

Interpretation of Cellular Proliferation Data: Avoid the Panglossian

Mario Roederer*

• Abstract

There are several statistics that may be calculated to characterize a cellular proliferation experiment. By far, the most commonly-reported statistic is the percent of cells in the final culture that have divided; however, this statistic has significant limitations. Other statistics provided by software modeling provide a much richer characterization of the biological response; however, their use also comes with caveats. Here, I discuss the practical application of these statistics, including their limitations and interdependencies, using hypothetical data. The goal of this perspective is to prevent the blind reliance or overly optimistic (“panglossian”) interpretation of the statistics generated by software, so that researchers and reviewers have a more-informed basis for drawing conclusions from the data. Published 2011 Wiley-Liss, Inc.†

• Key terms

CFSE; mathematical modeling; precursor frequency; cell division

AN important assessment of cellular function, particularly in immunology, is the proliferative capacity of cells under different conditions. A flow cytometric assay based on the dilution of carboxy fluorescein diacetate succinimidyl ester (CFSE, also known as CFDA-SE) fluorescence during culture has become the standard for this purpose (1); when done well, analysts can enumerate cells that have divided as well as quantifying how many divisions each cell has undergone (2–4). The basic principle is well-described, and relies on the two-fold dilution of fluorescence accompanying each cellular division as the fluorescently-tagged cellular proteins are allocated equally to each daughter cell.

Prior to the introduction of this assay, there were two dominant methods for quantifying proliferation: cell counting and ³H-thymidine (³H-T) uptake. Cell counting provides an assessment of the total expansion of the culture (i.e., how many cells are present at termination compared to initiation); ³H-T provides an assessment of how many cells were in cell cycle during the pulse phase (usually the last few hours prior to termination). Neither assay could be used to determine the precursor frequency (i.e., what fraction of original cells ever underwent division) nor the distribution of the number of divisions that responding cells underwent. Both of these values have biological relevance.

The introduction of the CFSE assay provided researchers with the ability to derive those values, and, with sophisticated mathematical models, estimate values such as death rates, change in proliferation rates over time, and more (5–7). When good staining techniques are used, the CFSE assay can be used to separately quantify the number of cells in the culture that have not divided vs. those that have divided a specified number of times. In general, the limit on the number of resolvable generations is about 8; this is governed by two limitations. First, there is a limit to how brightly-stained the original population can be; seven divisions results in cells that have 1/128th (<1%) of this fluorescence, which may not be much above autofluorescence. Second, the ability to resolve generations depends on the CV of each fluorescent peak (CV^F); often, as divisions increase, the CV^F may increase, leading to a “smearing” of the populations in higher generations.

ImmunoTechnology Section, Vaccine Research Center, NIAID, NIH, Bethesda, Maryland

The author has a financial interest in FlowJo, one of the software packages discussed in this report.

Received 13 October 2010; Revision Received 11 November 2010; Accepted 20 November 2010

Grant sponsor: National Institute for Allergy and Infectious Diseases, NIH.

*Correspondence to: Mario Roederer, Vaccine Research Center, NIH, 40 Convent Dr., Room 5509, Bethesda, MD 20892-3015, USA.

E-mail: Roederer@nih.gov

Published online 12 January 2011 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.21010

Published 2011 Wiley-Liss, Inc.†

This article is a US government work and, as such, is in the public domain in the United States of America.

Table 1. List of common statistics used to characterize proliferation assays

STATISTIC ^a	TYPE ^b	FORMULA ^c	VALUE RANGE		INTERPRETATION
			MIN	MAX	
Fraction diluted	Ext	$Dil = \frac{\sum_0^i N_i}{\sum_0^i N_i}$	0	1 ^d	Fraction of cells in the final culture that divided at least once
Precursor frequency, %dividing cells ^{efg}	Int	$PF = \frac{\sum_0^i \frac{N_i}{2^i}}{\sum_0^i \frac{N_i}{2^i}}$	0	1	Probability that a cell will divide at least once
Proliferation index ^{ef} ; expansion index ^g	Ext	$EI = \frac{\sum_0^i N_i}{\sum_0^i \frac{N_i}{2^i}}$	1	128 ^h	Fold expansion during culture (ratio of final cell count to starting cell count)
Division index ^f ; replication index ^g	Ext	$RI = \frac{\sum_1^i N_i}{\sum_1^i \frac{N_i}{2^i}}$	1	128 ^h	For responding cells, fold expansion over the culture time
Proliferation index ^g	Int	$PI = \frac{\sum_1^i i \times \frac{N_i}{2^i}}{\sum_1^i \frac{N_i}{2^i}}$	1	7 ^h	For responding cells, average number of divisions they have undergone
Division index ^g	Ext	$DI = \frac{\sum_0^i i \times \frac{N_i}{2^i}}{\sum_0^i \frac{N_i}{2^i}}$	0	7 ^h	Average number of divisions for all cells in the culture
Proliferation variance ^g	Int	$SD^D = \sigma(PI)^i$	0	7 ^h	Standard deviation of the number of divisions (PI)

^aAll statistics except “Fraction Diluted” require modeling to estimate the fraction of cells in each generation. Names given by different software packages are shown.

^bType of value: Ext = extrinsic, dependent on the frequency of responding cells and culture time; Int = intrinsic, a property of individual cells.

^cIn the formulas, *i* is the generation number (0 is the undivided population), and *N_i* is the number of events in generation *i*.

^dValues above 0.6 (60%) may not be comparable (see Fig. 3).

^eAs defined in ModFit.

^fAs defined in FCSExpress.

^gAs defined in FlowJo.

^hIn general, modeling of CFSE data is usually not valid for cells beyond the 7th division, thereby limiting the maximum reliable values for the Index statistics.

ⁱSD^D is computed as the standard deviation on the distribution of number of cell divisions for responding cells in the original population.

Statistical Analysis of CFSE Distributions

Table 1 lists a number of easily-calculable statistics to characterize CFSE distributions; some confusion arises from the same name having been given to distinct statistics by various software packages. Thus, authors should be careful to clarify which statistic they are reporting. Figure 1 and Table 2 show values for these statistics for hypothetical CFSE distributions.¹ All figures and analyses in this article are based on hypothetical data generated for illustrative purposes.

The most commonly reported statistic is the percent of cells in the final population that have divided (Dil). This is easily computed by creating a univariate gate on the CFSE distribution, including all events below the undivided peak (Generation 0).

Perhaps the most important of the other statistics is the precursor frequency (PF), or percent dividing cells. This

defines what fraction of the original population divided at least once during the culture period. For example, when stimulating bulk T cells with antigen, PF is the proportion of T cells that are antigen-specific and respond by proliferating.

A statistic termed proliferation index (ModFit, FCS Express) or expansion index (FlowJo), EI, determines the fold-expansion of the overall culture. For example, if one million cells were stimulated, and EI is 3.0, then the resulting cell count at the end of the culture is three million. This statistic is therefore identical to that obtained by manual cell counting.

A fourth metric, termed division index (FCSExpress) or replication index (FlowJo), RI, determines the fold-expansion of only the responding cells. This reports the expansion capability of the replicating cells rather than for the entire culture (which is EI), and is predicted to be closely related to a ³H-T assay of the culture.

¹Hypothetical distributions were generated by simulation using a script for JMP version 8 (SAS Institute, Cary, NC). Inputs to the script include precursor frequency, average number of divisions for responding cells, variance on the number of divisions (SD^D), variance on the fluorescence distribution within each generation (CV^F), mean fluorescence of the undi-

vided population, and background (autofluorescence). For most simulations, the background and CV^F were held constant across all generations. In general, simulations generated 10³ to 10⁵ events each; statistics were calculated on the generated events using formulas in Table 1. Distributions were also graphed using JMP. The script is freely available from the author.

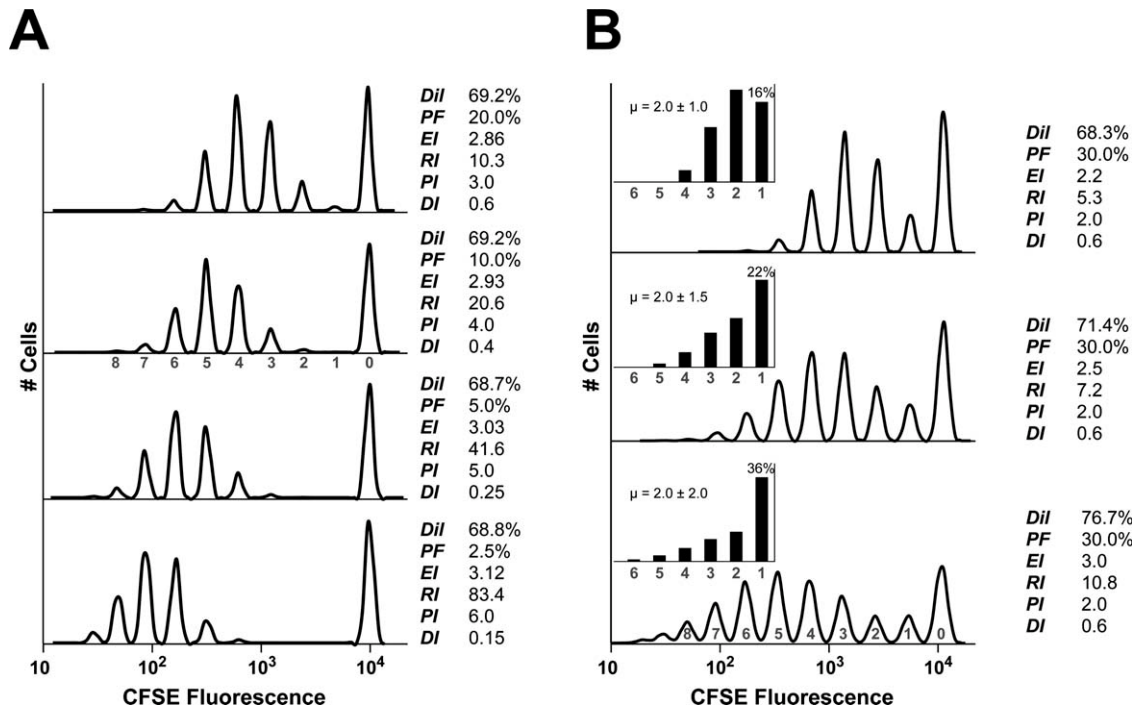


Figure 1. Effect of culture time, precursor frequency, and division variance on proliferation distributions. (A) Shown are hypothetical CFSE distributions for cell cultures, in which the precursor frequency (PF) ranges from 20% (top) to 2.5% (bottom), and the average number of divisions that a responding cell has undergone (PI) increases from 3 (top) to 6 (bottom), as might happen with longer culture times. Grey numbers indicate the generation number. (B) Hypothetical CFSE distributions for cell cultures, for which PF is 30% and PI is 2. The distribution of the number of divisions for cells in the starting population is shown in each inset; the variance on this distribution (SD^D) increases from ± 1 (top) to ± 2 (bottom). The inset shows the fraction of cells that undergo one to six divisions (in each case, 70% undergo no divisions); the percentage of cells that undergo only one division is shown above the right most bar. Notice how the CFSE distribution (distribution of divided cells) looks completely different than the distribution of precursor cells. Very few cells in the starting population undergo more than four divisions (insets), but these rare cells are highly disproportionately represented in the CFSE distribution. For this reason, it is impossible to estimate visually that the average precursor only underwent two divisions in these distributions.

These four statistics are related by the following equations:

$$EI = \frac{1 - PF}{1 - Dil}$$

$$RI = \frac{EI \times Dil}{PF} = \frac{(1 - PF) \times Dil}{(1 - Dil) \times PF}$$

Since both EI and RI can be expressed in terms of Dil and PF, they do not provide independent information; among these four variables are two independent values.

Two additional statistics are computed by FlowJo. FlowJo's proliferation index, PI, is the average number of divisions that all responding cells have undergone since the initiation of the culture. This statistic reflects the proliferative capacity of responding cells. FlowJo's division index, DI, is the average number of divisions for all cells in the original culture (i.e., averaging in zeros for all non-dividing cells). These two values are related by the formula:

$$DI = PI \times PF$$

Thus, the three values only provide two independent pieces of information. Notably, the information provided by PI is

Table 2. Example CFSE modeling statistics

PF (%)	PI	DI	DIL (%)	EI	RI	SD^D
Culture time (no. of divisions)						
10	2	0.2	30.8	1.3	4	0
10	4	0.4	64.0	2.5	16	0
10	6	0.6	87.7	7.3	32	0
Interaction of variables—Figure 1A						
2.5	6	0.15	68.8	3.12	83.4	1
5	5	0.25	68.7	3.03	41.6	1
10	4	0.4	69.2	2.93	20.6	1
20	3	0.6	69.2	2.86	10.3	1
Variance of division capacity—Figure 1B						
30	2.0	0.61	68.3	2.21	5.3	1
30	2.0	0.59	71.4	2.45	7.2	1.5
30	2.0	0.60	76.7	3.00	10.8	2

Notes: The first set of conditions illustrates the relationship of the statistics to a hypothetical culture that was allowed to go for enough time to support 2, 4, or 6 divisions. The other two sets correspond to hypothetical cultures discussed in Figure 1. Inputs to the simulation were the precursor frequency, the average number of divisions a cell in the starting population was to, and the variance on the number of divisions (SD^D). The four statistics, PI, DI, Dil, EI, and RI were computed from the output of the simulation, which generated a sample distribution of 10^5 or 10^6 cells.

independent from Dil and PF (and thus independent from EI and RI).

Finally, another parameter may be useful to report, which is equal to the variance of PI and termed SD^D . It defines how much variability there is in the proliferative capacity of cells in the starting culture. For example, a proliferating population with an SD^D of zero would have all dividing cells in a single division peak (with or without undivided cells present); as the SD^D increases, progeny would be spread out through more and more division peaks. It is more difficult to assign a direct biological meaning to this parameter, and therefore perhaps less useful to report on a routine basis. It can be thought of as a measure of how tightly-regulated the proliferation process was, after the responding cells were stimulated.

Of these statistics, Dil, EI, DI and RI are extrinsic (i.e., depend on the number of cells that proliferate and culture time). On the other hand, PF, PI, and SD^D are intrinsic, and reflect pre-programmed biological properties of the individual responding cells (Table 1).

In summary, among the distribution statistics listed in Table 1, there are as many as four independent values that can be used to characterize cellular proliferation data. These include two of the values from [Dil, PF, EI, RI], one of the values from [PI, DI], and SD^D . Each of these statistics is associated with an underlying biological process; thus, deciding which to report (if not all) should be done with the intent to best convey the scientific conclusions.

Limitations on Proliferation Statistics

Software packages usually use standard nonlinear least-squares modeling to estimate values for these statistics. The use, optimization, and limits of such modeling are not discussed here; this article is focused on the limitations and applicability of the statistics themselves.

Computation of all of statistics except Dil requires the ability to accurately enumerate cells in each generation. This is true whether the enumeration of the cells is performed manually (by gating on each peak), or by software modeling. If distinct peaks cannot be defined, then such enumeration is less likely to give robust, reliable measures, and the only statistic that should be used is Dil. Typically, if the CV^F of the cells within a generation is more than 40%, generational modeling may be difficult. In such a case, modeling requires an unstimulated control (to estimate the CV^F and the position of the undivided peak), with the assumption that these two variables do not differ from the stimulated population, and the assumption that each subsequent division has exactly half the fluorescence of the previous generation (e.g., no loss of fluorescence due to other factors); an example of such modeling is given by Bantly et al. (8). With a small CV^F , these assumptions are unnecessary as the model can estimate the variables directly from the data. A small CV^F can usually be achieved with careful staining technique.

These statistics are usually used to convey the degree of proliferation that has occurred in a culture. However, care must be taken in how they are interpreted. This comes about from the exponential nature of proliferation: a single cell that

divides six times results in 64 cells in the final culture, whereas 16 cells that divide twice also result in 64 cells. Thus, visual interpretation of CFSE histograms is often highly suspect: for these two examples, there are equal-sized peaks to the left of the undivided population; however, the underlying biology is completely different.

Consider the distributions in Figure 1A. These four conditions represent a range of biological conditions: precursor frequencies between 2.5 and 20%, and average divisions between 3 and 6. These distributions illustrate the limitations of Dil and EI statistics: both remain the same across this wide range of proliferative responses. Hence, comparison of these statistics for different cultures should be done while also comparing PF, PI, DI, or RI—only if these latter statistics are also matched between cultures could a conclusion of biological similarity be drawn.

The distributions in Figure 1B reveal limitations on PI and DI. In these hypothetical cultures, a 30% precursor population is responding by dividing, on average, twice. However, the variance on distribution of divisions (SD^D) differs; this could be interpreted that the cultures with smaller SD^D have more tightly-regulated control over how many times a cell divides. Across these three distributions, the values for PI and DI are the same. However, the cultures end up with different numbers of cells (as reflected by EI). The reason for this is that the culture with greater variance has a greater proportion of cells that have divided once, but also a greater proportion cells that have divided (e.g., four or five times). Because of the exponential effect of proliferation, those more-highly dividing cells contribute far more to the final culture, resulting in a greater apparent expansion. It remains to the researcher to define whether or not this is biologically relevant: i.e., while the cultures have the same proportionate response (PF) and, on average, the same proliferative response in terms of divisions (PI), they have a distinct overall proliferative response (EI).

The reliability of all of the statistics except Dil depend on the accuracy with which cells within each generation can be enumerated. As shown in Figure 2, two variables affect this: the relative staining intensity and variance. The intensity of the undivided population (compared to the autofluorescence background) determines how many generations can be identified. Once the diluted fluorescence overlaps with autofluorescence, no further divisions can be distinguished and the ability to quantify further divisions is eliminated. (It should be noted that, for the purposes of calculating the statistics, far more information is derived from the earlier generations than from later. Thus, distributions with reasonable proportions of events in the lower-numbered generations will still generate values that are reasonable estimates. However, as the significant majority of cells enter the low end of the fluorescence distribution near background, the metrics increase asymptotically to their maximum values.)

The staining variance (CV^F) also impacts the ability to resolve generations (Fig. 2B). Here, modeling of cell numbers in each division could be performed by making assumptions as to the position of each generation—but the error in these

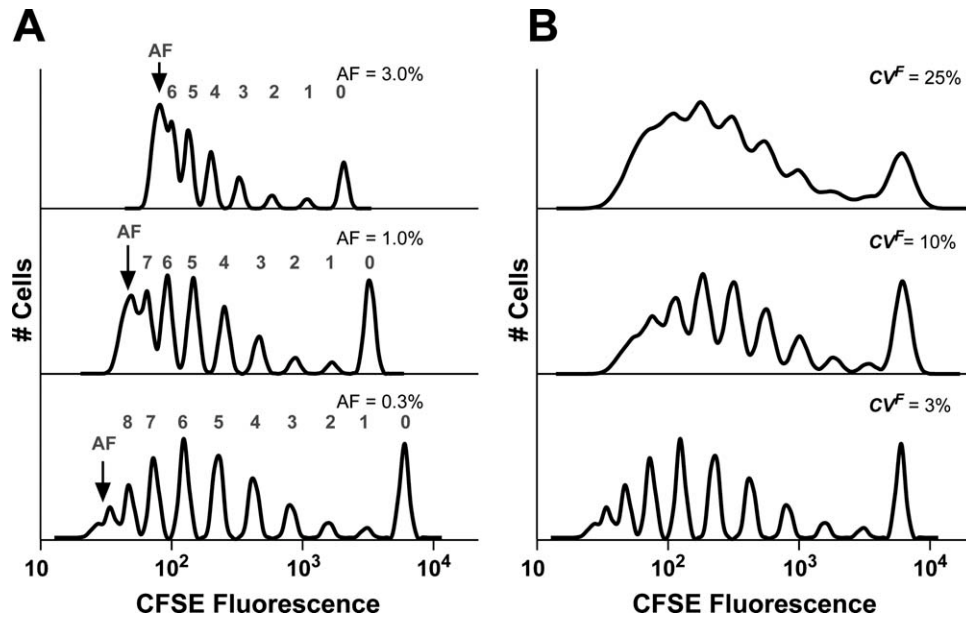


Figure 2. Effect of staining intensity and variance on distributions. (A) Hypothetical distributions were generated for a 20% precursor frequency (PF), undergoing 3 ± 2 ($PI \pm SD^D$) divisions. The effect of weaker staining (or, equivalently, higher autofluorescence background) is illustrated. In the top graphic, the autofluorescence (AF) is modeled as 3% of the fluorescence of the undivided population; sensitivity is increased 3-fold and 10-fold for the other two distributions. The ability to resolve the higher generations disappears as the diluted CFSE fluorescence becomes invisible against the autofluorescence background. Grey numbers indicate to the generation number. (B) Similar distributions as in (A), but with increasing variance of the fluorescence distribution (CV^F). Resolution of generations becomes impossible as CV^F increases to the point where adjacent generations overlap significantly.

estimations will reduce precision in the statistics. It is difficult to assess the accuracy of models under conditions where the staining variance is large, and therefore difficult to determine the accuracy of the reported statistics. Certainly, care must be taken when interpreting any statistics (other than Dil) based on such data.

For pragmatic reasons, Dil is the most-commonly reported statistic, and in some case is the only statistic that can be reliably computed (e.g., Fig. 2). However, Dil conveys a limited amount of information. Dil is perforce limited to 1 (100%) as a maximum value. As its value rises above 60%, the ability to distinguish biological variation becomes far more difficult. As shown in Figure 3, Dil rises quickly with the number of divisions (PI) and linearly with the precursor frequency (PF), up to about 50%. After this, however, it asymptotically rises toward 100%, with the result that fairly large deviations in divisions or precursor frequency become difficult to resolve in the face of experimental error. As such, comparison of distributions with Dil values above 60% should be done with caution.

Impact of Cell Death and More Complex Analyses

The simple modeling described above makes a major assumption: there is no cell death occurring during the culture. Differential death of cells depending on their stimulation history can significantly impact the interpretation of the proliferation statistics. Quantifying cell death is straightforward, but requires absolute cell counts at the time of analysis (C) and at the initiation of the culture (C_0). Assuming equal cell death

across all generations, the fraction of cells that have died (D) is given by these counts and the expansion index (EI):

$$D = 1 - \frac{C}{C_0 \times EI}$$

Nonzero values of D are indicative of unaccounted for cell death and may impact the accuracy of the estimation of statistics such as those in Table 1. Graphical methods have been employed to estimate cell death for proliferation studies, to achieve more accurate estimates of division capacity (9); if the data reveals substantial cell death, consideration to application of more sophisticated analyses should be given.

Several papers have been published that propose and compare more sophisticated modeling of CFSE distribution data. From such models, one can estimate additional parameters such as the lag time before initiation of division, cycling time, individual deaths rate for each generation, and other parameters (5–7). Fundamentally, these models all require accurate estimation of the representation of cells in each generation, as do the statistics shown in Table 1; in some cases analyses of the culture over time must be performed. Analysis of the more sophisticated models is beyond the scope of this perspective; researchers should qualify any such models within the scope of their own data using standard approaches.

Qualifying Statistical Analyses for Comparison of Distributions

The discussions above are meant to guide interpretation of CFSE statistical analyses, and are not meant to replace rig-

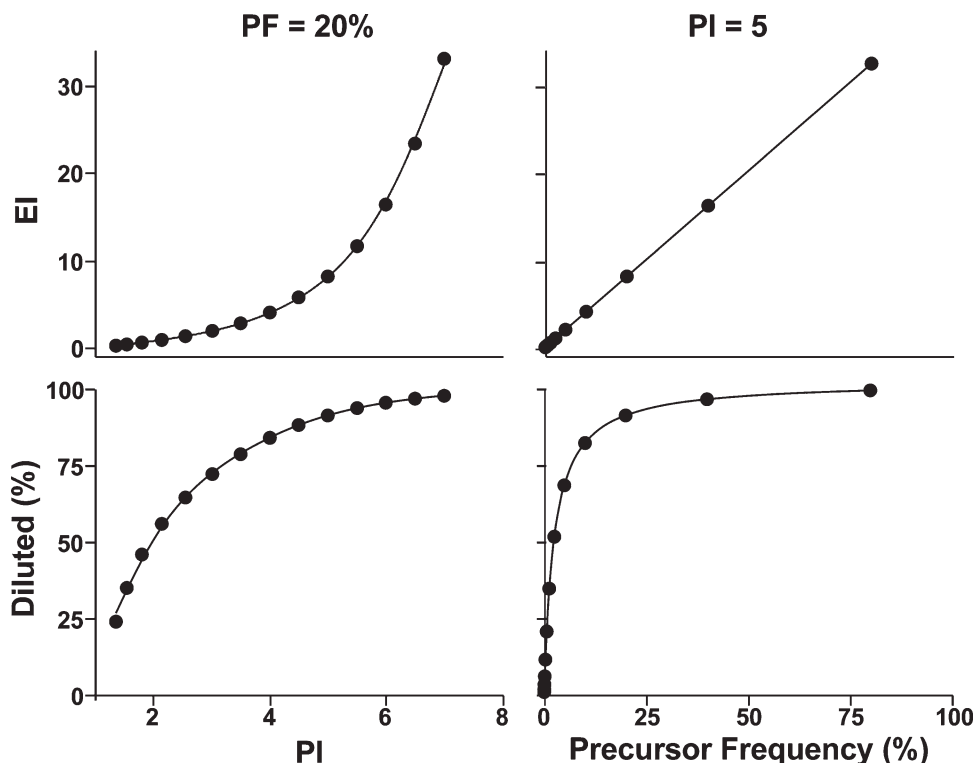


Figure 3. Interdependence of proliferation statistics. (Left panels). Hypothetical distributions were generated using a 20% precursor frequency (PF) with a range of average divisions (PI); for each, a $SD^D \pm 1$ division was used. EI is related exponentially to the number of divisions; Dil is non-linearly related. (Right panels). Hypothetical distributions were generated for responding cells undergoing 5 ± 1 ($PI \pm SD^D$) divisions, with a range of precursor frequencies. EI is linearly related to precursor frequency over this range; Dil saturates quickly with increasing precursor frequency.

orous qualification of assays in the lab. Any of the statistics shown in Table 1 can be used for the purposes of comparing proliferative responses between samples. It is the responsibility of the researcher to establish, within the context of their specific protocols and conditions, what the reproducibility, sensitivity, precision, linearity (if applicable), robustness, and other performance characteristics of the particular statistic may be.

Dil remains the most commonly reported statistic in the literature. The primary limitation of this metric is that it conveys a limited amount of information, and it becomes significantly less informative at values above 60%. In many experiments, such as monitoring lymphocyte proliferation, cells divide very rapidly (the cycle time can be as little as 12 h). Thus, small experimental variability can lead to fairly large variance in actual cycle time during the culture—which itself can lead to exponentially increased variability in some computed statistics such as EI and (at lower values) Dil. Qualification of the assay in-house can assure both the researcher and the reviewers that reported differences in CFSE-measured proliferation assays are indeed due to biological variation rather than experimental error; for publications reporting only Dil, it is incumbent on the authors to demonstrate reproducibility, sensitivity, and precision.

Once the reproducibility (variance) for CFSE statistics have been established, it is possible to compare these values between culture conditions. Standard statistical comparisons

(e.g., Wilcoxon Rank or Student's t tests) can be used to infer differences. When comparing between test and control cultures, the difference or ratio of these statistics can be considered as a metric. Note that some of the statistics have a minimum value of 1, while others have a minimum value of 0; Munson's recent report (10) defines different comparison formulas to be used depending on this minimum.

SUMMARY

CFSE-based proliferative assays provide a rich amount of information to characterize the response and subsequent expansion capacity of cell cultures. However, the exponential expansion leads to an inability to infer the actual amount of proliferation that has occurred without careful application of the proper algorithms and statistics. In particular, visual representation of the distributions should not be used solely to convey relative proliferative capacity. Singular use of the “% diluted” statistic, easily derived by gating on the CFSE distribution, must be approached with great caution. There are a variety of statistical analyses that can be applied to such data; different types of statistics can reveal different underlying biological mechanisms. In any case, care must be taken to ensure that assays and analyses are performed in a rigorous, reproducible fashion, and publications of these types of analyses should indicate the degree to which the assays have been qualified.

The author thanks the members of the laboratory, the ImmunoTechnology Section at the Vaccine Research Center, for inspiration and constructive discussions, and the Tree Star Applications Scientists and Mark Munson for critical discussions about the statistics of CFSE modeling.

LITERATURE CITED

1. Lyons AB, Parish CR. Determination of lymphocyte division by flow cytometry. *J Immunol Methods* 1994;171:131–137.
2. Lyons AB, Doherty KV. Flow cytometric analysis of cell division by dye dilution. *Curr Protoc Cytom* 2004; Chapter 11, Unit 9.11.
3. Parish CR, Glidden MH, Quah BJ, Warren HS. Use of the intracellular fluorescent dye CFSE to monitor lymphocyte migration and proliferation. *Curr Protoc Immunol* 2009; Chapter 9, Unit 4.9.
4. Quah BJ, Warren HS, Parish CR. Monitoring lymphocyte proliferation in vitro and in vivo with the intracellular fluorescent dye carboxyfluorescein diacetate succinimidyl ester. *Nat Protoc* 2007;2:2049–2056.
5. De Boer RJ, Ganusov VV, Milutinovic D, Hodgkin PD, Perelson AS. Estimating lymphocyte division and death rates from CFSE data. *Bull Math Biol* 2006;68:1011–1031.
6. Nordon RE, Nakamura M, Ramirez C, Odell R. Analysis of growth kinetics by division tracking. *Immunol Cell Biol* 1999;77:523–529.
7. Yates A, Chan C, Strid J, Moon S, Callard R, George AJ, Stark J. Reconstruction of cell population dynamics using CFSE. *BMC Bioinform* 2007;8:196.
8. Bantly AD, Gray BD, Breslin E, Weinstein EG, Muirhead KA, Ohlsson-Wilhelm BM, Moore JS. CellVue Claret, a new far-red dye, facilitates polychromatic assessment of immune cell proliferation. *Immunol Invest* 2007;36:581–605.
9. Hawkins ED, Hommel M, Turner ML, Battye FL, Markham JF, Hodgkin PD. Measuring lymphocyte proliferation, survival and differentiation using CFSE time-series data. *Nat Protoc* 2007;2:2057–2067.
10. Munson ME. An improved technique for calculating relative response in cellular proliferation experiments. *Cytometry A* 2010;77A:909–910; erratum 77A:1177.