

Statistical Analysis of ELISPOT Assays

Marcus Dittrich and Paul V. Lehmann

Abstract

Cytokine ELISPOT assays have emerged as a powerful tool for the detection of rare antigen-specific T cells in freshly isolated cell material, such as blood. While ELISPOT assays allow one to directly visualize and count extremely low frequencies of cytokine-secreting T cells among millions of nonsecreting bystander cells, the interpretation of ELISPOT data can become ambiguous when (a) spot numbers in antigen-containing wells are low, (b) spot counts in negative control wells are elevated, and particularly (c) when both of the above occur simultaneously. Thus, the primary task, even before statistics are employed, must be the optimization of the basic assay parameters and reagents such that the assay yields low background signal in the negative-control wells and the maximal number of antigen-induced spots in test wells, i.e., the signal-to-noise ratio is maximized. Furthermore, the use of proper spot-size gating parameters for data analysis is indispensable for screening out irrelevant background spots, and thus increasing the signal-to-noise ratio. The goal of most ELISPOT experiments is to identify positive T-cell responses as defined by a significantly elevated spot count in antigen-stimulated wells over the nonstimulated medium-control or negative-control antigen. In this chapter, we conclude that – with some limitations – the *T*-Test and related statistical methods which rely on the assumption of normal distribution are suitable for identifying positive ELISPOT results.

Key words: ELISPOT, Statistical analysis, Signal to noise, Spot-size gating parameters, PBMC, ImmunoSpot, MATLAB, ANOVA, Wilcoxon Rank Sum Test, *T*-Test, Poisson distribution

1. Introduction

Because ELISPOT typically aims for the detection of rare antigen-specific cells within a variable background, the notion of “fuzzy in, fuzzy out” applies to ELISPOT data – perhaps more so than for other immunoassays. Statistical analysis cannot itself substitute for experimental stringency in performing ELISPOT assays which provide clear, unambiguous results. Preceding the statistical analysis of ELISPOT data, therefore, great attention must be given to establishing spot counts (which are in part based on the statistical analysis of the spot-size distributions) before the counts obtained within an experiment are subject to further statistical analysis.

1.1. “Precise in”: Optimizing Signal- to-Noise

1.1.1. The Choice of Membrane

In cytokine ELISPOT assays, cytokine secretion from individual cells is measured. In typical experiments, 96-well PVDF membrane plates are precoated with a cytokine-specific capture antibody, e.g., a suitable anti-IFN- γ antibody. Due to its fractal surface and high hydrophobicity, PVDF membranes have been found to outperform most other membranes which had been used previously for cytokine ELISPOT assays (1), and therefore the use of PVDF membranes is highly recommended for obtaining optimal results when performing these tests. Historically, cytokine ELISPOT assays had been very fragile and hard to reproduce before the introduction of PVDF membranes.

1.1.2. The Choice of Cell Numbers Plated

In the next step of the assay, the test cell material (e.g., human PBMC) is plated into the precoated wells, both with and without antigen. For T-cell assays, the test cells need to be dense enough to allow for optimal contact between T cells and antigen-presenting cells (APC), but the cells must not be overcrowded, as it is essential that each T cell sits directly on the membrane so that its secretory product can be effectively captured. For human PBMC, 100,000–800,000 cells can be plated per well into 96-well plates with the number of antigen-induced spots per number of PBMC plated following a linear relationship (2). (In contrast, when a monolayer of APC is provided, even single T cells can be plated per well and tested (3)).

For the purpose of economizing cell utilization, ELISPOT assays are frequently performed with 100,000 PBMC per well. However, when low-frequency T cells are to be assayed, increasing cell numbers up to 800,000 per well increases the signal in a directly linear fashion without disproportionately introducing noise into the system. When ambiguous results are obtained through the testing of low cell numbers (e.g., 100,000 per well in 96-well plates), retesting these PBMCs with higher cell numbers can provide clear results. The cells can be readily retested, since protocols have been developed that permit an investigator to freeze PBMC such that their functionality in ELISPOT assays is retained and is effectively identical to fresh cells upon thawing (4). Also, 6-well PVDF membrane plates are being introduced that allow for increasing the sample size per well tenfold respective to 96-well plates, thus increasing 10 \times the signal-to-noise resolution of ELISPOT assays.

1.1.3. Establishing the Correct Spot Counts in Antigen-Stimulated Wells

Upon antigen stimulation, the antigen-specific T cells engage in the secretion of cytokine which is captured around the cell by the membrane-bound capture antibody. Overall, the size and density of the resulting cytokine “spot” is reflective of the quantity of cytokine produced by the cell, and the morphology of the spots reveals the secretion kinetics: i.e., fuzzy spots reflect a rapid secretion kinetics, whereas sharp spots are indicative of slow analyte release (5) (see also Chapter 11 in this volume). Invariably, T-cell ELISPOT assays provide a wide spectrum of spot sizes and morphologies, irrespective of whether T-cell clones or primary antigen-specific

T-cell populations are studied, and of the analyte (3, 6). Thus, the question arises as to how to reliably establish spot counts, i.e., how to distinguish whether larger spots result from cell clustering, and how to determine the smallest spots which still are to be counted. Tremendous variations in counts can be seen in the absence of clear, unambiguous criteria for counting. Such criteria can be objectively established, however, by understanding the rules that underlie spot-size distributions. For T-cell clones, as well as primary T cells secreting various cytokines, it has been determined that the spot-size distribution always follows log-normal distribution (3, 6). Thus, proper analysis establishes the spot-size distribution for an assay and subjects the size distribution curve to statistical analysis, automatically setting the gates at 99.7% confidence to establish upper and lower limits, respectively, for the largest and smallest spot sizes which still belong to that distribution. In this way, spots that exceed this size gate can be recognized as clusters, and spots that are smaller than the lower limit are gated out. Such statistical analysis permits objective, user-independent counting while spot counts established without such analysis are subjective, unreliable numbers contributing to “fuzzy in.”

1.1.4. Correct Spot Counts in Negative-Control Wells

ELISPOT assays aim at establishing frequencies of antigen-specific T cells, i.e., measuring antigen-induced spots over medium-control or irrelevant antigen. The negative-control wells, however, can also contain cytokine-secreting cells (which generate spots). Such background spots typically are produced by cells of the innate immune system which also secrete the cytokine in question. One of the primary challenges of ELISPOT analysis is to differentiate between such background spots and antigen-induced, T cell-derived spots. This frequently can be accomplished by size gating. T cells typically have a substantially higher cytokine secretion rate (resulting in larger spot-size distributions) than cells of the innate immune system (resulting in a smaller spot-size spectrum) (7). Thus, analogous to gating in flow cytometry, the smaller spot-size distributions seen in the negative controls can be gated out, allowing for identification of the larger, T cell-derived spots in the antigen-stimulated wells. ELISPOT gating can be done automatically (and thus, objectively) following statistical principles (for more on this topic, please refer to the Chapter 13 in this volume). Medium-control and antigen-stimulated spot counts generated without such counting principles are literally meaningless numbers (much like ungated flow cytometry frequencies), and provide a “fuzzy in” for which subsequent statistical analysis cannot compensate.

1.1.5. Lowering the Medium Background: Thus Increasing Signal to Noise

ELISPOT assays, like most cellular assays, have traditionally been performed using culture media which contain serum. Serum, however, is a major assay variable. It contains cytokines that stimulate or suppress the test PBMC, many times resulting in an increased

background spot production (triggering cytokine secretion by cells of the innate immune system) that is not necessarily accompanied by an increased antigen-specific response by T cells. Because the cytokines present in serum bind to high-affinity receptors on PBMC, even brief exposure of PBMC to a stimulatory or suppressive serum during freezing/thawing or washing the PBMC can effectively ruin an ELISPOT assay. The use of serum-free media for freezing, washing, and testing is, therefore, highly recommended providing typically higher signal-to-noise ratios than even the “best” sera selected for this purpose (2).

1.2. Testing Antigen- and Medium-Control Wells in Replicates

Because in PBMC, antigen-specific T cells typically occur in low frequencies, ELISPOT assays frequently need to detect small increases in antigen-induced spot counts versus the negative control. Therefore, to increase the statistical relevancy of the resulting spot counts, it has become convention to measure cytokine production in medium-control and antigen-containing wells in replicate wells, typically triplicates. After proper counting of replicate wells for each condition (assuring a “high resolution in” – see above), further analysis can be done to establish whether there are indeed more spots seen in the antigen-containing wells. Many times, the spot counts in replicate wells are consistent, and the differences between medium-control and antigen-stimulated wells are large – in such cases, the interpretation of the results is clear-cut. However, the interpretation of ELISPOT data can become ambiguous when (a) the spot numbers in antigen-induced wells are low, (b) spot counts in negative-control wells are elevated, and (c) when both of the former occur simultaneously. Here is where the grey area of ELISPOT data interpretation lies, with different solutions, mostly empirical, propagated to resolve the problem (8). Further, the choice of triplicates is entirely empirical, with its statistical foundation not having been sufficiently clearly delineated – more on this below.

1.3. Counting Apoptotic Cells in Addition to Live and Dead

Many times, immunizations cause only a moderate increase in the frequency of antigen-specific T cells. Such differences can be detected when comparing antigen-induced responses in preimmunization PBMC samples with PBMC obtained at various time points following immunization. Such longitudinal testing is mostly done with cells that have been stored/shipped for many hours, sometime days, before being frozen and/or tested. Protracted handling can damage the cells, resulting in increased numbers of dead as well as apoptotic cells. Standard Trypan Blue cell counting can discern between live and dead cells, but does not detect cells that have entered upon the irreversible path of apoptosis. Apoptotic cells are still alive, and are counted as such with Trypan Blue and other live/dead counting methods. Apoptotic T cells should be reckoned as effectively dead for functional assays, however, because they are refractive to antigen stimulation and will be dead before

they could produce cytokine. ELISPOT assay harmonization panels have, therefore, repeatedly emphasized the need to count apoptotic cells, in addition to live and dead cells, and to correct the live count by subtracting the apoptotic count. Various dyes are available to stain apoptotic, live, and dead cells with different colors for visual counting using standard hemocytometers and UV microscopy or, more conveniently and precisely, by image analysis or flow cytometry.

1.4. Empirical Versus Statistical Evaluation of Medium-Control Versus Antigen-Induced Spot Numbers

A key issue in ELISPOT data evaluation is setting the threshold beyond which a response should be considered positive. Several different approaches are described in the literature which may be classified mainly into two different categories: empirical and statistical (8). Empirical approaches generally do not have sound theoretical justifications, but can rather be regarded as rules of thumb. Such commonly used approaches require an arbitrarily designated minimum difference between the mean spot counts for the antigen-containing wells and the negative-control wells, mostly medium. Other empirical rules demand the ratio of antigen-elicited spots to spot numbers in the medium control to exceed a certain value – here, often a twofold increase is considered to be the cutoff for a positive response. Various combinations of these rules have also been proposed, such as combining the minimum threshold R_0 for spot count per 10^6 cells with a minimum threshold C_0 ratio of antigen to control; responses exceeding these values are designated as positive, yet the two thresholds may be chosen in such a way as to obtain a false-positive rate <0.01 (9). The advantage of empirical rules is their simplicity. Their major drawback is their lack of reliability when it comes to detecting weak responses. Since they have no precise theoretical foundation, empirical rules have no claim to universality, since they do not rely on explicit model assumptions which could be tested for and thus ensure the general applicability of the proposed procedure.

In contrast to empirical rules, statistical tests rely on a theoretical background. Here, the observed data (or a statistical result of the data, such as the mean) is tested for its compatibility with a (usually, purely random) null model (null hypothesis) to obtain a p -value. The p -value is consequently a measurement of the probability of the observed data under the chosen null model, where low p -values indicate evidence which is highly improbable to be observed under the null model and suggests a rejection of the null hypothesis. Thus, the p -value itself delivers a quantitative measurement of probability, and not a binary yes/no decision for the identification of a positive response. A call for a positive response is then usually claimed when the p -value is below a certain threshold called the significance level α , where values smaller than 0.05 or 0.01 are typically considered significant. Hence, given the null hypothesis is

true, the probability of Type I errors (false-positive calls) would be 5% (for $p=0.05$) or 1% (for $p=0.01$).

Note that with this definition the p -value always needs to be interpreted in relation to the particular null model under consideration; in fact, it could also be regarded as a measurement of distance from a certain null model. Moreover, the null model often relies on specific assumptions, and the violation of this assumption leads to invalid results. Exactly this point is presently the bottleneck for scientifically validated statistical analyses of ELISPOT data, since no data or predictions are currently available regarding the distribution and inter-/independence of ELISPOT data points (spots).

Since as of yet, distributional assumptions are not available for ELISPOT data evaluation, some researchers advocate the use of resampling-based statistical approaches that do not involve such assumptions (10). It allows to perform single or multiple testing corrections against a randomly permuted null distribution in one step. Simulation studies with data drawn from Poisson and Negative Binomial Distributions (which account for an overdispersion effect) have been conducted for various assays. The authors advocate the usage of permutation tests which require the generation of a background distribution using a large number (e.g., 10,000) of random permutations and comparison of the value of the observed test statistics from the real (unpermuted) data against this background distribution. An empirical p -value can then be computed as the fraction of test statistics with a value greater than or equal to the value from the permuted background dataset. To generate a sufficiently large background distribution (in the case of low replicate numbers), the permutations can be performed over all different antigens tested (10). A general problem with the global permutation approach is that the p -value of one antigen is dependent on the responses to other antigens, so the methods may fail to recognize a moderate response in the presence of a large response against one of the other antigens (11). Thus, the permutation of the controls and of each antigen separately has been proposed (8, 11). Generally, resampling-based techniques are rather computationally intensive (which, however, is not a great problem with modern computers) and often not as easily available to many experimental researchers as standard testing methods. Furthermore, when only a low number of replicates is available, as is typically the case in ELISPOT experiments, the p -value distribution can become “granular” and, due to the limited number of different permutations possible, it may become impossible to obtain p -values below a certain threshold. Empirical evaluation of ELISPOT data or resampling-based statistics is a valid approach to ELISPOT analysis only when the distributional properties of spots are not known. Therefore, we set out to establish those experimentally, allowing for the introduction of assumption-based statistics.

2. Materials

For ELISPOT counting, and thus for generating the numeric data that can be subject to further statistical analysis, we recommend the ImmunoSpot Series 5 Instrumentation and Software by CTL, Shaker Hts, Ohio, because it offers user-independent, statistics-based setting of counting parameters for spot recognition and gating. Serum-free media for freezing, thawing, and testing of PBMC, as well as PBMC reference samples with established antigen-specific response levels, are also available from CTL.

1. Regular *T*-Test statistics can be computed using standard spreadsheet software, such as Microsoft Excel which is included in the Microsoft Office Suite.
2. For advanced statistical analysis of ELISPOT data (like fitting linear models, testing distributional assumptions, or permutation tests), real statistical software packages are required. A prominent example is the freely available statistical software package R (12) which can be downloaded from <http://cran.r-project.org/>. This is a very powerful statistical computer language which is broadly extensible by a wide range of additional packages. To exploit the full power of R, some skills in computer programming are recommended. Similarly, the commercial MATLAB software (<http://www.mathworks.com/>) includes a statistics toolbox which may also be used for analyzing ELISPOT data.
3. Alternatively, other commercial statistical software packages featuring a graphical user interface are available, like GraphPad Prism® (<http://www.graphpad.com/prism/prism.htm>), SPSS (<http://www.spss.com>), SAS (<http://www.sas.com>), or Statistica (<http://www.statsoft.com/>).

3. Methods

3.1. Normal Distribution of ELISPOT Data for >30 Spots per Well

To address the distributional properties of ELISPOT data, we performed experiments with a transfected cell line that constitutively secretes IFN- γ . Increasing numbers of such cells were plated into an IFN- γ ELISPOT assay, with 192 replicate wells for each cell number, and the distribution of the spot counts was studied in detail. The results showed that, for lower number of cells (<20), the spot counts approximated the Poisson distribution closer than the normal distribution, whereas wells with spot counts higher than 30 gave an excellent fit to the normal distribution (Fig. 1). Several statistical tests are available to verify whether the normality assumption holds. These include the Shapiro-Wilks, Kolmogorov-Smirnov, or the Anderson-Darling tests – the latter, in particular, is

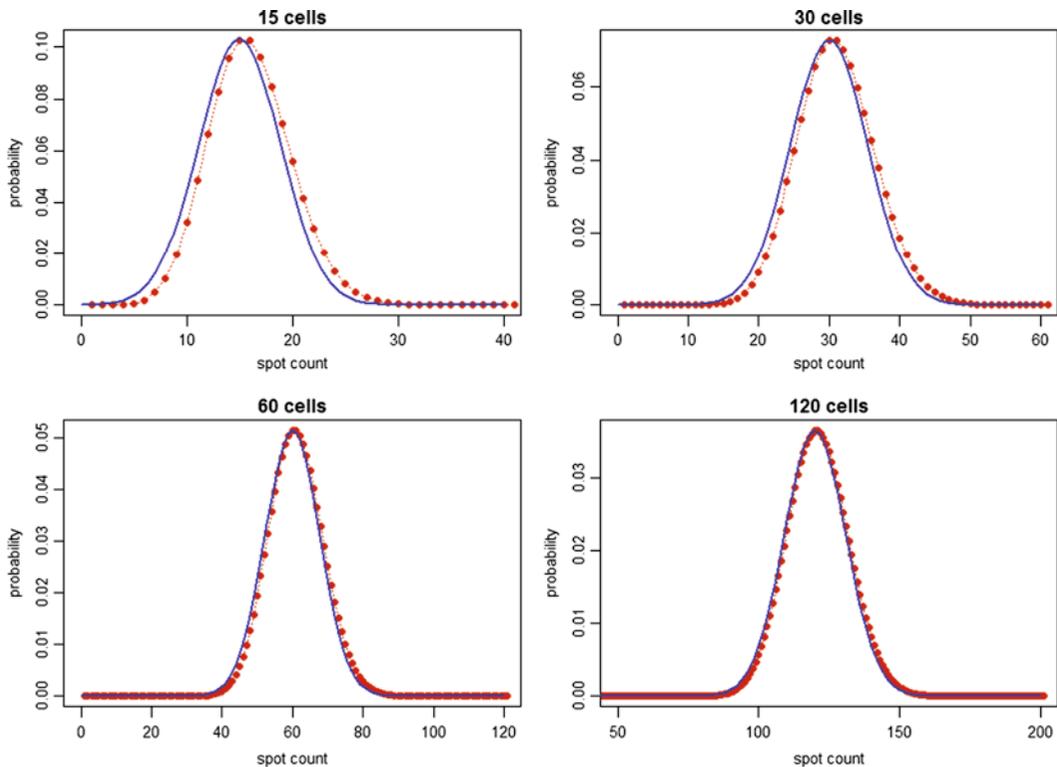


Fig. 1. Measured versus theoretical distribution of ELISPOT counts. IFN- γ -transfected CHO cells were plated into an IFN- γ ELISPOT assay at 15, 30, 60, and 120 cells per well, with 192 replicate wells per cell number, developed and counted with ImmonoSpot[®] Software. Distributions for spot counts are depicted for a Poisson distribution (*dotted lines*) and a normal distribution (*the solid line*) with the means and a standard deviation corresponding to the square root of the means.

one of the most powerful tools for detecting most departures from normality (13). Applying the Anderson Darling test to the above data, we found no significant deviation from the normal distribution for the set of wells with 30 (p -value 0.058), 60 (p -value 0.52), and 120 (p -value 0.23) cells per well. For the wells with 15 cytokine-secreting cells, we found only a slight deviation from the normal distribution (p -value 0.02). Therefore, we conclude that for ELISPOT data in the range of >15 spot counts per well a normal distribution of the data can safely be assumed, and statistical tests which require normally distributed data should be applicable. Thus, the T -Test is well-suited for the statistical analysis of such data, and also all T -Test-related statistics used in linear models and, in particular, the Analysis of Variance (ANOVA), should apply.

3.2. Poisson Distribution of ELISPOT Data with 15 Spots or Fewer per Well

Since optimized ELISPOT assays measure absolute numbers of secreting cells, they produce real counts of individual cells, and not semiquantitative data or ratios. Thus, one could expect ELISPOT data to follow a Poisson distribution. Indeed, analyzing the ELISPOT counts from our transfected cell line, we found no

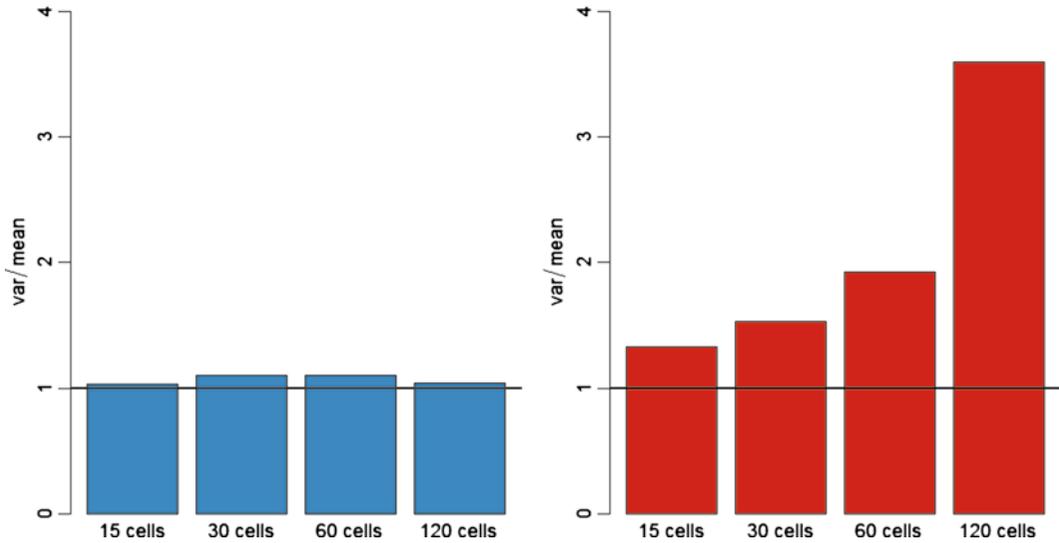


Fig. 2. Overdispersion of ELISPOT count data. *Left panel:* Simulated count data drawn from a Poisson distribution with mean parameter λ for 15, 30, 60, and 120 cells per well ($n=192$ each) exhibit a variance equal to the mean as expected. *Right panel:* Real spot count data obtained with the transfected cell lines display a classical overdispersion effect, that is, the variance increases with the mean.

significant deviation from the Poisson distribution at low counts (15 spots) per well. In the low-count range (<15 spots), ELISPOT data should thus be assumed to have a Poisson distribution. In contrast to the normal distribution, the Poisson distribution supports only positive values and is skewed rightward in the range of low intensities, i.e., small mean values (especially in the range <5). For higher mean values, however, the Poisson distribution becomes more symmetric and approximates a normal distribution (Fig. 1). Furthermore, it should be noted that for Poisson-distributed data there is a coupling between the mean and the variance (with variance = mean) as the Poisson distribution is characterized by only a single parameter. In effect, this means that the variance increases proportionally as the mean increases, meaning that the ratio of the variance to the mean is the same for any response intensity (Fig. 2; left side). As is the case with real count data from many different areas, one can also observe an overly large variance with increasing mean values for ELISPOT count data, an effect called overdispersion (Fig. 2; right panel).

The use of the Negative Binomial Distribution has been proposed to model count data for ELISPOT counts (10). This distribution can be regarded as an extension of the Poisson distribution. It has two parameters which allow modeling of both the variance and the mean, and it is thus able to handle the overdispersion problem. A common used remedy in cases of assumed deviations from the normal distributions is the usage of nonparametric tests, such as the well-known Wilcoxon Rank Sum Test. This might entail

a loss of power in case of normally distributed data, but can provide an improved power over the *T*-Test in cases of considerable deviation from the normal distribution. A further concern is that data originating from two different distributions need to be compared when the medium background falls into the <20 spots category while the antigen-induced spots are >30 (see Note 1).

**3.3. The Pragmatic
Solution: The
Applicability of the
T-Test for ELISPOT
Data Analysis**

For pragmatic purposes, it is important to know how the Poisson distribution of low-count data affects the performance of the *T*-Test (which requires normally distributed data) in the low-count range. Our data from simulations indicate that even for Poisson-distributed data in the low-count range, no increased number of false positives for a significance level of 0.05 is to be expected. However, the *T*-Test may fail to identify weak responses. The extent to which the power of the *T*-Test is affected by a Poisson distribution in the low-signal range needs further investigation. Increasing the number of replicates in any case increases the probability of detecting weak responses. The triplicates typically used are certainly a minimum. The determination of a reasonable number of replicates, however, depends on several variables and often requires an advanced statistical power analysis for a particular experimental setup and expected effect size (see Note 2).

In conclusion, the usage of empirical rules, such as using fold changes with a fixed threshold, should in general be discouraged. Albeit enjoying some popularity, such empirical rules do not incorporate any information about variance, and thus provide no measurement of confidence. Instead, the application of solid statistical tests is highly recommended. The question of which statistical tests are best suited for ELISPOT evaluation is an important one, and is certainly not completely resolved (see Note 3). Our results from transfected cell lines indeed suggest a Poisson distribution, but only for data in the low-count range. For spot counts >15, no significant deviation from the normal distribution could be observed, indicating that spot counts in that range can be assumed to be normally distributed. This implies that most standard statistical tests, which are valid for normally distributed data, like the *T*-Test and ANOVA models, are applicable in this setting. In any case, more investigations (experimental data as well as computer simulations) are needed not only to assess potential effects of deviations from normality in the low spot-count range, but also to develop novel methods for the analysis of ELISPOT data.

4. Notes

1. Although being essentially a nonparametric test, the Wilcoxon Rank Sum Test requires the data to be independent and identically distributed (iid). While the first assumption of independence

is not affected, the second assumption of identical distributions is obviously violated. Whether this has practical implications for the statistical analysis of ELISPOT data has not yet been investigated, and still requires more in-depth research.

2. On a practical note, it is recommended to perform all tests using at least triplicates, and if the results fall along the borderline, to retest cryopreserved aliquots at higher cell numbers, and with more replicates.
3. An important aspect here is the understanding of the distribution of ELISPOT data. Since the assay provides counts of individual antigen-specific T cells that are randomly distributed during pipetting into the wells, on theoretical grounds, one might expect ELISPOT data to follow a Poisson distribution.

References

1. Forsthuber, T., Yip, H. C., and Lehmann, P. V. (1996) Induction of TH1 and TH2 immunity in neonatal mice. *Science* **271**, 1728–1730.
2. Zhang, W., Caspell, R., Karulin, A. Y., Ahmad, M., Haicheur, N., Abdelsalam, A., et al. (2009) ELISPOT assays provide reproducible results among different laboratories for T-cell immune monitoring – even in hands of ELISPOT-inexperienced investigators. *J Immunotoxicol* **6**, 227–234.
3. Hesse, M. D., Karulin, A. Y., Boehm, B. O., Lehmann, P. V., and Tary-Lehmann, M. (2001) A T cell clone's avidity is a function of its activation state. *J Immunol* **167**, 1353–1361.
4. Kreher, C. R., Dittrich, M. T., Guerkov, R., Boehm, B. O., and Tary-Lehmann, M. (2003) CD4+ and CD8+ cells in cryopreserved human PBMC maintain full functionality in cytokine ELISPOT assays. *J Immunol Methods* **278**, 79–93.
5. Ott, P. A., Tary-Lehmann, M., and Lehmann, P. V. (2007) The secretory IFN-gamma response of single CD4 memory cells after activation on different antigen presenting cell types. *Clin Immunol* **124**, 267–276.
6. Nowacki, T. M., Kuerten, S., Zhang, W., Shive, C. L., Kreher, C. R., Boehm, B. O., et al. (2007) Granzyme B production distinguishes recently activated CD8(+) memory cells from resting memory cells. *Cell Immunol* **247**, 36–48.
7. Guerkov, R. E., Targoni, O. S., Kreher, C. R., Boehm, B. O., Herrera, M. T., Tary-Lehmann, M., et al. (2003) Detection of low-frequency antigen-specific IL-10-producing CD4(+) T cells via ELISPOT in PBMC: cognate vs. nonspecific production of the cytokine. *J Immunol Methods* **279**, 111–121.
8. Moodie, Z., Price, L., Gouttefangeas, C., Mander, A., Janetzki, S., Lower, M., et al. (2010) Response definition criteria for ELISPOT assays revisited. *Cancer Immunol Immunother* **59**, 1489–1501.
9. Dubey, S., Clair, J., Fu, T. M., Guan, L., Long, R., Mogg, R., et al. (2007) Detection of HIV vaccine-induced cell-mediated immunity in HIV-seronegative clinical trial participants using an optimized and validated enzyme-linked immunospot assay. *J Acquir Immune Defic Syndr* **45**, 20–27.
10. Hudgens, M. G., Self, S. G., Chiu, Y. L., Russell, N. D., Horton, H., and McElrath, M. J. (2004) Statistical considerations for the design and analysis of the ELISpot assay in HIV-1 vaccine trials. *J Immunol Methods* **288**, 19–34.
11. Moodie, Z., Huang, Y., Gu, L., Hural, J., and Self, S. G. (2006) Statistical positivity criteria for the analysis of ELISpot assay data in HIV-1 vaccine trials. *J Immunol Methods* **315**, 121–132.
12. R Development Core Team. (2009) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.
13. Anderson, T. W., and Darling, D. A. (1952) Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes. *Ann Math Stat* **23**, 193–212.