Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Virological Methods 135 (2006) 118–123
Journal of Virological Methods
www.elsevier.com/locate/jviromet

A statistical method for comparing viral growth curves

Gary P. Wang, Frederic D. Bushman *

University of Pennsylvania School of Medicine, Department of Microbiology, 3610 Hamilton Walk, Philadelphia, PA 19104-6076, United States

Received 16 December 2005; received in revised form 10 February 2006; accepted 20 February 2006
Available online 27 March 2006

Abstract

Viral replication is often analyzed by growth curves, in which viral multiplication in the presence of host cells is measured as a function of time. Comparing growth curves is one of the most sensitive ways of comparing viral growth under different conditions or for comparing replication of different viral mutants. However, such experiments are rarely analyzed in a statistically rigorous fashion. Here a statistical method is described for comparing curves, using replication of HIV in the presence of an integrase inhibitor as an example. A complication in the analysis arises due to the fact that sequential measurements of virus accumulation are not independent, which constrains the choice of statistical method. In the recommended approach, the values for virus accumulation over time are fitted to an exponential equation, then the means of the extracted growth rates compared using a nonparametric test, either the Mann–Whitney *U*-test for two samples or the Kruskal–Wallis test for multiple samples. A web-based tutorial for implementing this method is available at http://microb230.med.upenn.edu/tutorials/wangTutorial.

Keywords: HIV; Integrase; Inhibitor; Antiviral; Statistics; Growth curve

1. Introduction

Studies in virology are often conducted and reported in a fashion that lacks statistical rigor. As a result, conclusions reached are sometimes wrong. For most experiments in virology, however, there are statistical procedures that allow a rigorous assessment of the possible significance of trends in the data. An excellent introduction can be found in Richardson and Overbaugh (2005).

Growth curve studies are widely used in virology. Growth curve experiments involve infecting cells with a viral stock and monitoring viral replication over time (Delbruck, 1940). Often two conditions are compared, for example, a mutant and wild-type virus or replication in the presence and absence of an antiviral inhibitor. Some experiments monitor a single cycle of viral replication, which can vary from 20 min for bacteriophages up to several days for some animal viruses (Fields and Kinpe, 1996). Alternatively, growth can be monitored over many cycles of viral replication, sometimes for many weeks. New virus production can be measured by removing an aliquot of supernatant and analyzing the number of infectious units present (Delbruck, 1940), the amount of viral antigen (Michael and Kim, 1999) or the amount of viral nucleic acid (examples can be found in Butler et al., 2001; Michael and Kim, 1999). Alternatively, the fraction of infected cells can be measured, as with an immunofluorescence assay for cell-associated viral antigen (e.g. Michael and Kim, 1999). For a virus that integrates its DNA into the host genome, the number of prophages or proviruses per cell can be measured (Butler et al., 2001; O'Doherty et al., 2002).

Viral growth curves are rarely if ever compared in a statistically rigorous fashion (for previous work and references; see Richardson and Overbaugh, 2005; Spouge and Layne, 1999; Weinberg and Lagakos, 2001). To analyze trends in growth curve data, multiple independent cultures are analyzed over time and the data for each virus at each time point are typically plotted with error bars. Often, if the error bars do not overlap, the data is judged to be significantly different (in the following, "significant" is used to imply "statistically significant"). This is unsatisfactory. For one thing, there is wide variation in the formulas used to determine the values for the error bars. The popular statistical package GraphPad Prism provides no fewer than four alternatives and the option selected is rarely specified. Furthermore, there can be cases where error bars overlap but the difference is significant or where the error bars do not overlap but the data is not significant (Richardson and Overbaugh, 2005). An

Abbreviations: HIV, human immunodeficiency virus; IN, integrase

* Corresponding author. Tel.: +1 215 573 8732; fax: +1 215 573 4856.
E-mail address: bushman@mail.med.upenn.edu (F.D. Bushman).

0166-0934/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jviromet.2006.02.008

added difficulty is that experiments carried out on different days may have different extents of infection with all viruses tested (including the control) due to slight differences in the experimental manipulations. Averaging of experiments from different days can thus lead to an unreasonable increase in the apparent error. For these reasons, researchers often select a "representative" growth curve they believe represents the true response. Of course, such data selection is very risky and probably accounts for some cases of irreproducibility between laboratories.

A method for statistical analysis of viral growth curves is presented below. The mathematical machinery behind the methods is well worked out—readers seeking more background are referred to the following reference (Sokal and Rohlf, 1995, pp. 423–541). Curves are fit to the exponential phase of viral growth and the rate extracted. The collection of rates is then compared between treatments using a Mann–Whitney *U*-test or Kruskal–Wallis test, and significance reported as a *P*-value. A tutorial describing implementation of this method in detail can be found at http://microb230.med.upenn.edu/tutorials/wangTutorial.

2. Methods

2.1. Measurements of HIV replication

HIV NL4-3 viral stock was prepared by transient transfection in 293T cells using calcium phosphate. Viral supernatant was harvested at 48 h, then filtered through a 0.45-μm filter, concentrated and stored frozen at −80 °C. The amount of p24 viral capsid antigen in the viral stock was determined by enzyme-linked immunosorbent assay.

Viral titers were determined using the GHOST indicator cell assay as described (Morner et al., 1999). Under conditions of optimal infection (spinoculation and addition of 20 μg/ml DEAE dextran), infection of 40,000 GHOST cells with 100 ng p24 resulted in 11.9 ± 1.0% infection. Under the conditions used in the example (no spinoculation or addition of DEAE dextran), infection of 40,000 cells with 100 ng p24 yielded 0.34 ± 0.043% infection.

Jurkat cells at a density of 8×10^5 ml^{-1} were inoculated with 100 ng of HIV NL4-3 p24 capsid antigen, in the presence of 100 nM, 2 μM or 20 μM of integrase inhibitor L-731,988 or without the inhibitor. Each culture condition was performed in quadruplicate. Cells were cultured for 3 h, washed and transferred to 2.5 ml fresh RPMI 1640 medium containing 10% fetal calf serum, 50 μg/ml gentamicin, 2 mM L-glutamine and integrase inhibitor at the same concentration, and cultured at 37 °C with 5% CO_2. One hundred microliters of supernatant was removed from each culture daily, and fresh medium containing the same integrase inhibitor concentration was added back to the culture. The amount of p24 capsid antigen in the supernatant was measured by ELISA. To maintain similar cell density for the duration of the experiment, cells were split every 4 days by diluting one-fourth the culture with four-fold of fresh medium containing the same inhibitor concentration. The cell density prior to splitting reached about 2–2.5 × 10^6 ml^{-1}, and was consistent throughout the experiment. A culture that was not split over the course of the experiment was also performed in parallel. In this case, culture supernatant was obtained daily for p24 analysis as described above; however, three-quarter of the supernatant was exchanged for fresh medium every 4 days without removing cells in the culture.

2.2. Statistical analysis

The statistical package GraphPad Prism was used for fitting exponential growth data, the Mann–Whitney *U*-test and the Kruskal–Wallis test. The detection of correlated error structure in the growth curve data was carried out by Dr. Charles Berry as follows. The log-transformed data was fit to linear mixed-effects models using R, and an AR1 model was found to fit the data better than a repeated measures model.

3. Results

3.1. Sample HIV growth curves for statistical analyses

As a model system for analysis, we have analyzed replication of HIV in a T-cell line over multiple replication cycles (Fig. 1). HIV stocks (from strain NL4-3) were used to infect Jurkat T-cells. The cells were incubated with the stock for 3 h, then washed and transferred to fresh medium. Samples were withdrawn over 20 days and analyzed for viral output by quantifying the amount of p24 capsid antigen present in the culture medium. Three levels of an HIV-1 integrase inhibitor, L-731,988 (Hazuda et al., 2000), were added for comparison.

As can be seen in Fig. 1A (no inhibitor), the amount of p24 in the medium increased over the time of culture. A technical issue is introduced by the long duration of the experiment. A typical culture of uninfected Jurkat cells must be "split" periodically to keep the density of cells roughly constant (in this case, one-quarter the culture was diluted with three volumes of fresh medium). However, splitting dilutes the amount of virus in the culture medium and reduces the number of virus-producing cells. As a result, the curves in Fig. 1 show a periodic drop in p24 antigen levels.

A culture that was not split during the course of the experiment was also tested (data not shown). In this case, the output of p24 antigen was reduced, consistent with reduced health of the cells under these conditions. Tests by trypan blue exclusion showed that the number of dead cells was greater in the culture that was not split (e.g. about 25% of cells from the split cultures did not exclude trypan blue when tested after day 4, whereas about 70% of cells failed to exclude trypan blue in the unsplit culture). Thus in this system it is possible to dispense with regular splitting, but only at the expense of accepting suboptimal viral and cellular growth. For this reason statistical methods that can be used with the irregular curves in Fig. 1 are desirable.

As can be seen in Fig. 1C, the inhibitor was highly effective at 20 μM. In fact, the effectiveness of inhibition introduces a challenge in statistical analysis, because the p24 values were below the level of detection (for this reason, fewer time points were measured than for the control due to the expense of the assay). At an intermediate concentration of inhibitor (2 μM; Fig. 1B),

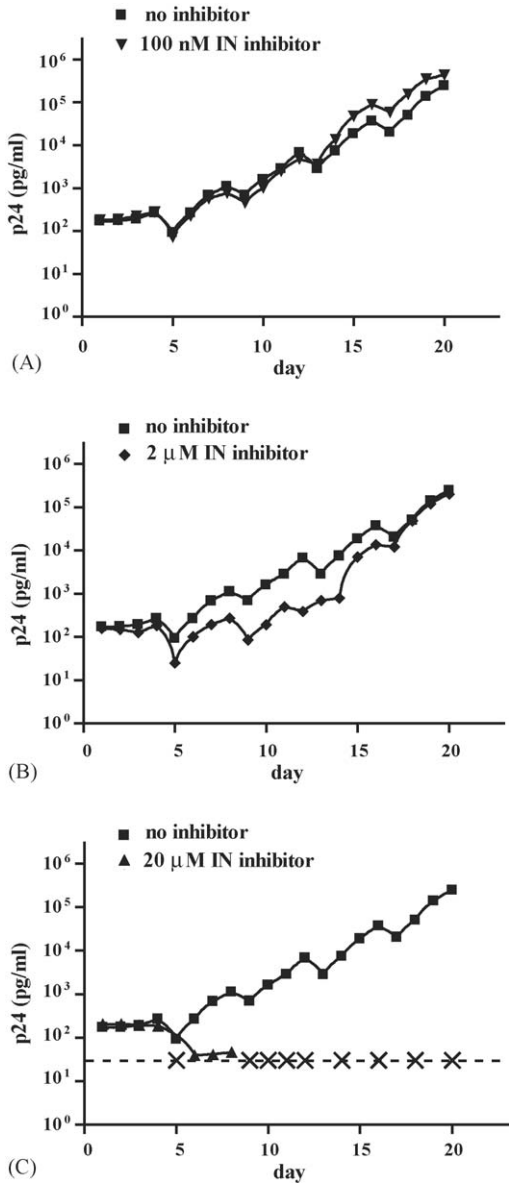

Fig. 1. Concentration of HIV p24 capsid antigen production in infected cultures of Jurkat T-cells. (A) Output of p24 in infected Jurkat cultures split every 4 days. Cells were cultured in the absence of inhibitor (filled squares) or in the presence of 100 nM concentration of the integrase inhibitor L-731,988 (inverted triangles). (B) As in (A) but 2 μM of inhibitor was used (filled diamond). (C) As in (A) but 20 μM of inhibitor was used (filled triangles). The '×' marks indicate cases where at least one of the four measurements was below the level of detection of the p24 assay. Each curve summarizes the values of four replicates.

a modest reduction in p24 production was seen. Whether this difference is statistically significant is unclear from the raw data. At 100 nM inhibitor, the curve describing p24 production appears to be indistinguishable from the wild-type. It is desirable to apply statistical analysis to determine whether there is any basis for claiming differences among these curves.

3.2. Curve-fitting by linear regression

A logical method for analyzing growth of viruses involves modeling the data as an exponential process. This approach has several advantages and disadvantages. For viruses that grow rapidly, and for which an exponential growth phase can be readily measured, this method is optimal. A considerable advantage is that the numerical value extracted – the growth rate – is a

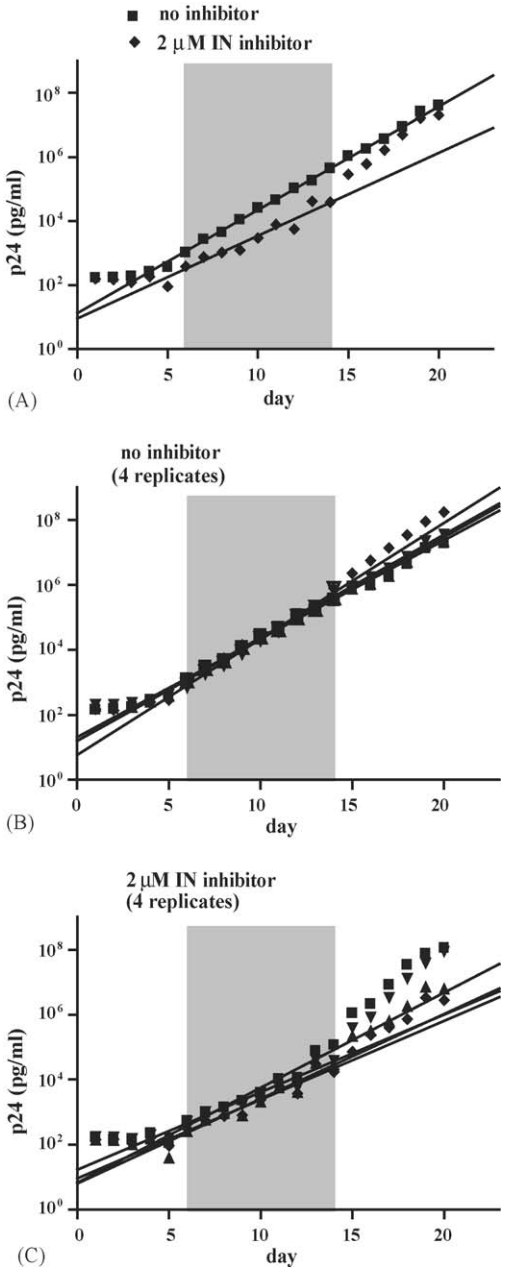

Fig. 2. Analysis of the rates of p24 production in the absence of inhibitor (filled squares) or in the presence of 2 μM inhibitor (filled diamond). Curves were fit by linear regression against the log of the p24 concentration in the culture supernatant. The effects of splitting the curves were smoothed by the following calculation: days 1–4: no change; days 5–8: each p24 value was multiplied by 4; days 9–12: p24 values were multiplied by 16; days 13–16: p24 values were multiplied by 64; days 17–20: p24 values were multiplied by 256. These correction factors were based on the expected p24 and virus-producing cells had the cultures not been split. Data from days 6 to 14, an estimation of the exponential growth phase, were fit by linear regression against the log of the p24 concentration in the culture supernatant. (A) Pooled replicates (average of four per time point), (B) no inhibitor, four replicates shown separately and (C) 2 μM inhibitor, four replicates shown separately. The gray shading indicates the time points (days 6–14) used in the linear regression.

biologically meaningful quantity and not a purely mathematical concept. However, cultures of many viruses, such as the HIV cultures in Fig. 1, require experimental manipulations (splitting the cell cultures) that cause the measurements to depart from the expectations of simple exponential growth. This requires correction of the data. Another complication is that the boundaries of the exponential growth phase must be determined by a preliminary inspection of the data, which has the potential to introduce subjective factors. This is in any case an approximation, since the fraction of uninfected cells will be changing during the culture period. Thus, culture conditions must be arranged to allow detection of the early exponential growth phase.

As an example, consider the data in Fig. 1B, which compare the culture with no inhibitor to one containing 2 μM inhibitor. The data in Fig. 1B showed periodic decreases in p24 antigen due to splitting of the cell culture. Fig. 2A shows artificially smoothed data, in which the amounts of virus and virus-producing cells expected to have been removed by splitting are added back computationally (see figure legend for details). Fig. 2B and C show the four individual replicates under each condition.

An exponential phase could be identified in the data in Fig. 2A. In the first 5 days, there was little growth above background, so this phase has been removed. After day 14, the growth in the culture with 2 μM inhibitor was not much different from that of the culture with no inhibitor. The basis for this was not studied, but it is possible that inhibitor resistance mutants accumulated in the viral population, as has been seen in previous studies of this inhibitor (Hazuda et al., 2000). Thus, data from days 6 to 14 was selected for the analysis.

As can be seen in Fig. 2B and C, the artificially smoothed data for each replicate could be fit to a linear exponential model. The natural log of the p24 value was used instead of the initial value, so that the slopes are proportional to the growth rate.

$$Y = e^{Kt}, \text{ so } \ln(Y) = Kt$$

The extracted growth rate K (in units of (ln(p24 pg/ml))/day) for all the data in Fig. 1 are summarized in Table 1.

For the 20 μM inhibitor data, the measurements of p24 were low at the start and became lower than the detection limit of the assay with time. This creates a problem for the analysis. Excluding these values would obviously bias the 20 μM measurements in favor of higher values. How to treat such data has been the topic of extensive discussion (e.g. Beal, 2001). In order to allow fitting of these data, the undetectable values were set to 1. This resulted in negative values for the slopes in the 20 μM data, because the values went down over time. A consequence of the treatment of undetectable values is that the rate of decline of p24 in the cultures is not estimated very accurately, but the rate of loss of virus under conditions of efficient inhibition is not important in evaluating the experimental outcome in this case.

3.3. Mann–Whitney U-test

The growth rates extracted from this analysis could then be compared using the Mann–Whitney *U*-test. This is a nonparametric test, which means that there is no assumption of a normal or Gaussian distribution of the error term, as is the case in a *t*-test or *F*-test. The measurements of growth rate are independent, since each replicate represents a different culture dish, as is required for this and subsequent tests. As can be seen from Table 1, the collection of growth rates in the presence of 100 nM inhibitor was indistinguishable from the collection of rates in the absence of inhibitor. The 2 and 20 μM values, however, were significantly different from the control. That is, the *P*-values were below 0.05. Note that the *P*-value for 2 and 20 μM values were the same ($P = 0.0286$). This is because the Mann–Whitney *U*-test works by ranking the collection of values, and in both cases the four measured rates in the presence of inhibitor were below the four rates in the absence of inhibitor.

However, as is discussed below, the Mann–Whitney test is only appropriate when two conditions are compared, for example, no inhibitor and one concentration of inhibitor.

3.4. Using the Kruskal–Wallis test and Dunn's post-test to control for multiple comparisons

An issue arises when more than two comparisons are made, because the chances of obtaining a significant difference by chance are increased as the number of comparisons increase. For this reason, it is appropriate to apply the Kruskal–Wallis test, which compares the mean values among the four conditions tested and allows control of the inflation of false-positive error rate due to multiple comparisons with Dunn's post-test.

Table 1
Statistical analysis of HIV growth rates in the presence and absence of the integrase inhibitor L-731,988

Replicate	No inhibitor	Inhibitor concentration		
		100 nM	2 μM	20 μM
A	0.7295 ± 0.0520	0.8151 ± 0.0563	0.6732 ± 0.1309	−0.1863 ± 0.3136
B	0.6968 ± 0.0420	0.7382 ± 0.0724	0.6007 ± 0.1483	−0.1577 ± 0.2967
C	0.7220 ± 0.0326	0.7296 ± 0.0708	0.5495 ± 0.1298	−0.2331 ± 0.3502
D	0.8223 ± 0.0510	0.9162 ± 0.1271	0.5580 ± 0.1532	−0.1644 ± 0.3028
P-value vs. no inhibitor				
Mann–Whitney *U*-test		0.2	0.0286	0.0286
Kruskal–Wallis test		>0.05	>0.05	<0.05

Growth rates (ln(pg/ml)/day) were determined from the exponential phase of the data (days 6–14) by linear regression, and are shown as the best fit values with 95% confidence intervals. The replicates for no inhibitor and each inhibitor concentration are unpaired, and represent independent measures of growth rates.

As with the Mann–Whitney test, the Kruskal–Wallis test is non-parametric and so does not require that the error terms of the measurements be normally distributed. Three comparisons were chosen for analysis by this method, in which each of the drug treatments were compared to the untreated control.

The overall Kruskal–Wallis test returned $P = 0.0041$, indicating that there were significant differences in the mean growth rates. Pairwise comparisons, taking advantage of Dunn's correction, showed that the control curve differed significantly from the 20 μM sample, but that the 100 nM and 2 μM curves did not differ. Thus, when multiple comparisons are taken into account, the relatively modest difference between the control and 2 μM-treated sample becomes insignificant.

To investigate possible significance of inhibition in the presence of 2 μM inhibitor further, it would be necessary to carry out another experiment. To maximize the statistical power of the experiment, it would be best to compare only the control and 2 μM inhibitor conditions, and to test more replicates of each condition.

3.5. Consequences of different choices for the exponential phase

The exponential phase in the log-transformed data was identified by eye as the linear range (days 6–14) in a plot of ln(p24) versus time, which raises potential concerns. How much difference does it make if different choices are made? To investigate this question, the analysis was repeated but selecting different time intervals for analysis.

If all the data is used (days 1–20), the uncertainty in the fit of the regression line to the data is increased, and there is no significant difference detected ($P > 0.05$) among any of the pairs of samples after Dunn's correction for multiple comparisons (though the overall Kruskal–Wallis test does achieve significance of $P = 0.014$). Using days 5–15, which include points that diverge from the linear fit, resulted in a slight increase in significance ($P = 0.011$), but again none of the comparisons to the no-inhibitor control achieved signficance after Dunn's test.

Narrowing the time interval from 6–14 to 7–13 days, yielded a more significant value for the Kruskal–Wallis test ($P = 0.0045$), and the no inhibitor rates were significantly different from the 20 μM rates. Thus, comparisons indicate that it is important to restrict the analysis to the exponential phase of the growth curve, and it is best to select relatively narrower time intervals.

4. Discussion

Growth curve analysis is widely used to compare viral mutants, antiviral agents, cellular mutants and other determinants of viral growth rates. These curves are rarely analyzed in a rigorous way for significant differences. This paper presents one approach to statistical analysis of such data.

The analysis is considerably complicated by the correlated error structure in the measurements (Charles Berry, G.P.W. and F.D.B., data not shown). That is, if one measured value is higher than the unknown "true" value for that time point, then there is an increased chance that the next measured time point will be erroneously high as well. This was found to be the case by empiral analysis of the data in Fig. 1, and is expected to be the case generally (Richardson and Overbaugh, 2005). This means that tests such as the Sign test on paired values or ANOVA-based methods will potentially yield incorrect *P*-values due to the correlated error. One way around this complication is to generate a summary measure for each replicate – the growth rate – and carry out the statistical analysis on the collection of summary measures.

To carry out this test, it is necessary to identify the exponential phase of viral growth. In the example (Figs. 1 and 2), the cultures showed little HIV growth in first 5 days, and later growth in the 2 μM inhibitor-treated culture (days 15–20) was similar to the no-inhibitor control, possibly due to viral mutation to drug resistance. Modeling the exponential phase required removing points outside the exponential range to achieve significance. Comparison of analysis using different time intervals indicated that it is best to take relatively narrow intervals.

Another potential complication arises in cases where a virus overgrows the cell culture. In that case, the measure of viral growth approaches a plateau value as the uninfected cells become depleted. In this case, it is necessary to analyze the early (more nearly exponential) phase of the curve or use a mathematical approach to determining the growth rate that takes saturation into account. Because the reasons for saturation can be diverse, and the mathematical model can be influenced by these assumptions, it is more straightforward to estimate growth rate from the exponential phase only.

A few practical recommendations for implementing this method.

The statistical power is determined by the number of replicates. Four replicates is the minimum that can achieve $P < 0.05$ in the Mann–Whitney *U*-test, and this is only when all of the rates from one condition are greater than all of the rates in the other condition. If one value overlaps between two replicates, then five tests are needed to obtain significance. It is helpful to carry out as many independent determinations as is reasonably possible to maximize the opportunity for detecting significant differences.

For modeling as an exponential process, the number of points in the exponential phase should be maximized. A pilot experiment may be helpful to identify the optimal times and concentrations of virus and cells.

Often experiments are carried out on different days. It is essential that significance be determined over an entire body of experiments, but modest differences in the viral inoculum, the passage number of cells and other parameters can result in differences in virus output. However, the growth rates will usually be more similar between days, so the rates may be pooled to allow an overall analysis of significance across all trials.

Careful thought should be given to the question of single cycle growth curves versus multicycle growth curves. Allowing the virus to grow for multiple cycles has the potential advantage of magnifying small differences, and so may be more sensitive. However, at least for animal cell culture, this will more often require splitting of cells and consequently corrections in the data. Single-cycle growth curves have the potential advantage

of allowing assignment of rates to production of stage-specific replication intermediates. However, care must be taken to ensure that the underlying process really is exponential. An alternative method for analyzing single-cycle data would be to quantify the maximum amount or rate of production of a replication intermediate. Mean values for the different conditions or mutants could then be compared directly using the Mann–Whitney test or Kruskal–Wallis tests with Dunn's correction. Alternatively, if the data is normally distributed, a *t*-test or ANOVA with Tukey's correction for multiple comparisons could be used (Richardson and Overbaugh, 2005).

In closing, this study is intended as a point of departure for developing useful approaches for statistical analysis of viral growth curves. Further suggestions are welcome. For example, for many data sets it might be possible to design linear mixed-effects models that would allow more sensitive detection of statistical significance. However, the models needed will likely differ for different experiments and implementing such methods requires advanced mathematical training. Similarly, there are more sophisticated ways of handling data points below the limit of detection than that employed here (Beal, 2001), but many of these too require advanced mathematical methods. The goal of this report is to introduce relatively simple methods that can be carried out in typical virological laboratories. More detailed instructions for implementing this method can be found in the tutorial at http://microb230.med.upenn.edu/tutorials/wangTutorial.

Acknowledgements

We are extremely grateful for crucial input from Charles Berry and Peter Petraitis, without whom this work would not have been possible. This work was supported by NIH grants AI52845, AI34786, AI057168, the James B. Pendleton Charitable Trust, and Robin and Frederic Withington (to F.D.B.). Gary P. Wang was supported by NIH NIAID T32 AI07634 (Training Grant in Infectious Diseases).

References

Beal, S.L., 2001. Ways to fit a PK model with some data below the quantification limit. J. Pharmacokinet. Pharmacodyn. 28, 481–504.

Butler, S., Hansen, M., Bushman, F.D., 2001. A quantitative assay for HIV cDNA integration in vivo. Nat. Med. 7, 631–634.

Delbruck, M., 1940. The growth of bacteriophage and lysis of the host. J. Gen. Phys. 23, 643–660.

Fields, B.N., Kinpe, D.M., 1996. Virology. Raven Press, New York.

Hazuda, D.J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J.A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., Miller, M.D., 2000. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. Science 287, 646–650.

Michael, N., Kim, J.H., 1999. HIV Protocols. Humana Press, Totowa, NJ.

Morner, A., Bjorndal, A., Albert, J., Kewalramani, V.N., Littman, D.R., Inoue, R., Thorstensson, R., Fenyo, E.M., Bjorling, E., 1999. Primary human immunodeviciency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage. J. Virol. 73, 2343–2349.

O'Doherty, U., Swiggard, W.J., Jeyakumar, D., McGain, D., Malim, M.H., 2002. A sensitive, quantitative assay for human immunodeficiency virus type 1 integration. J. Virol. 76, 10942–10950.

Richardson, B.A., Overbaugh, J., 2005. Basic statistical considerations in virological experiments. J. Virol. 79, 669–676.

Sokal, R.R., Rohlf, F.J., 1995. Biometry. W.H. Freeman and Company, New York.

Spouge, J.L., Layne, S.P., 1999. A practical method for simultaneously determining the effective burst size and cycle times of viruses. Proc. Natl. Acad. Sci. U.S.A. 96, 7017–7022.

Weinberg, J.M., Lagakos, S.W., 2001. Efficiency comparisons of rank and permutation tests based on summary statistics computed from repeated measures data. Stat. Med. 20, 705–731.