Statistical Treatment of Fluorescence *in Situ* Hybridization Validation Data to Generate Normal Reference Ranges Using Excel Functions

Allison L. Ciolino, Mary E. Tang, and Ron Bryant

From the Department of Pathology, University of Vermont College of Medicine/Fletcher Allen Health Care, Burlington, Vermont

Fluorescent *in situ* hybridization has become an essential tool for diagnosing and monitoring hematological disease. Testing for minimal residual disease requires precise and accurate normal cut-offs. There is no consensus in the field on the correct method of establishing a normal reference range. We discuss and compare several proposed statistical methods to calculate normal reference ranges, including Gaussian statistics, the β inverse function, and a binomial treatment of the data. We demonstrate that a binomial treatment of the data is an accurate and simple method to calculate a normal reference range. (J Mol Diagn 2009, 11:330–333; DOI: 10.2353/jmoldx.2009.080101)

Fluorescent *in situ* hybridization (FISH) was first used in the clinical lab as a qualitative test to determine the presence or absence of a chromosomal abnormality, such as a microdeletion or a simple translocation. With the expanding role of cytogenetics in hematology, FISH has become an essential tool to monitor disease. Although FISH is not as sensitive as quantitative reverse transcriptase PCR, it can detect variant translocations¹ and is reliable for monitoring unusual fusion variants.² The result is that FISH has become a quantitative test that requires a reference range to distinguish minimal residual disease from background noise. Several statistical methods have been proposed to calculate accurate and precise reference ranges.^{3–5} These methods include Gaussian statistics, the β inverse function, and a binomial treatment of the data. This paper will discuss the theory and application of statistics to establish reference ranges for FISH assays.

Determining the frequency of false positives for each probe and each signal pattern is the key to establishing good reference ranges. This is best accomplished by a validation study as described by the American College of Medical Genetics.⁶ Results from the validation study can be used to set cut-off limits for each probe pattern of each probe used.

Probability Theory and FISH

False-positive signals occur infrequently in FISH, but have many causes. They may be due to probe positioning in a three-dimensional nucleus that allows probe signals to overlap, diffuse probe signals in cells with less condensed chromatin, extraneous fluorescent signals, or random loss or gain of signals. In addition, the false positive rate is dependent on probe size and design, sample type, slide preparation technique, and observer bias.^{4,7} Strict attention to slide preparation and cell selection criteria can minimize but not eliminate the problem of false positive signals.

Dual-fusion probes with essentially no false positives are available for a large variety of abnormalities. With dual-fusion probes, any classic positive probe signal pattern, commonly 1 red, 1 green, 2 fused, is almost certainly a true positive cell. Accurate reference ranges are most critical for single fusion, enumerator, deletion, and break-apart probe strategies as well as nonclassical patterns from dual-fusion probes.

To establish a reference range (normal cut-off) a laboratory must estimate the false positive rate from a sample of clinically normal specimens and choose a cut-off that is higher than that encountered in “most” normal specimens. “Most” is usually defined as 95% or higher and is referred to as the confidence limit of the FISH assay.⁸ The level of confidence, however, should be chosen based on the clinical indications for performing the test. Higher confidence intervals (>95%) should be used when greater specificity is needed and the number of cells counts increased when greater sensitivity is required.⁹

Despite the many factors contributing to the false positive rate, the probability of a false positive can be expressed as a single number p . The p value of a probe pattern is calculated by dividing the total number of positive nuclei by the total number of nuclei evaluated (Equation 1) from a normal validation study.

$$p = \frac{\text{Total Positive Nuclei}}{20 \times 500} = \frac{\text{Total Positive Nuclei}}{10000} \quad (1)$$

Accepted for publication March 4, 2009.

Address reprint requests to Mary E. Tang, Department of Pathology and Laboratory Medicine, Fletcher Allen Health Care, 111 Colchester Avenue, Burlington, Vermont 05401. E-mail: mary.tang@vtmednet.org.

Table 1. Definition of Symbols Used

Symbol	Definition
p	Probability of a positive nucleus
n	Number of nuclei scored on a specimen
k	Number of positive nuclei in a specimen of sample size n
k_{max}	Largest number of positive nuclei found in a validation sample of size 500

A separate value of p should be determined by each laboratory for each probe, each pattern, and each sample type. It is very important to document all abnormal patterns found for each specimen type so that a separate statistical analysis can be performed for each pattern observed.

The number of positive signals found in a FISH assay is designated as k . The distribution of the number k is dependant on p (the probability of a positive signal) and n (number of nuclei scored in a FISH assay. See Table 1). Over many assays the number of positive signals takes on a binomial distribution (Equation 2).⁴

$$\Pr\{\text{positives} \leq k\} = \sum_{m=0}^k \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} \quad (2)$$

Using the value of p estimated by the normal validation study, Equation 2 can be inverted to find the normal cut-off for a FISH assay of n nuclei. $\Pr\{\text{positives} \leq k\}$, represents the confidence limit and the corresponding k is the normal cut-off (reference range) for the assay. For example, to find the 95% confidence limit it is necessary to find the value of k that produces a result of 0.95 for Equation 2. This process is best illustrated with a graph. Figure 1 illustrates the binomial distribution (Equation 2) for $p = 0.013$. In Figure 1A, the number of nuclei scored is $n = 100$ and in Figure 1B, $n = 500$. The normal cut-off value at the 95% confidence limit is the k -coordinate of the first point above the horizontal line drawn at $\Pr\{\text{positives} \leq k\} = 0.95$. Thus for n equals 100 or 500, $k_{0.95}$ is 3 or 11 respectively.

This graphical method of finding normal cut-off levels is a theoretically accurate but cumbersome method of finding the confidence limit. The mathematical solution is

Figure 1. Binomial distribution functions with 95% confidence limits indicated. **A:** $p = 0.013$, $n = 100$, $k_{0.95} = 3$; **B:** $p = 0.013$, $n = 500$, $k_{0.95} = 11$.

complex, but Microsoft Excel provides a function that inverts the binomial distribution function for a given n , p , and confidence limit. The function is named *CRITBINOM* (Equation 3).

$$k_{\Pr\{\text{positives} \leq k\}} = \text{CRITBINOM}(n, p, \Pr\{\text{positives} \leq k\}) \quad (3)$$

For example, assume a validation study obtains a value of $p = 0.013$. There are 200 nuclei scored in the test assay. The normal cut-off at the 95% confidence limit ($k_{0.95}$) is obtained in Microsoft Excel by entering:

$$= \text{CRITBINOM}(200, 0.013, 0.95)$$

The result displayed will be 5. Since 95% of normal specimens will be expected to have 5 or fewer positive nuclei, a specimen with more than 5 positive nuclei is interpreted as a positive result (eg, disease is present).

β Inverse Function

The β inverse function in Microsoft Excel has been proposed as an alternative function to calculate normal FISH cut-off values.^{4,10} The β inverse function inverts the β distribution, which is a very robust statistical distribution. Under a limited set of circumstances the β distribution is closely related to the binomial distribution.¹¹ The β distribution uses k , n and p to generate $\Pr\{\text{positives} > k\}$ rather than $\Pr\{\text{positives} \leq k\}$. Therefore, the binomial distribution equals $1 - \beta$ distribution that is equivalent to:

$$\Pr\{\text{positives} \leq k\} + \Pr\{\text{positives} > k\} = 1 \quad (4)$$

The first term in Equation 4 is the binomial distribution and the second term is the β distribution. The mathematical equation for the β distribution is complex but Microsoft Excel contains a function that does the calculation.

$$\Pr\{\text{positives} > k\} = \text{BETADIST}(p, k+1, n-k) \quad (5)$$

The binomial distribution (Equation 3) and the β distribution (Equation 5) sum give a result of 1 when k , n , and p are defined as above. The relationship between the binomial distribution and β distribution infers that the inverse β distribution can be used to calculate normal cut-off values of FISH. The proposed inversion uses the Microsoft Excel *BETAINV* function shown in Equation 6.^{4,10}

$$k_{\Pr\{\text{positives} \leq k\}} = \text{BETAINV}(\Pr\{\text{positives} \leq k\}, k_{max} + 1, n) \times n \quad (6)$$

k_{max} is the maximum number of positive nuclei found among the 20 validation samples. Equation 7 shows the correct use of the *BETAINV* function for the parameters

Table 2. Data Used in Validation Simulation

Random numbers simulating a validation of 20 samples of size $n = 500$ with $p = 0.013$	5, 5, 7, 5, 7, 6, 6, 7, 11, 5, 14, 9, 7, 4, 1, 4, 12, 7, 7, 2
Estimated p from simulated data*	0.014 (1.4%)

*Note that the estimated p (0.014) is not exactly equal to the p value (0.013) used to generate the random numbers.

Figure 2. Normal cut-off values at the 95% confidence limit calculated from the simulated validation data in Table 2 calculated by three methods. Binomial: 'o'; Gaussian (DeMoivre-Laplace): '*'; and β inverse: 'x'.

defined previously (Excel 2000 in a nutshell, accessed February 2009, <http://proquest.safaribooksonline.com/1565927141>).^{12,13}

$$p = \text{BETAINV}(\text{Pr}\{\text{positives} > k\}, k + 1, n - k) \quad (7)$$

Comparison of the two BETAINV functions reveals three errors in its use in Equation 6:

1. The "confidence limit" used in Equation 6 is $\text{Pr}\{\text{positives} \leq k\}$, but the value called for in the β inverse function is $\text{Pr}\{\text{positives} > k\} = 1 - \text{Pr}\{\text{positives} \leq k\}$ as shown in Equation 4.
2. Equation 6 uses k_{max} not k .
3. The last term should be $n - k$, and not n . This difference is negligible with the large n values commonly used for FISH.

For these reasons the BETAINV function should not be applied to FISH cut-off values.

Gaussian Treatment

Many laboratories have attempted to calculate normal FISH reference ranges using Gaussian statistics. The Gaussian treatment will work if done correctly but failure to appropriately do the mathematics has caused many laboratories to abandon the Gaussian statistics in FISH applications. The binomial treatment approach using Microsoft Excel function CRITBINOM (Equation 3) is the simplest and most precise method. The Gaussian approach is included for illustrative purposes.

Following a validation study there are 20 values ($k_1, \dots, k_{20}$) representing the number of positive nuclei found in each of the 20 specimens of 500 nuclei. Many laboratories simply find the mean and SD (sd) of these 20 values to calculate a normal cut-off value for a FISH assay. This approach is acceptable if two conditions are met. First the cut-off applies if the number of nuclei counted in the validation study equals the number of nuclei counted in the clinical assay. For a validation study of 20 specimens of 500 nuclei, the calculated cut-off, or its percentage, only applies to clinical assays with sample size $n = 500$. A separate validation study using 200 cell counts must be performed if the clinical assay uses 200 cell counts. The second condition is that a one-sided confidence limit should be calculated, not the more common two sided limit that has been used to establish other laboratory reference ranges. For a 95% confidence limit, the cut-off level is set 1.65 sd above the mean, not 1.96 sd above the mean. In addition, the cut-off value must be rounded upwards since k must be an integer.

FISH data can be treated in a Gaussian fashion for test sample sizes other than $n = 500$ when the DeMoivre-Laplace theorem is applied.¹¹ The DeMoivre-Laplace Theorem states that the Gaussian distribution approximates a binomial distribution when n is large, and $n \times p$ is significantly bigger than one. To use the Gaussian approach, the mean ($\bar{x}$) and SD (sd) of the Gaussian must be estimated as shown in Equations 8 and 9 below. This mean and SD with a z-score of 1.65 will encompass 95% of observations. Rounding the answer up to the next highest integer will give an accurate estimate of the binomial cut-off level (Equation 10).

$$\bar{x} = n \times p \quad (8)$$

$$sd = \sqrt{p(1 - p)n} \quad (9)$$

$$k_{0.95} = \bar{x} + 1.65 \, sd \quad (10)$$

Alternative one-sided z-scores are needed for other confidence limits.

FISH assays satisfy the DeMoivre-Laplace theorem because we are interested in confidence limits in the range of 95% to 99% where $k = n \times p$ is usually larger than one. Even if p is so small that $n \times p$ is less than one, the above method overestimates the confidence limits by at most one event.

Demonstration

A mock reference range determination is used to compare the binomial, β inverse, and Gaussian methods. A

Table 3. Normal Cut-Off Values and Cut-Off Percentages at the 95% Confidence Limit Calculated from the Simulated Validation Data in Table 2 by Three Methods

Method	$n = 100$		$n = 500$		$n = 1000$		$n = 1500$	
	$K_{0.95}$	Percent nuclei	$k_{0.95}$	Percent nuclei	$k_{0.95}$	Percent nuclei	$k_{0.95}$	Percent nuclei
Binomial	4	4.0%	12	2.4%	20	2.0%	29	1.9%
Gaussian	4	4.0%	12	2.4%	21	2.1%	29	1.9%
Beta inverse	19	19%	21	4.2%	22	2.2%	22	1.4%

probe validation is simulated by generating 20 random binomial numbers with $p = 0.013$ and $n = 500$ (Table 2). These random numbers are then used to calculate the 95% confidence cut-off limits ($k_{0.95}$) for test sample sizes of $n = 100, 200, 300, \dots, 1500$, by each of the methods discussed above. The binomial method uses *CRITBINOM* (Equation 3). The previously proposed β inverse method uses Equation 6. The Gaussian method uses Equation 10 and Equations 8 and 9 to calculate the mean and SD (Figure 2 and Table 3).

Results of the simulation experiment demonstrate that the Gaussian and binomial methods are essentially equivalent. The β inverse method calculates much higher cut-off values with sample sizes of less than 1000 nuclei. In addition, the β inverse method shows minimal change with increasing n .

Discussion

There is overall agreement in the cytogenetics community that FISH data takes on a binomial distribution.⁴ There are currently several methods being considered to calculate cut-off values after performing a validation study. The binomial distribution is theoretically correct and is the simplest method to use (Equation 3). This method, or any reliable method, is heavily reliant on a precise estimation of p to obtain accurate cut-off values. The American College of Medical Genetics standard validation with 500 nuclei from each of 20 normal specimens suffices for most circumstances.

The demonstration shows that the Gaussian method approximates the binomial method very well when large numbers of nuclei are scored. The Demoivre-Laplace Theorem indicates that the difference between the methods can be made statistically insignificant by scoring many nuclei and in practice this condition is easily met in FISH assays. While a properly done Gaussian analysis is an acceptable alternative to a binomial treatment, today's availability of software tools make the binomial treatment easier.

The β inverse method is published in the literature as a solution for calculating FISH reference ranges, but this demonstration indicates that the cut-off values that it produces do not correspond to the theoretical binomial results.

Accurate reference ranges for the classic patterns of dual-fusion probes pose a problem. It is very difficult to obtain a good estimate of p since positive signals are generally not found during the validation.^{3,14} If there are no positive signals in a 10,000 cell validation, then one can assume that p must be very small. Based on the binomial distribution (Equation 2) one can be 99% confident that if $p = 0.0005$ there should be one or more positive signals in 10,000 nuclei. Therefore, if no positive nuclei are seen in a validation study then certainly $p = 0.0005$ or less. A positive rate of $p = 0.0005$ results in cut-off values of $k_{0.99} = 1, 1, 2, 2$, and 2 for $n = 100, 200, 300, 400$, and 500 cells, respectively, at a 99% confidence limit. The estimate of $p = 0.0005$ in the above example is probably the worst case scenario. The majority of laboratories do not

report any positive cells in a 10,000-cell validation study of the dual fusion probes.^{3,14} Assuming that one in five laboratories were to find a positive signal during a validation study, this would correspond to $p = 0.00002$ for a dual-fusion probe. Using this revised value of p , any positive signal in 500 cells or less is statistically significant. Although the statistics argue for the significance of any positive nuclei, consideration of clinical implications is urged when interpreting a single positive nucleus.

In summary, a binomial treatment of the data is the theoretically most accurate method to determine cut-off values of FISH assays and can be easily done in Microsoft Excel using the *CRITBINOM* function. A properly done Gaussian treatment is as accurate, but is more complex to perform. The use of the β inverse function should be abandoned.

References

1. Ou J, Vergilio J, Bagg A: Molecular diagnosis and monitoring in the clinical management of patients with chronic myelogenous leukemia treated with tyrosine kinase inhibitors. *Am J Hematol* 2008, 83:296–302
2. Park IJ, AeLim Y, Lee WG, Park JS, Kim HC, Lee HJ, Cho SR: A case of chronic myelogenous leukemia with e8a2 fusion transcript. *Cancer Genet Cytogenet* 2008, 185:106–108
3. Dewald G, Stallard R, Alsaadi A, Arnold S, Blough R, Ceperich TM, Elejalde BR, Fink J, Higgins JV, Higgins RR, Hoeltge GA, Hsu WT, Johnson EB, Kronberger D, McCorquodale DJ, Meisner LF, Micale MA, Oseth L, Payne JS, Schwartz S, Sheldon S, Sophian A, Storto P, Tuinen PV, Wenger GD, Wiktor A, Willis LA, Yung JF, Zenger-Hain J: A multicenter investigation with D-FISH BCR/ABL1 probes. *Cancer Genet Cytogenet* 2000, 116:97–104
4. Wiktor AE, Van Dyke DL, Stupca PJ, Ketterling RP, Thorland EC, Shearer BM, Fink SR, Stockero KJ, Majorowicz JR, Dewald GW: Preclinical validation of fluorescence in situ hybridization assays for clinical practice. *Genet Med* 2006, 8:16–23
5. Thall P, Jacoby D, Zimmerman S: Estimating genomic category probabilities from fluorescent in situ hybridization counts with misclassification. *Appl Stat* 1996, 45:431–446
6. American College of Medical Genetics Standards and Guidelines for Clinical Genetics Laboratories, 2006 edition. Section E10.3.1
7. McClatchey, KD: Clinical laboratory medicine, second edition. Edited by McClatchey KD. Philadelphia, Lippincott, Williams, & Wilkins, 2002, pp. 658–685
8. American College of Medical Genetics Standards and Guidelines for Clinical Genetics Laboratories, 2006 edition. Section E10.3.2
9. Landstrom AP and Tefferi A: Fluorescent in situ hybridization in the diagnosis, prognosis, and treatment monitoring of chronic myeloid leukemia. *Leuk Lymph* 2006, 47:397–402
10. Wolff DJ, Bagg A, Cooley LD, Dewald GW, Hirsch BA, Jacky PB, Rao KW, Rao PN, the Association for Molecular Pathology Clinical Practice Committee and the American College of Medical Genetics Laboratory Quality Assurance Committee: Guidance for fluorescence in situ hybridization testing in hematological disorders. *J Mol Diagn* 2007, 9:134–143
11. Papoulis, A: Probability, random variables, and stochastic processes, third edition. Edited by Howell RL and Morriss JM. New York, McGraw-Hill, 1991
12. Evans M, Hastings N, Peacock B: Statistical distributions, third edition. Edited by Evans M. New York, John Wiley and Sons, Inc., 2000, pp. 34–42
13. Simon J: The statistical functions, Chapter 15 in Excel 2000 in a Nutshell [internet]. O'Reilly Media, Inc., 2000, 404–473
14. Dewald G, Wyatt W, Juneau A, Carlson R, Zinsmeister A, Jalal S, Spurbeck J, Silver R: Highly sensitive fluorescence in situ hybridization method to detect double BCR/ABL fusion and monitor response to therapy in chronic myeloid leukemia. *Blood* 1998, 91:3357–3365