could be achieved with experiments designed to determine age-specific fertility under certain prescribed conditions. Key variables in these quantitative viral reproduction assays would include (*i*) choice of HIV strains; (*ii*) use of lymphocytes, monocytes, macrophages, or dendritic cells; (*iii*) growth factors used in cell cultures; and (*iv*) number of CC and CXC receptors expressed on cell surfaces.

Antiviral Drugs and Immune Activity. Viral reproduction assays for assessing antiviral drugs, immunoglobulins, solubilized receptor proteins, and cytotoxic cellular killing would consist of two arms (i.e., control and experiment) but, in other respects, they would be similar to experiments just described. The control arm would measure reproductive statistics in the absence of antiviral agents. The corresponding experimental arm would measure reproductive statistics in the presence of antiviral agents, with all other conditions kept similar to controls. The output from the two parallel arms would therefore consist of determinations of age-specific fertility over a series of agent concentrations. If the agent has specific activity against HIV, it would perturb one or more reproductive statistics (m, μ, σ, and τ) in a concentration-dependent pattern that is characteristic of underlying action. If the agent has no specific activity, reproductive statistics would demonstrate only random fluctuations or possibly patterns attributable to nonspecific cellular toxicity.

Previous screening and evaluation methods for antiviral agents have often generated partial data that are incomparable to other data sets. In many instances, it has thus proven difficult to compare results from one infectivity assay to another. Such problems stem from uncontrolled variations in viral stocks (2) or arbitrary assignments regarding cutoff points, such as reporting ID_{50} vs. ID_{90}. Furthermore, most infectivity assays are scored at a single time point only, which provides isolated "snapshots" of how antiviral agents work over time. On the other hand, our methods could produce data that are comparable from one laboratory to another and could also help in addressing the questions posed in the Introduction.

Further Applications. Throughout this paper, we have emphasized methods relating to virology and reproduction of infectious agents. Our methods are equally applicable, however, to the screening and identification of new anticancer drugs. In this case, time-series data and markers are based on the cell itself and not on some cellular parasite.

Our methods require larger data sets than are customary at present, and so they can not be conveniently produced by humans. Fortunately, biological research is becoming automated and high-throughput research activities are becoming commonplace. Our methods are ready to take advantage of this trend (20).

We thank George Lewis and Michael Merges for useful discussions. This work was supported by the National Library of Medicine and the University of California.

1. Crick, F. (1981) *Life Itself: Its Origin and Nature* (Simon and Schuster, New York).
2. Layne, S. P., Merges, M. J., Dembo, M., Spouge, J. L., Conley, S. R., Moore, J. P., Raina, J. L., Renz, H., Gelderblom, H. R. & Nara, P. L. (1992) *Virology* **189,** 695–714.
3. Dimitrov, D. S., Willey, R. L., Sato, H., Chang, L. J., Blumenthal, R. & Martin, M. A. (1993) *J. Virol.* **67,** 2182–2190.
4. Fenyo, E. M., Albert, J. & Asjo, B. (1989) *AIDS* **3,** Suppl. 1, S5–12.
5. Allison, A. C. & Valentine, R. C. (1960) *Biochim. Biophys. Acta* **40,** 393–399.
6. Mandel, B. (1967) *Virology* **31,** 238–247.
7. Spouge, J. L. (1994) *J. Virol.* **68,** 1782–1789.
8. Wu, S. C., Spouge, J. L., Conley, S. R., Tsai, W. P., Merges, M. J. & Nara, P. L. (1995) *J. Virol.* **69,** 6054–6062.
9. Karlin, S. & Taylor, H. M. (1975) *A First Course in Stochastic Processes* (Academic, San Diego).
10. Lighthill, M. J. (1978) *Fourier Analysis and Generalized Functions* (Cambridge Univ. Press, Cambridge, U.K.).
11. Kendall, M. & Stuart, A. (1979) *The Advanced Theory of Statistics* (Griffin, London).
12. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1994) *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, U.K.).
13. Dulbecco, R., Vogt, M. & Strickland, A. G. R. (1956) *Virology* **2,** 162–205.
14. Layne, S. P., Spouge, J. L. & Dembo, M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4644–4648.
15. Layne, S. P. & Dembo, M. (1992) *Int. Rev. Immunol.* **8,** 1–32.
16. Baggiolini, M. (1998) *Nature (London)* **392,** 565–568.
17. D'Souza, M. P. & Harden, V. A. (1996) *Nat. Med.* **2,** 1293–1300.
18. Pantaleo, G., Soudeyns, H., Demarest, J. F., Vaccarezza, M., Graziosi, C., Paolucci, S., Daucher, M., Cohen, O. J., Denis, F., Biddison, W. E., *et al.* (1997) *Proc. Natl. Acad. Sci. USA* **94,** 9848–9853.
19. Rosenberg, E. S., Billingsley, J. M., Caliendo, A. M., Boswell, S. L., Sax, P. E., Kalams, S. A. & Walker, B. D. (1997) *Science* **278,** 1447–1450.
20. Layne, S. P. & Beugelsdijk, T. J. (1998) *Nat. Biotechnol.* **16,** 825–829.
21. Weston, S. A. & Parish, C. R. (1990) *J. Immunol. Methods* **133,** 87–97.