YALE JOURNAL OF BIOLOGY AND MEDICINE 68 (1995), pp. 101-117.

Amniotic Fluid Polarization of Fluorescence and Lecithin/Sphingomyelin Ratio Decision Criteria Assessed

Larry H. Bernstein[a], Robert Stiller[b], Cheryl Menzies[c], Malcolm McKenzie[b,d] and Clark Rundell[a,d]

Departments of [a]Pathology and Laboratory Medicine, [b]Obstetrics and Gynecology and [c]Pediatrics, Bridgeport Hospital, Bridgeport, Connecticut; and [d]Maine Medical Center, Portland, Maine

(Received May 5, 1995; sent for revision October 18, 1995; accepted November 20, 1995)

A negative finding of amniotic fluid (AF) phosphatidyl glycerol (PG) does not eliminate the need for determining the lecithin/sphingomyelin ratio (LSR). We use a novel approach to classify fetal lung maturity (FLM) data, and to validate the fluorescence polarization (FP) surfactant assay (Abbott), which replaces the PG assay and reduces the frequency of repeat LSR. This method finds the values (decision points) of these tests that allow for classifying the data with least errors. These tests best identify the risk of respiratory distress syndrome (RDS) from fetal lung immaturity. We find the decision values for tests by exploring the data for information content and optimize their selection using group-based reference. We previously defined normal reference as the maximum entropy set with no information. The uncertainty resolved by information provided in the data allows formation of syndromic classes. This is greatest at the values for the variables (decision-points) associated with the greatest decrease in entropy. Decision-values found for PF, EGA, PG, LSR that classify amniotic fluids into the mature and not-mature classes are in agreement with the results of ROC analysis. We validate the replacement of PG by the PF method. We also find a level of FP below which LSR might be required to resolve uncertainty and above which the FP indicates maturity. We confirm the ability to evaluate fetal maturity methods using information analysis.

INTRODUCTION

This laboratory has triaged amniotic fluid (AF)[e] specimens for several years using the Amniostat-FLM procedure (Irvine Scientific, Santa Ana Calif.) for determining the presence of phosphatidyl glycerol (PG) before carrying out the thin-layer chromatographic analysis of lecithin/sphingomyelin ratio (LSR) (Helena Laboratories, Beaumont, Texas). This "triage" approach results in an unacceptable confirmation rate of PG negative, LSR positive specimens because of the poor sensitivity of the screening method. On the other hand, a PG value scored as 1+ or 2+ by this method correlates with a positive LSR so that an LSR may be done only in the absence of a PG value of at least 1+.

The desirability of replacing the PG method was motivated in large part because performing the LSR is labor intensive. Therefore, an automated screening method was needed with an accuracy comparable to the LSR with a low confirmation rate. This became feasible with the introduction of a fluorescence polarization assay (FP) (based on surfactant to albumin ratio) by Abbott Diagnostics, Inc.

[a]*To whom all correspondence should be addressed:* Larry H. Bernstein, M.D., Department of Pathology and Laboratory Medicine, Bridgeport Hospital, 267 Grant Street, Bridgeport, CT 06610. Tel: (203) 384-3208; Fax: (203) 384-3237; E-mail: lhbernsl@ix.netcom.com.

[e]*Abbreviations:* AF, aminiotic fluid; PG, phosphatidyl glycerol; LSR, lecithin/sphingomyelin ratio; FP, fluorescence polarization; RDS, respiratory distress syndrome; EGA, estimated gestational age; FLM, fetal lung maturity; ROC, receiver-operator characteristic curve; IUGR, intrauterine growth retardation; BH, Bridgeport Hospital; MMC, Maine Medical Center.

We evaluate these assays of fetal lung maturity (FLM) using test data to form a classification related both to FLM and to risk of RDS. The tests of FLM are combined to form syndromic classes. The classification is determined by the information content in the data from which decision values for the tests are determined. The results obtained by ROC analysis using RDS as the training variable and the entropic method are the same.

PRELIMINARY STUDY

An initial study of 158 AF analyses was done at Bridgeport Hospital. The PG and LSR were both done because of negative or indeterminate PG determinations by Amniostat FLM for 158 analyses at Bridgeport Hospital. The enzymatic method for PG analysis that measures to an LSR equivalent of 2 was not published at the time of this study [1]. This contributed to a bias in immaturity of the Amniostat method with an expected positive result at an LSR well above 2. The results were compared with estimated gestational age, which was available for all tests. In many of these patients maturation of fetal lung surfactant was induced prior to amniocentesis. The data were analyzed by the method of group-based reference [2]. We compared the entropy in the data after randomizing the variables to remove correlation with the entropy of the ordered data. We obtained decision-points or cut-points (decision values) from the difference in entropies for the data EGA, PG, LSR, separating the population into syndromic classes. By this measure, the information was greatest at the value of each variable used for separation for which the difference between the data entropy and the entropy of the randomized data is greatest. These were: EGA, 33.8 weeks; PG, 0; LSR, 2.5.

The PG was scaled at 0, 0.5 and 1 for values of not-mature, indeterminate and mature, respectively. A score of 0.5 was equivalent to a 0. The EGA and LSR, respectively, were scaled at less than 33.8 weeks (not-mature) and 33.8 weeks or more (mature), and less than 2 (not-mature), and 2 or more (mature). The number of groups that can be defined by three variables with two possible outcomes is eight. The classes with the least uncertainty (EGA, PG, LSR) were: 000, 101, 011, 100, 111 (a 110 pattern is excluded).

The result, an LSR decision point at 2.5, even though RDS is likely at an LSR well below 2, follows from the paired comparisons using PG. Table 1 shows the means and standard deviations for 102 not-mature and 55 mature results. The high negativity of the PG is shown at nearly equivalent EGAs. This contributes to significant uncertainty about FLM without the LSR.

POLARIZATION OF FLUORESCENCE (FP) ASSAY

We then used the estimated gestational age, PG and LSR to evaluate a new test, the fluorescence polarization (FP) [3-5] assay with the expectation that the new method could reduce the number of LSR tests done because of false-negative PG tests. The population defined by the analyses of AFs has data which define at least two different populations based on FLM. The risk of RDS of the newborn decreases as lecithin increases and LSR approaches and exceeds two with fetal lung surfactant maturation.

The new method can be evaluated by comparison with previous methods. A cross-tabulation is better than regression analysis for this comparison because normal distribution cannot be assumed for the data or for the variances. The linear regression model has an inherent problem in requiring the errors to be normally distributed. An approach based on outcomes evaluation (RDS) requires a study population of substantial size because of a low incidence of RDS in the not-mature population. We therefore combined the study of two institutions in Bridgeport, Connecticut and Portland, Maine to obtain a study sample of sufficient size with 496 patients (147 from Bridgeport and 349 from Portland) for outcomes determination. Additional data were provided by Dr. Alan Wu from his study at Hermann Hospital in Houston, Texas, bringing the study sample to 571.

We have suggested that if there is sufficient information in the data, it is possible to examine the data and determine the optimum decision values without referencing each value to a prior classification [2]. The determination of sensitivity and specificity requires a prior assignment of RDS and the confidence intervals under the ROC are problematic at a low rate of RDS. The entropic method is not dependent on either a prior assignment of RDS, or on distributional characteristics of the data. The decision-values are entirely derived from the informational characteristics of the data.

We analyze the entropy of the data set to find the decision-values for each variable that best classifies the data. The data which distinguish the mature from not-mature population in terms of EGA, PG, LSR and FP have a decreased entropy associated with uncertainty relieved at the critical decision-values. These values are where the difference is greatest between the entropy of the ordered and randomized (eliminating association) data, allowing these tests to separate the populations. This critical-value is that which allows best separation into mature and not-mature classes using EGA, PG, LSR and FP. The entropic method has been successfully applied to evaluation of tests used to diagnose other conditions [2,6-10]. The effect of the relationship between a surfactant test and lung maturity is an expected significantly different rate of RDS in the two groups.

STUDY SAMPLE

An initial retrospective study used 158 patients and three years of data collected by the Bridgeport Hospital (BH) laboratory on LSR, PG and EGA without regard to outcome. The second study at Bridgeport Hospital was prospective and involved collection of the same data, fluorescence polarization data (FP) on the Abbott FLM test and review of outcomes for 147 patients with a very low RDS occurrence. An additional 349 patients were usable from a similar study carried out at the Maine Medical Center (MMC) in Portland, Maine and 75 patients were from Dr. Alan Wu in Houston, Texas. In the BH study the PG test was done by Amniostat, but in the MMC study the PG values were read from the thin-layer chromatographic plates. Outcomes were established, such as RDS, transient tachypnea or apnea, intrauterine growth retardation, diabetes, premature rupture of membranes, etc., as much as possible in both studies except when the patient was lost to follow-up.

The data sets were combined without any assumptions of normal reference. This was not inappropriate because the Abbott method is automated and has no bias across laboratories. In the case of the PG there is a difference in sensitivity between the precipitin and chromatographic assay. The analysis is carried out independently of laboratory definitions of reference range.

METHODS

Entropy

We have reported [2] on how to explore data for information to discover critical decision-points at which the entropy decreases, suggesting a separation between distinct populations. In this case we can expect to discover at least two distinct populations that are described by the data: FP, PG, LSR and EGA. There are more than two when one considers intrauterine growth retardation (IUGR) as an outcome, and the tests are not sufficient to distinguish all of the populations. A finding where each of the four variables is positive (+,+,+,+) is always mature, and a finding where each value is negative (-,-,-,-) is always not-mature. The importance of intermediate combinations of these have to be discovered. The LSR is scaled to 0 for less than two, 2 for > 2.5 and 1 for equivocal. The PG is either 0 or 1. The optimum value for FP is determined from the entropy in the data using EGA, PG and LSR.

Our method [2] extends the work by Rypka [11] in the domain of medical clustering, and is directly traceable to the relationship between entropy and missing information in both physics (e.g., Boltzmann, 1894; Szilard, 1925; von Neumann, 1932) and in communication theory elaborated by Shannon [12]. In information theory the quantitative measure of information in a message source is entropy, H:

$$H = P_1\log_2(1/P_1) + P_2\log_2(1/P_2) + \ldots P_n\log_2(1/P_n),$$

where H is a function of the probabilities (P) of the choices in a set of messages having different amounts of information. We have described [2, 6] the treatment of test combinations as a diagnostic message set, the variable combinations of which are a message transmission. If there is no information in the data-set supplied, the data-set is at maximum entropy and has no information. If there is information in the data-set supplied, the information supplied is that which is necessary to relieve uncertainty in the message.

The analysis of information in the data supplied is made possible by examining the entropy of the entire set of messages transmitted (which continue to be supplied). As the message set becomes complete or the number of messages becomes sufficient, the amount of information in the data and the amount of information needed to relieve uncertainty becomes more clear. We stress that information *not* provided in the data may be as important as that which *is* discovered in this process.

We measure the entropy of the data at maximum uncertainty by randomizing the data to determine its entropy without associative relationships between the variables. We determine the uncertainty in the ordered data by measuring its entropy. The difference between the maximum entropy of the unordered data and the lower entropy of the ordered data is the effective information. This difference is greatest at the decision-point for each variable in the message set for which the most information is supplied. We refer to this as the *optimum decision-value*. The effectiveness of the message, or information provided, is dependent on the information provided by each test and by the selection of tests. It follows that any combination of n tests may be inaccurate if it provides insufficient information to resolve uncertainty in the message. One can then assume that a sufficient *number* and *quality* of tests can minimize uncertainty in any diagnostic problem.

The data were treated in a matrix format, with four columns corresponding to EGA, PG, LSR and FP and the patients in rows. Entropy calculations were made from a binary data matrix produced by applying decision values as described [2]. The decision-points determined the reference-limits for the study population. The relative frequencies of the binary patterns were used to calculate the entropies. The decision-values for scaling of LSR and PG were derived from a prior study without FP of 158 AFs using a three-letter word message set. Eight binary patterns were possible. The present study with FP uses a 93 (of 121 or 470) line four-letter message (including FP) with 16 possible patterns. All possible patterns occur with equal frequency when there is no information in the message set, all messages being equally probable. If the average entropy, H, in the data is 1 (for true-false outcomes) then probabilities of eight equally likely messages to yield one unit of information is 0.125, and the sum of the entropies of these events takes on its maximum value. In the case of binary expression of three variable combinations, the maximum entropy is the number of attributes, or 3, and the frequency distribution of a flat distribution for each binary expression is 0.375. If the variables that are used to form the binary pattern classes are randomly and independently associated, then the outcomes are equally likely for any pattern. This results in a decision-value for each variable at its median at maximum entropy with equal frequencies of the binary patterns at maximum entropy.

The finding of maximum uncertainty in the data set indicates that there is no effective information in the data. The finding of much greater frequency of the 0000 and 1111

pattern is highly significant for four tests. This distribution is different from the flat maximum entropy partition. It indicates there is information in the message; the basis for a discovery process. This results in a measured entropy that is less than four bits. The difference between the measured entropy and the expected maximum entropy is the effective information present.

We have stressed that information can only be found in multivariate relations. There is not enough information in a single test to resolve uncertainty when two choices are possible, and even when a single test is used as a benchmark in a diagnostic process, it is not used as a random independent test. In other words, the use of the test is not from an assumption of two equally likely outcomes. In our case of four tests, we explore the ability to predict fetal maturity with outcomes that are not independent for FP, EGA, PG and LSR. We define the effective information in the data as the difference between the maximum uncertainty of uncorrelated (randomized) data and the actual uncertainty in the data obtained. Examination of this relationship reveals decision-points that decrease the uncertainty in the relations and at the same time produces binary words that classify the data. In this sense, we find a definition of "limits-of-normal" that are not arbitrary, that are defined by endogenous information in the data itself. In examining the entropy of each variable, the entropy of each variable is calculated in increments using the variable under study vs. the medians of the other variables as fixed values. The entropy is calculated from the probability distribution of the binary patterns produced for each increment. This entropy is subtracted from the entropy derived from destroying the association between the variables.

We have referred to the effective information in the data relation as uncertainty "relieved" or the missing information at the different decision-limits. We also find that there may be more than one decision-limit for a set of data which is related to different classification problems. When such a situation exists, it becomes important to have sufficient tests to reduce mis-classification errors.

Statistics

Statistical analyses were carried out on a 486DX IBM-compatible PC on Statgraphics software (Manugistics, Rockville, Maryland) and included: normal probability plots and distribution analyses, cross-tabulation, one-way analysis of variance, Kruskal-Wallis analysis, regression analyses, and the Kolmogorov-Smirnov two-sample test. Analysis of variance was used when variances of continuous predictors (LSR and FP) were examined between categorical groups, i.e., RDS, no RDS, and intrauterine growth retardation (IUGR). Regression was examined when both the predictor and response were continuous variables.

In the case of examining outcomes, a regression model can be used having more than one predictor. This is problematic if the response is not linear or if the errors are not normally distributed. In each of these cases, however, a model can be constructed only if a prior, supervisory classification is defined. We have selected a method which is independent of the distribution of the errors and which minimizes the entropy.

A receiver-operator characteristic curve (ROC) determines sensitivity and specificity at different cut-points. The ROC depends on the selection of an *a priori* outcome. One might also plot the diagnostic efficiency, or errors, against the reference outcome as the value of the predictor is varied. An inherent problem is the low level of RDS in the study population and the range of values for which the predictor is positive with RDS. Reliable confidence intervals under the ROC curve may be hard to come by. There is no inherent advantage to deriving the cut-points using the ROC curve. First, the approach requires a prior reference to a supervisory classification. This is not the case with entropic method which uses the information in the data to create a self-learning matrix. Second, the outcome

Table 1. Statistics for fetal maturity.

Test	Mature (mean + SD)	Not mature (mean + SD)
EGA	34.6 + 1.1	34.3 + 1.2
PG	0.2 + 0.3	0.05 + 0.15
LSR	3.0 + 0.44	1.8 + 0.43
Patterns (no 110)	111, 011, 101, 001	000, 100, 010, 001
Size	56	102

Table 2. Simplified ROC analyses of FLM data.

Test	Sensitivity	1-Specificity
FP		
70 mg/g	0.967	0.485
50 mg/g	0.70	0.208
35 mg/g	0.567	0.70
LS		
2.5	0.967	0.362
2.0	0.70	0.146
1.5	0.50	0.042
PG	0.909	0.547

Table 3. Truth table of classified data.

Class	Pattern	Frequency	Likelihood	Uncertainty
0	0000	15	0.30	3.333
	1000	2	0.04	25.0
		17	**0.34**	2.941
1	0010	3	0.06	16.667
	0011	5	0.10	10.0
	0012	1	0.02	50.0
	1010	2	0.04	25.0
	1021	1	0.02	50.0
		12	**0.24**	4.167
2	1122	10	0.20	5.0
	1121	2	0.04	25.0
	1022	3	0.06	16.667
	1021	1	0.02	50.0
	0122	3	0.06	5.0
	0021	1	0.02	50.0
	0022	1	0.02	50.0
		21	**0.42**	2.381

of interest, RDS, is dependent on the association between EGA and surfactant level (measured by LSR and FP). Therefore, examining the data for information establishes the same cut-points that are optimum for RDS determined by the ROC curve.

RESULTS

ROC Analyses

The FP, LSR, PG and EGA were analyzed by cross-tabulation and a simplified ROC analysis was carried out using the table generated. The FP and LSR are continuous variables. The FP and LSR were converted to scores prior to cross-tabulation based on age range and presence or absence of RDS (IUGR was excluded). FP converted to PFSCORE is: less than 35 mg/g = 0; 35-49 mg/g = 1; 50-70 mg/g = 2; 71-85 mg/g = 3; greater than 85 mg/g = 4. LSR converted to a LSCORE is: less than 1.5 = 0; 1.5-1.9 = 1; 2.0-2.4 = 2; 2.5-2.9 = 3; equal to or greater than 3 = 4.

Table 2 is the result of the simplified ROC analysis from the cross-tabulated data.

Entropy decision-values

The initial study of 158 amniocenteses resulted in the classification with EGA, PG and LSR (Table 1). The LSR between 2.0-2.5 (indeterminate) maturity allows for an intermediate class arising out of the paired comparisons with PG (PG-, LSR+). The entropy determined decision-values that classify for risk of RDS were: FP, 50-80; EGA, 35 weeks; PG, 1; LSR, 2.5. Figures 1a and 1b are plots of effective information (bits) in the L/S ratio and the new FLM test, FP, showing decision points at 2.5 and 68, respectively (data from-first 93 cases at BH).

Binary pattern classes

The number of groups that can be defined by four variables with two possible outcomes is 16. These are compressed into three categories defined by the binary classes of the four variables.

Table 3 is a truth-table showing the test patterns in the three groups using a ternary assignment for LSR and for FP as follows: LSR: not-mature, less than 2 = 0; probable, 2.0-2.5 = 1; mature, greater than 2.5 = 2; FP: not-mature, less than 50 mg/g = 0; probable, 50-80 mg/g = 1; mature, greater than 80 mg/g = 2.

The group assignments are determined by the binary patterns. The Fisherian likelihood for each pattern is 1/n, where N is the sum of n's, and the uncertainty is represented as the proportion n/N. Table 3 constructs these representations of the pattern classes.

Using the decision-values for assigning mature *vs.* not-mature, it is possible to calculate the likelihoods for surfactant activity with FLM (specificity) and or its absence with immaturity (sensitivity), and the likelihood ratios for each variable. The low level of RDS (30) in a population sample of 290 known outcomes makes the calculation of incremental risk unattractive. In this study the negative LSR or FP surfactant assay should be associated with RDS of the newborn and reflected in the likelihood ratio, which decreases significantly at FP values above 50 mg/g. The odds-ratio was calculated for RDS with the LSR and FP, respectively, scaled using -2, -1, 0, 1 and 2 as dummy scores for the intervals less than 1.5; 1.5-1.9; 2.0-2.4; 2.5-2.9; and greater than 3.0; for LSR, and less than 35 mg/g; 35-49 mg/g; 50-70 mg/g; 71-85 mg/g; and greater than 85 mg/g for FP. The Goldminer universal regression program was used (J. Magidson, Statistical Innovations, Inc., Belmont, Mass.) [13,14] to obtain odds-ratios of 24.66 ($p = 1.2 \times 10^{-15}$) for LSR and 14.831 ($p = 4.0 \times 10^{-13}$) for FP.

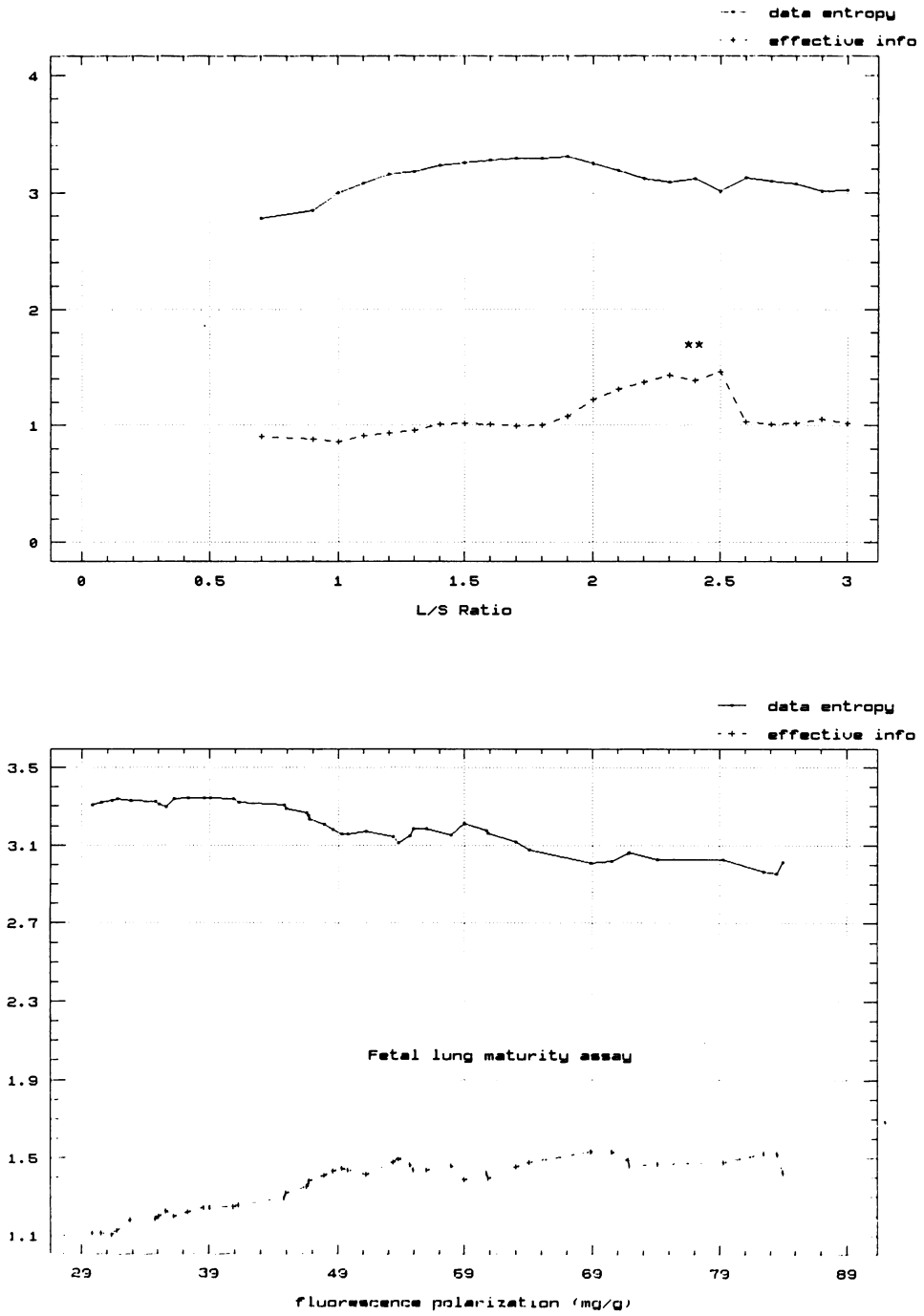

Figure 1. a) Effective information (bits) (ordinate), the entropy of the ordered data subtracted from the entropy of the randomized data is plotted against the lecithin/sphingomyelin ratio (LSR) (abscissa); b) Effective information (bits) (ordinate), the entropy of the ordered data subtracted from the entropy of the randomized data versus fluorescence polarization (mg/g albumin) results (abscissa). The data entropy (solid line) is plotted above the effective information (broken line). The optimum decision value is the value of the variable at the effective information maximum.

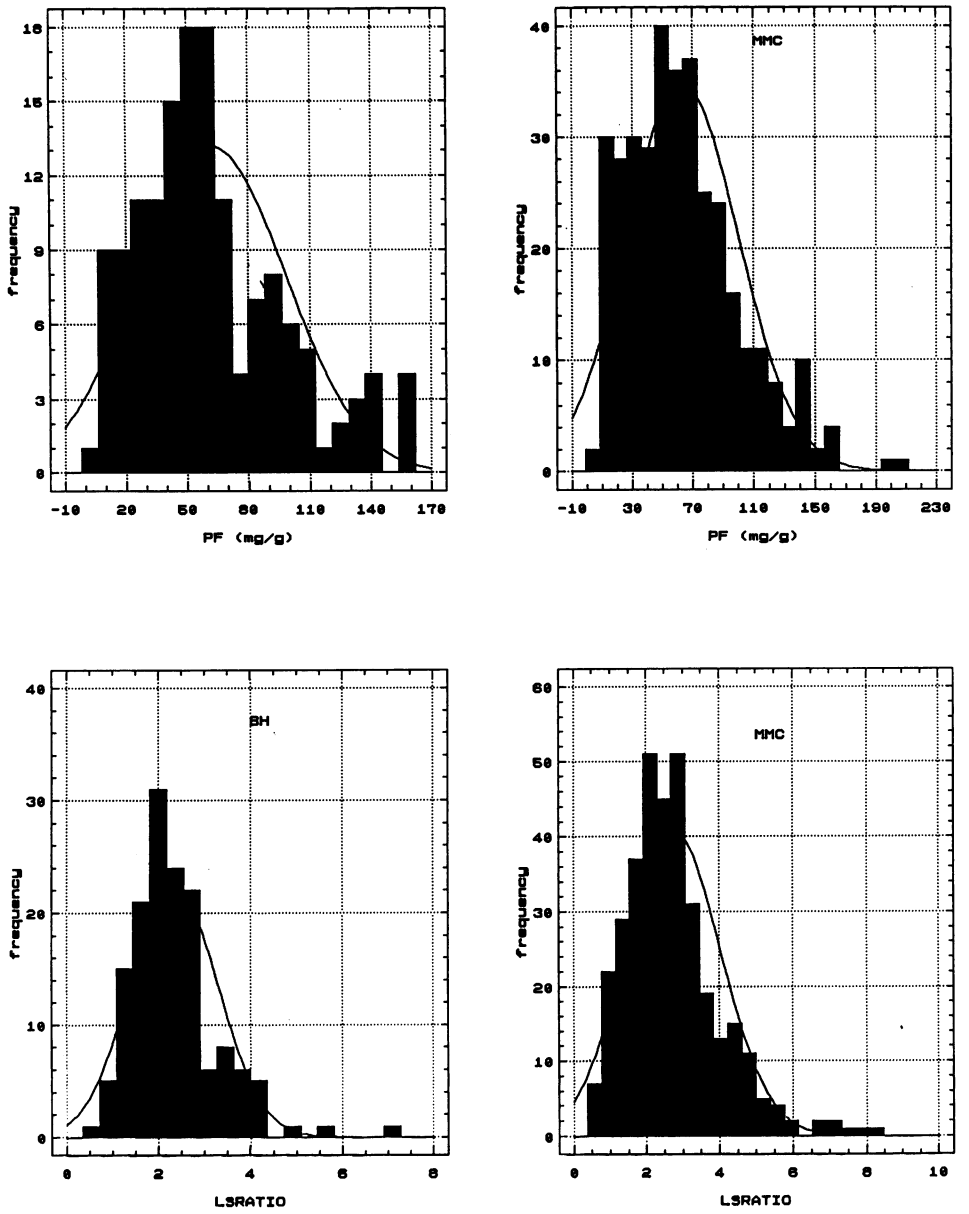

Figure 2. Histograms show distribution of: FP (PF) for Bridgeport Hospital sample (a); FP (PF) for Maine Medical Center (b); LSR (LSRATIO) for Bridgeport Hospital (c); LSR (LSRATIO) for Maine Medical Center (d).

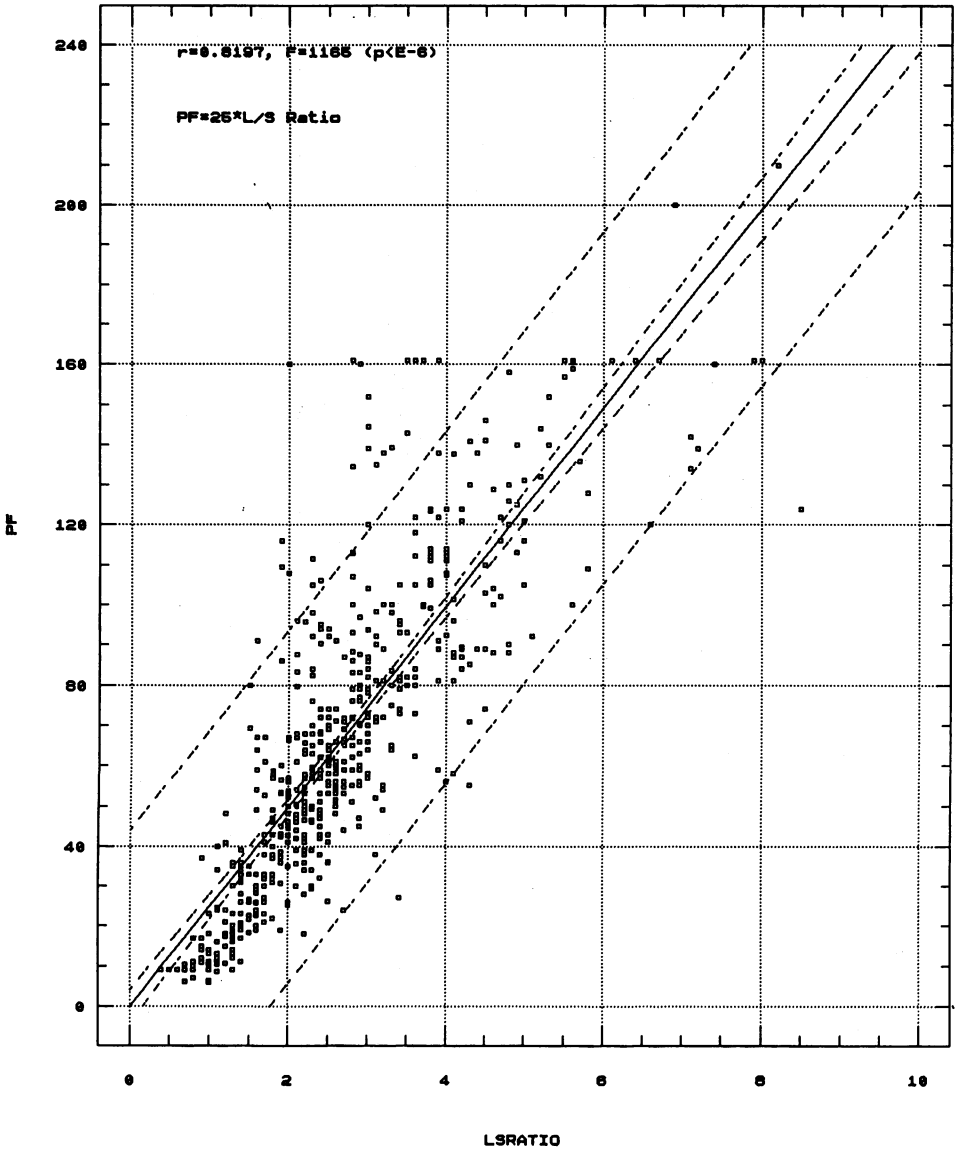

Figure 3. Linear relationship between TDX FP (PF) surfactant/albumin assay (ordinate) and LSR (LSRATIO) (abscissa).

The effect of the PG determinations on the matrix, which contributes to the maximum at LSR of 2.5 and FP of 70 mg/g, reflecting the use of the test in the diabetic population, supports the intermediate range suggested by the manufacturer.

Data evaluation

Figure 2 is a histogram comparing the FP and LSR distributions for BH and MMC. The LSR, not the FP, is normally distributed. Kruskal-Wallis analysis for FP (9×10^{-9}) and LSR (1.7×10^{-5}) by data is significant.

FP and LSR analyses

The relationship between the FP and LSR (Figure 3) as determined by the combined data is described by the equation:

$$FP = 24.909 \times (LSR) - 0.1237$$
$$(r^2 = 0.8197, F(570) = 1165, p < 10^{-6})$$

Maturity class is predicted from LSR and FP by the equation:

$$\text{maturity class} = -0.60654 + 0.014099 \times FP + 0.33762 \times (LSR),$$
$$(r^2 = 0.79530, F(570) = 179.68, p < 10^{-5})$$

Figures 4a and 4b are scatterplots of FP and LSR, respectively, in those cases with and those without respiratory distress syndrome. FP values are below 40 mg/g with RDS and above 50 mg/g without RDS, and LSR are below 2 with RDS and above 2 without RDS. Figure 5 is a comparison of the FP and LSR by component scatterplots for: (a) RDS only; (b) No RDS; (c) PG positive; (d) PG negative. With rare exception, RDS is included in the range FP less than 60 mg/g and LSR less than 2.5. In addition, PG positivity is usually associated with FP exceeding 40 mg/g, and LSR exceeding two. The PG-negative range, however, is very dense at FP above 40 mg/g and LSR exceeding 2.

Analysis of classified data

The classification of the BH and MMC data according to outcomes is evaluated in Figure 6. The classes, in order, are: no RDS = 0; RDS = 1; transient tachypnea or apnea = 2; intrauterine growth retardation (IUGR) = 3. The analysis of variance and box plots of FP and LSR shows separation of RDS and no-RDS at 50 mg/g and 2, respectively, based on the upper and lower confidence limits for the groups. IUGR (3) and unclassified (9) have low FP and LSR, but IUGR is not as low as RDS. The Kruskal-Wallis analysis of groups by rank is significant for FP and LSR ($t = 114.4$; $t = 111.3$; $p < 10^{-6}$). The difference between RDS and no-RDS is significant for FP and LSR by the Kolmogorov-Smirnov nonparametric two sample test ($p < 3.9 \times 10^{-8}$; $p < 9.3 \times 10^{-8}$).

DISCUSSION

This study supports the finding that EGA and LSR are covariates for assessing FLM. The greatest difficulty is the inaccurate dating of fetal gestational ages between 32 and 37 weeks. We found fewer RDS in the BH AF classified as not-mature, but a comparable proportion in the combined BH and MMC sample to that of Steinfeld et al. [5], whose FP, not-mature, at less than 50 mg/g was associated with RDS present in 18 of 350 AF, and who found no cases of RDS at LSR less than 2 or FP greater than 50 mg/g. There was no correlation between any predictor and outcome of RDS or transient tachypnea. However, it is notable that transient tachypnea occurs in the same range as no RDS.

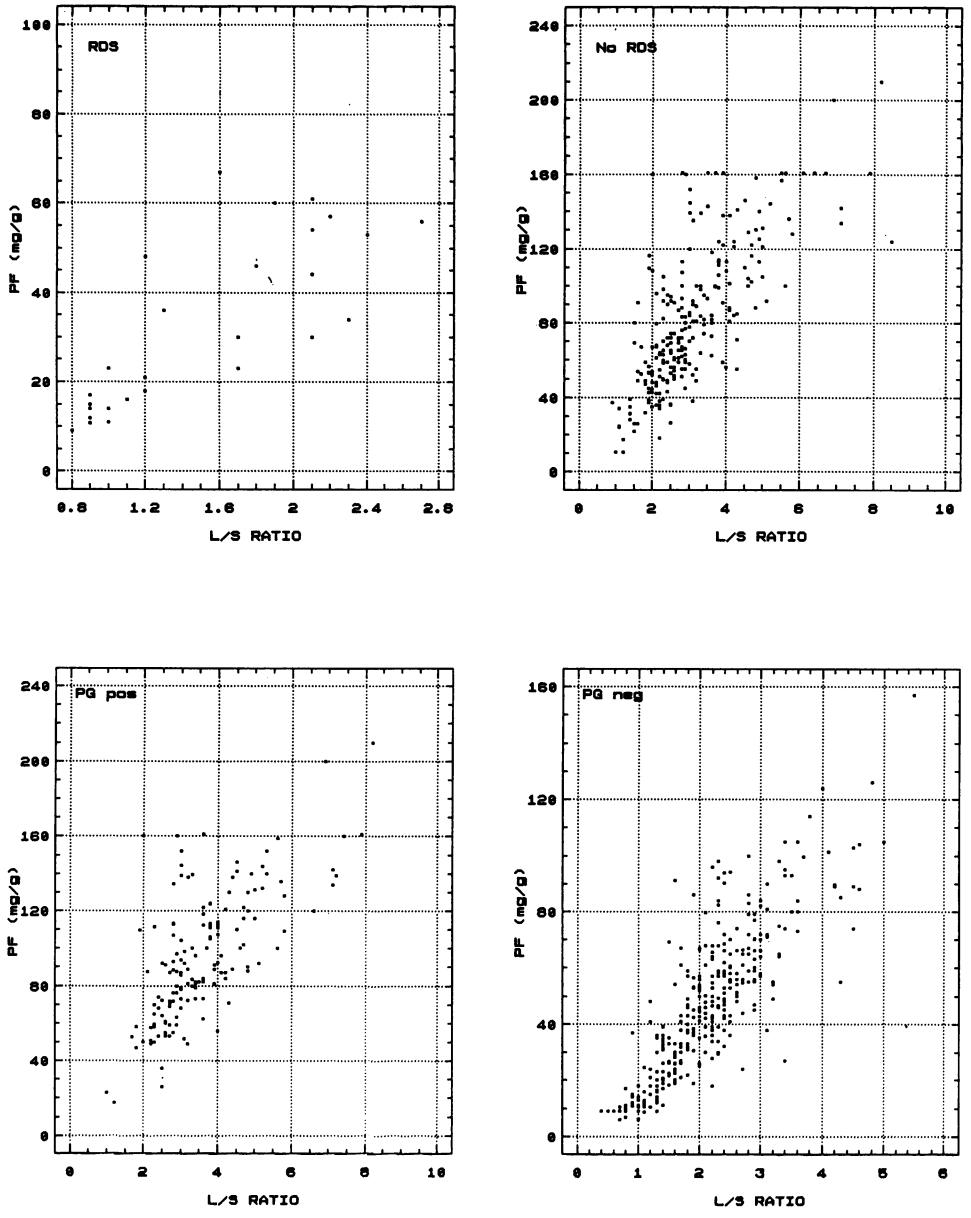

Figure 4. Scatterplots of FP (a) and LSR (b) by class (abscissa). No RDS, 0; RDS; 1; TTN, 2; IUGR, 3.

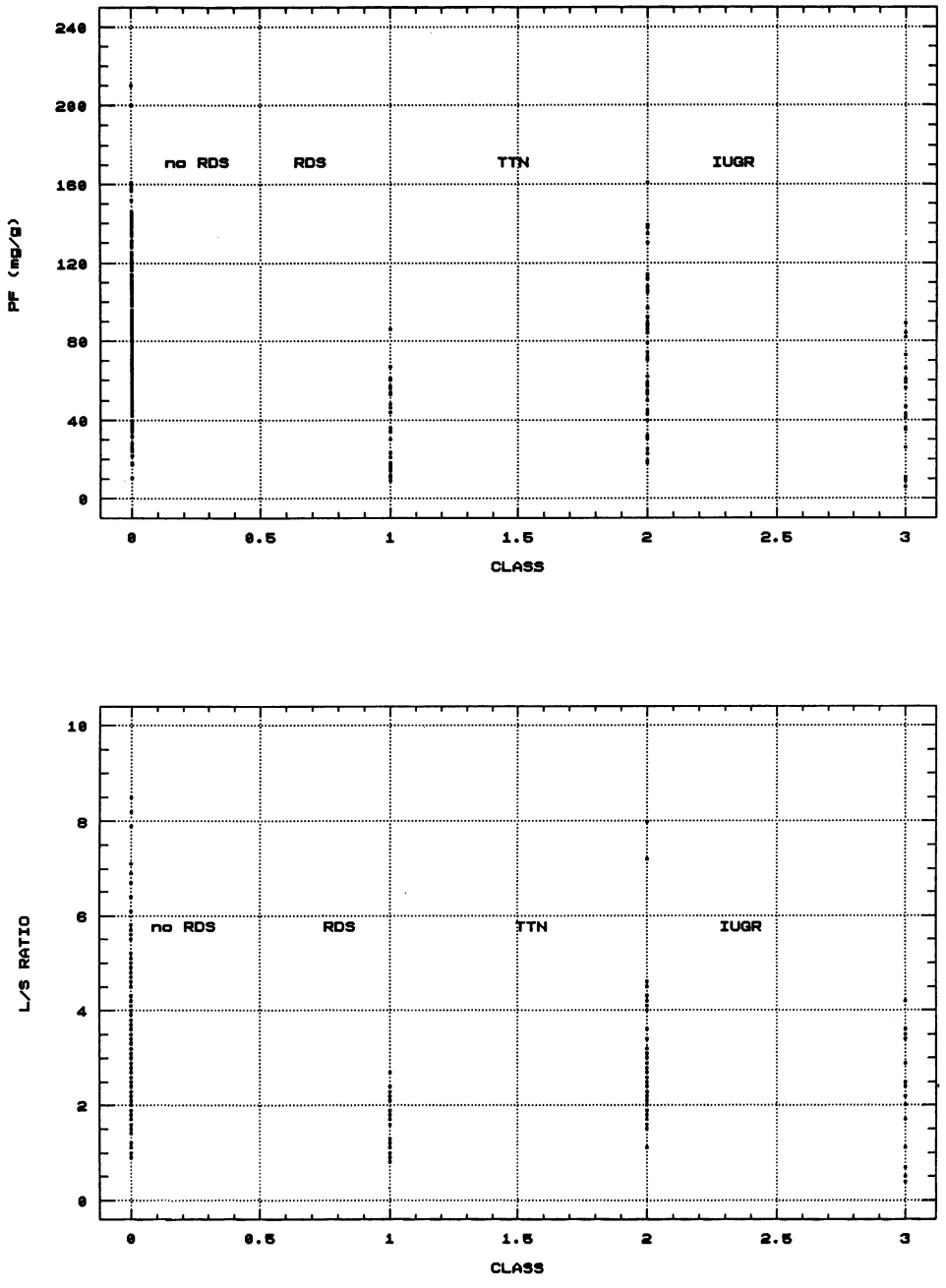

Figure 5. Plots of FP (ordinate) against LSR (abscissa). (a) RDS; (b) no RDS; (c) PG positive; (d) PG negative.

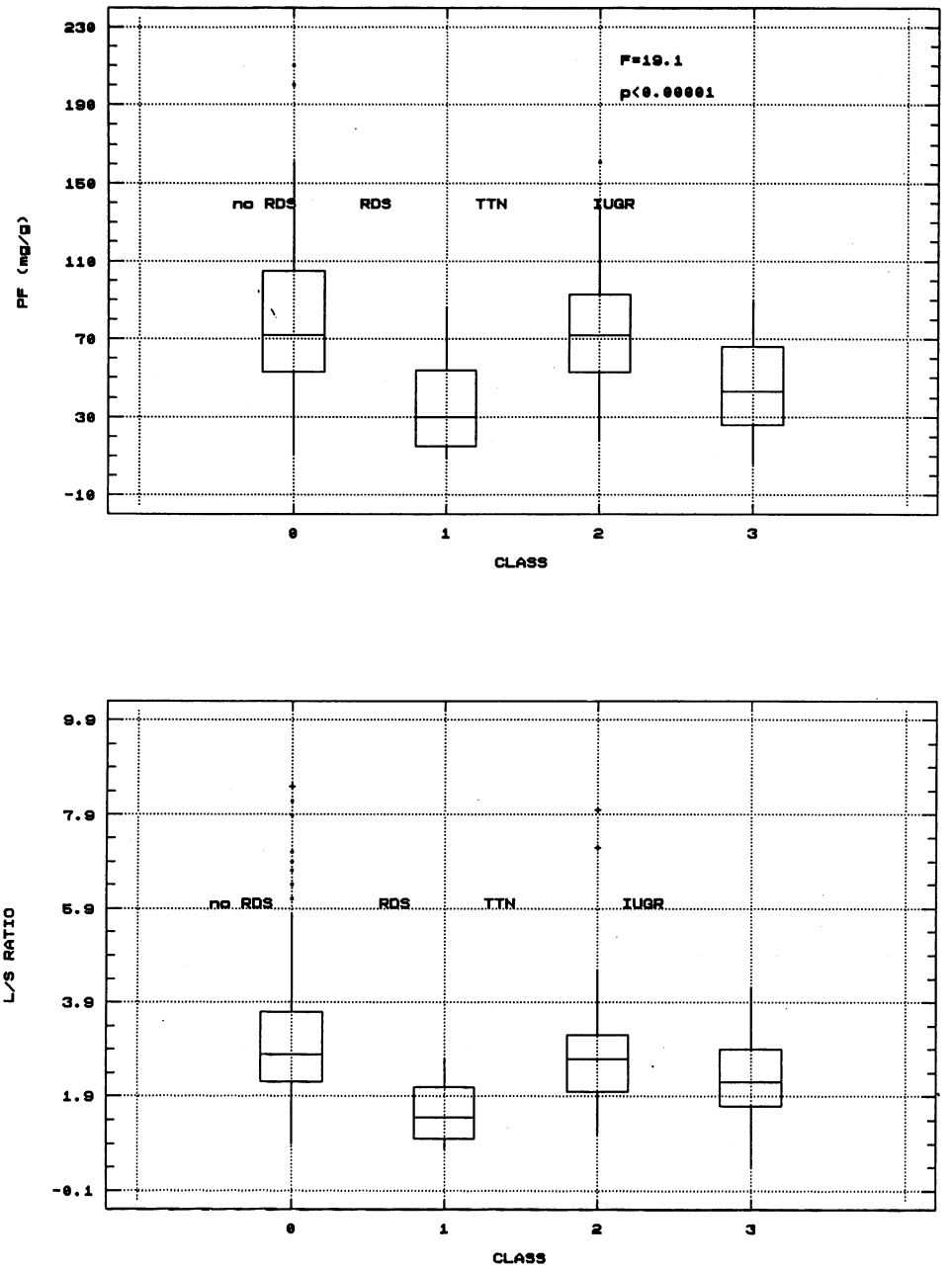

Figure 6. Means ± standard error of means (SEM) by class: (a) FP; (b) LSR. No RDS, 0; RDS, 1; TTN, 2; IUGR, 3.

This analysis only considers IUGR with a low FP and LSR, not drug use or other disease factors which may accelerate or retard maturation. While an LSR exceeding 1.9 is usually considered mature, this study establishes the uncertainty at LSR ratio below 2.5 and FP below 70 mg/g. A PG that is not 1+ in the Amniostat method is the same as a negative and precludes the elimination of LSR. This is of interest in the group that has a negative PG and an LSR between 2.0 and 2.5.

This study illustrates the methodological problems posed by evaluation of tests for fetal lung maturity. It is customary to validate the LSR ratio using a ROC curve which compares the diagnostic efficiency or sensitivity and specificity of the test against the LSR cutoff at assigned ratios [15]. How is the efficiency measured? In this case, all of the cases of RDS (not transient tachypnea) of the newborn would be used for assigning disease class, but there is significant uncertainty about the negative tests vs. outcome.

The ROC curve is a model that requires the comparison of each test outcome with a reference classification defining mature and not-mature. In this case, RDS is the defining factor. This means that very large studies would have to be carried out for achieving validity because of the low frequency of the expected outcome. An important aspect of this issue is the fact that an LSR result that is not-mature does not invariably result in RDS. It usually does not. Therefore, other models for assessing maturity are needed.

Methods using fuzzy set logic where uncertainty is scaled into prediction of the outcome [16] are closely related to the model used here. The problem posed by this study illustrates the difference between measuring efficacy and effectiveness. The demonstration of efficacy requires controlled studies and invariably suffers from lack of adequate scope of the study population, thus introducing bias. The consequence is that tests or treatments have poorer performance when introduced into broader usage. This is expected when the validation study has insufficient information to define effects that occur outside of the study. Studies of effectiveness have to take into account utilization in situations in which confounding factors are present. This necessitates use of multivariate methods to determine the importance and effects of the confounding factors. In the case of LSR, the test was introduced based on the correlation between a biochemical measurement and development of fetal lung surfactant activity, which is measureable in amniotic fluid. Soon after its introduction it became apparent that both contamination issues (red cell phosphatidyl choline) and disease factors (diabetes) were important in its interpretation. In addition, the method limits were defined in relationship to RDS; mature and not-mature were both compared with respect to a low frequency event at a reference limit that is predetermined. PG is a correlated test unaffected by specimen contamination and it is useful for eliminating the need for LSR when the result is positive, but it has a high false-negative rate, making it costly in the situation with a significant prevalence rate of complicated pregnancies and premature deliveries. On the other hand, there is a high concordance of positive PG with positive LSR. Unlike the finding of a positive PG or LSR, a negative PG or LSR may be considered insufficient information for predicting RDS. EGA is necessary information that is only accurate in a three week range. The model using information-induction introduced by Rudolph et al. [2] allows an examination of the endogenous information in the data without reliance on some predetermined definition of outcome, such as RDS.

There are other models that are being used for effectiveness research. The most common methods are Cox regression and proportional hazards [17, 18]. These methods are used for comparing alternative treatments in oncology trials to determine the effect on survival when prognostic factors and therapy independently improve outcome. In this case, the selection of distinct patient subsets has posed problems requiring additional methods, such as recursive partitioning and amalgamation [19, 20]. This method successively partitions the population into two subgroups until no available variable produces a survival

difference and then joins subgroups of patients who do not differ with regard to survival potential. The recursive partitioning method has also been used to predict myocardial infarction [21], but while its goal is to predict patients with similar survival outcome, it can be used to identify important factors without their being specified prior to analysis. Clearly, the proportional hazards model might be adapted for studies such as this, but the unfavorable outcome is not non-survival. An interesting use of the proportional hazards model is in the finding of a ten-fold increase in relative risk of death associated with low cholesterol level in nursing home residents [22]. The model also identified a co-variable effect of hemoglobin and albumin, and a co-morbidity effect of decubitus ulcers.

The opportunity to look at data relational-linkages is especially attractive for an information-based model. This is the important question with respect to the effectiveness of alternative methods of defining fetal maturity: PG, LSR, FP, EGA. The recent study by Tanasijevic et al. [23] is of some interest because of the use of a combination of FLM surfactant/albumin ratio with obstetric estimates of gestational age. They used FLM with EGA in a prediction rule determined by logistic regression.

The ability to identify risk using binary pattern recognition has been demonstrated while confirming the value of the FP assay.

REFERENCES

1. Jones, G.W. and Ashwood, E.R. Enzymatic measurement of phosphatidylglycerol in amniotic fluid. Clin. Chem. 40:518-525, 1994.
2. Rudolph, R.A., Bernstein, L.H., and Babb, J. Information-induction for the diagnosis of myocardial infarction. Clin. Chem. 34:2031-2038, 1988.
3. Foerder, C.A., Tait, J.F., Franklin, R.W., and Ashwood, E.R. Improved fluorescence polarization assay for use in evaluating fetal lung maturity. II. Analytical evaluation and comparison with the lecithin/sphingomyelin ratio. Clin. Chem. 32:255-259, 1986.
4. Ashwood, E.R., Tait, J.F., Foerder, C.A., Franklin, R.W., and Benedetti, T.J. Improved fluorescence polarization assay for use in evaluating fetal lung maturity. III. Retrospective clinical evaluation and comparison with the lecithin/sphingomyelin ratio. Clin. Chem. 32:260-264, 1986.
5. Steinfeld, J.D., Samuels, P., Bulley, M.A., Cohen, A.W., Goodman, D.B.P., and Senior, M.B. The utility of the TDX test in the assessment of fetal lung maturity. Obstet. Gynecol. 79:460-464, 1992.
6. Rudolph, R.A., Pinto, M.M., and Bernstein, L.H. Measuring decision values for CEA and CA125 in effusions. Lab. Med. 21:574-578, 1990.
7. Bernstein, L.H., Pinto, M.M., Viner, N., Zuckerman, H., and Rudolph, R.A. Decision values for serum concentrations of prostate-specific antigen (PSA) determined. Clin. Chem. 36:515-518, 1990.
8. Bernstein, L.H., Leukhardt-Fairfield, C.J., Pleban, W., and Rudolph, R.A. Usefulness of data on albumin and prealbumin concentrations in determining effectiveness of nutritional support. Clin. Chem. 35:271-274, 1989.
9. Spiekerman, A.M., Rudolph, R.A., and Bernstein, L.H. Determination of malnutrition in hospitalized patients using group-based reference. Arch. Pathol. Lab. Med. 117:184-186, 1993.
10. Bernstein, L.H. Relationship of nutritional markers to length of hospital stay. Nutrition (suppl)10:205-209, 1994.
11. Rypka, E.W. Clustering by truth table comprehension for optimal interpretive reporting. Lab. Med. 13:20-41, 1982.
12. Shannon, C.E. and Weaver, W. The Mathematical Theory of Communication. Urbana, IL: University of Illinois Press, 1949.
13. Magidson, J. Chi-squared analysis of a scalable dependent variable. Proceedings of the 1992 Annual Meeting of the American Statistical Association. 1992, Section on Statistical Education, 242-247.
14. Haberman, S.J. and Magidson, J. Universal regression models. Submitted to J. Am. Stat. Assoc. August 1995 (Statistical Innovations, Inc.).
15. Zweig, M.H. ROC analysis: assessment of diagnostic accuracy in the laboratory. Lab. Med. 22:708-711, 1991.

16. Klir, G.J. and Folger, T.A. Fuzzy Sets, Uncertainty, and Information. Prentice Hall, Englewood Cliffs, NJ, 1988.
17. Kaplan, E.I. and Meier, P. Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. 53:457-481, 1958.
18. Cox, D.R. Regression models and life tables. J. Roy. Stat. Soc. B 34:187-202, 1972.
19. Sagman, U., Maki, E., Evans, W.K., Warr, D., Shepherd, F.A., Sculier, J.P., Haddad, R., Payne, D., Pringle, J.F., Yeoh, J.L, Crampi, A., DeBoer, G., McKinney, S., Ginsberg, R., and Feld, R. Small-Cell Carcinoma of the lung: Derivation of a prognostic staging system. J. Clin. Oncol. 9:1639-1649, 1991.
20. Albain, K.S., Crowley, J.J., LeBlanc, M., and Livingston, R.B. Survival determinants in extensive-stage non-small-cell lung cancer: The Southwest Oncology Group experience. J. Clin. Oncol. 9:1618-1626, 1991.
21. Goldman, L., Cook, E.F., Brand, D.A., Lee, T.H., Rouan, G.W., Weisberg, M.C., Acampora, D., Stasiulewicz, C., Walshon, J., Terranova, G., Gottlieb, L., Kohernick, M., Goldstein-Wayne, B., Copen, D., Balery, K., Brandt, A.A., Jones, D., Mellors, J., and Jakenkowski, R. A computer protocol to predict myocardial infarction in emergency department patients with chest pain. N. Engl. J. Med. 318:797-803, 1988.
22. Verdery, R.B. and Goldberg, A.P. Hypocholesterolemia as a predictor of death: A prospective study of 224 nursing home residents. J. Gerontol. 46:M84-90, 1991.
23. Tanasijevic, M.J., Wybenga, D.R., Richarson, D., Greene, M.F., Lopez, R., and Winkelman, J.W. A prediction model for fetal lung maturity employing gestational age and test results. Am. J. Clin. Pathol. 102:788-793, 1994.