Journal of Carcinogenesis

Special Issue - Colorectal Carcinogenesis and Prevention

Review Article

Statistical methods for assays with limits of detection: Serum bile acid as a differentiator between patients with normal colons, adenomas, and colorectal cancer

Bonnie LaFleur*, Wooin Lee[1], Dean Billheimer[2,4], Craig Lockhart[3], Junmei Liu[4], Nipun Merchant[5]

Department of Epidemiology and Biostatistics, [2]Department of Agricultural and Biosystems Engineering, [4]Statistical Consulting Laboratory, BIO5, University of Arizona, Tucson, AZ, [1]Department of Pharmaceutical Sciences, University of Kentucky, Lexington KY, [3]Department of Medicine, Washington University, St. Louis, MO, [5]Department of Surgery, Vanderbilt University Medical Center, Nashville, TN, USA

E-mail: blafleur@email.arizona.edu
*Corresponding author

Published: 16 April, 2011
Journal of Carcinogenesis 2011, 10:12
This article is available from: http://www.carcinogenesis.com/content/10/1/12

Received: 28 February, 2011
Accepted: 01 March, 2011

Abstract

In analytic chemistry a detection limit (DL) is the lowest measurable amount of an analyte that can be distinguished from a blank; many biomedical measurement technologies exhibit this property. From a statistical perspective, these data present inferential challenges because instead of precise measures, one only has information that the value is somewhere between 0 and the DL (below detection limit, BDL). Substitution of BDL values, with 0 or the DL can lead to biased parameter estimates and a loss of statistical power. Statistical methods that make adjustments when dealing with these types of data, often called left-censored data, are available in many commercial statistical packages. Despite this availability, the use of these methods is still not widespread in biomedical literature. We have reviewed the statistical approaches of dealing with BDL values, and used simulations to examine the performance of the commonly used substitution methods and the most widely available statistical methods. We have illustrated these methods using a study undertaken at the Vanderbilt-Ingram Cancer Center, to examine the serum bile acid levels in patients with colorectal cancer and adenoma. We have found that the modern methods for BDL values identify disease-related differences that are often missed, with statistically naive approaches.

Keywords: Bile acids, colorectal cancer, detection limits, statistical methods

INTRODUCTION

Methods that account for left-censoring have been utilized by statisticians for many years, particularly for measurement of environmental contaminates (e.g., arsenic) in water.[1,2] In biomedicine, vaccine studies and HIV[3] have driven methodological development. Despite the fact that many of the methods are well-established, new applications and extensions remain a statistician's research topic.[4] Recent research in biomarker discovery studies has provided a new impetus for development of methods.[5,6] This article, and the illustrative example, focus on situations where outcomes are measured with BDL values. Statistical treatment is

Access this article online

Quick Response Code:

Website: www.carcinogenesis.com

DOI: 10.4103/1477-3163.79681

similar whether the outcome or the exposure / predictor is measured with DL,[7] however, fewer software options are available when the exposure is measured with values below the DLs.

Reviews of statistical methods for estimation and analysis of data with DLs are abundant, particularly in the literature of environmental sciences.[8-10] Many of the reviews are concerned with estimation and provide estimates of bias (difference between the estimated mean under a given method and the true mean) and coverage probabilities of 95% confidence intervals. We focus our discussion on a subset of methods that closely link estimation and hypothesis tests. They include substitution, nonparametric, and maximum likelihood methods.

SUBSTITUTION METHODS

Naive methods for handling data of this nature include deletion or simple substitution with the DL, DL/2, or even zero. After substitution, the 'usual' statistical methods are used to obtain parameter estimates and test hypotheses, for example, linear regression and two-sample t-test. Although simple to implement, it has been demonstrated that parameter estimates resulting from these methods are biased;[8-10] furthermore, there is little mathematical justification for their use, as pointed out by Slymen.[2] Substitution of a BDL value with zero, biases the mean tolower than the true mean and substitution with the DL biases the mean to higher than the true mean. Substitution with half the DL has been shown to have reasonable properties under specific conditions, but in general, it has been shown to be biased;[8-10] furthermore, Slymen[2] accurately elaborates that this substitution assumes that the distribution of the BDL is uniform over the range between zero and the DL, and not a standard assumption. A very informative and thorough review can be found in Helsel.[11] It is noteworthy that the author, Helsel, wrote a similar review 15 years ago,[12] and has remarked that within the environmental sciences, substitution methods are still the most commonly employed. Helsel further points out in the earlier work that as methods are easily available and given the bias that is induced with estimation by methods of substitution, their continued use is simply 'not defensible'.[12]

Nonparametric methods

Regression on the order statistics (ROS) methods are based on a simple linear regression model using ordered detected values and distributional (normal) quantiles to estimate the concentration of the censored (BDL) values;[13] they are also called 'probability plot methods'.[12] This is actually a semi-parametric method, as the quantiles are obtained by assuming an underlying parametric distribution for the uncensored values. Often the data are transformed before the order statistics are obtained. For example, the NADA package in R defaults to the log transformation, but other transformations could be of interest. Once the censored values are estimated, then the usual methods of parameter estimation and hypothesis testing are employed. Although this method has been shown to be fairly robust to the proportion of BDL data, and in the presence of moderate skewness, the bias becomes problematic in the presense of highly skewed data.[1] Alternative nonparametric methods rely on Kaplan-Meier estimates and log-rank or Peto-Prentice test statistics.[14] The estimates are obtained using the usual empirical cumulative distribution function for censored data,[15] and the test statistics are the same, as is typically used with right- or interval-censored data. As with all nonparametric procedures, these estimates are relatively unbiased, although the test statistics are not as powerful as their parametric counterparts when the underlying distribution is known or can be approximated.

Maximum likelihood methods

Maximum likelihood methods are based on approximating the distribution of the observed data with a parametric distribution. Parameters are estimated to create a distribution under which the observations are 'most likely'. Hypotheses regarding plausible values of the parameters may be tested with regard to the observed data. The most common statistical model for biologic data posits a log-normal (or normal) probability distribution. Often the logrithmic transformation is preferred to help stabilize the variance and reduce large discrepancies in magnitude between smallest and largest values. The censoring mechanism can be easily incorporated into these parametric models, and most commercial statistical packages have some facility to estimate and test parameters from these models.[11,16] Additional methods that fall under this broad category of maximum likelihood include mixture models, and some methods of multiple imputation.

Mixture models are useful when the proportion of the values of BDL are higher than would be expected under other parametric models, or more importantly, when there are 'true' zero's in addition to the BDL values due to limitations in the assay or measurement technique. One of the most common applications for mixture distributions is measurement of serum antibody levels in infectious disease. In this situation, one may never have been exposed to the antigen or have lost acquired immunity, and having no (or zero) antibodies is a possibility. The basic approach is to model a mixture of two components: One component estimates the probability of having no measurable amount, and the second describes the distribution for values above the DL. The first

component is generally modeled by a binomial distribution (sometimes referred to as the Bernoulli component) and the second component generally uses the log-normal or normal distribution.[3,4] Estimation is performed independently for each component and hypothesis testing is done by combining the likelihoods for both components. One of the best introductions to the approach of combining likelihoods can be found in Lachenbruch;[17] although, censoring is not included in this description. Finally, another method that is based on likelihood is multiple imputation. Multiple imputation (MI) methods have become prevalent for use with data having missing values.[18] Imputation is used to 'fill in' the missing (or BDL) values, based on patterns from the non-missing data. Although most commonly used commercial statistical packages have imputation algorithms, it has been shown that these might be inappropriate when 'missingness' is due to values BDL.[16] Choosing the appropriate imputation model is the key in the utility of these methods, and if the wrong imputation model is used, the results are not much better than the other approaches described. Also, higher levels of censoring will result in greater bias in parameter estimates.

SIMULATION

To demonstrate the properties of competing statistical methods, we simulated data from the lognormal distribution and from the exponential distribution. For these simulations we were interested in estimating mean parameters in two groups and testing the differences between the parameters. We examined the 'bias' (difference between estimated parameter and true value from the underlying simulation) and power of the test to detect a difference. The underlying parameters for the simulation were chosen to provide approximately 80% power to samples sizes that were relevant to our bile acid example.

As the lognormal distribution is the most commonly used distribution with biomedical data, the exponential distribution provides a useful check for comparison of methods with model misspecification. The censored fraction ranges from 40 to 60%, a fraction commonly observed with bile acid measurements. Simulation parameters for the lognormal are $\mu 1 = 0.50$ and $\mu 2 = 1.0$ for the location parameter, and $\sigma 1 = \sigma 2 = 1.0$ for the scale parameter. Subscripts 1 and 2 denote the sample group identifier. Simulation parameters for the exponential distribution are means of 3 and 5 for groups 1 and 2, respectively, and a common scale parameter of 1.0. All simulations have been repeated 1000 times. We have evaluated bias and power under three types of substitution, zero, half-DL, and the DL, as well as estimated them under KM (using Peto-Prentice test statistics), ROS, and censored regression assuming lognormal distribution (labeled ML). Table 1 shows the estimated parameters, standard errors, bias (difference between true and estimated test statistics) and power under the three levels of censoring for substitutions from the lognormal distribution. Simulations from the exponential distribution can be found in Table 2.

These simulations are similar to those in previous studies.[1,8-10] When data come from a lognormal distribution, estimation (bias) and power (performance of inferential procedure) are both maximized. For simulations from a lognormal distribution the substitution methods have demonstrated decent power to detect pairwise differences. However, there is significant bias even at the lower levels of censoring. Both ROS and ML methods have performed better than the substitution methods in terms of power and bias, although ML methods are less biased, regardless of the amount of censoring. When underlying simulated data are generated from the exponential distribution, the results are mixed. Surprisingly, substitution with half-DL has performed as well as ML and ROS methods, with regard to bias and with little loss of power. We caution the reader, that the positive results for this method are likely due to characteristics of our underlying model (e.g., the exponential distribution is highly right-skewed, which limits the range between zero and the DL limit). We do not feel that this behavior will be consistent for arbitraty deviations from normality (or lognormality). Shumway[1] also shows that ML and ROS methods are similar, except in the case of highly-skewed distributions, where ML has outperformed the ROS methods. This is similar to what we have demonstrated with simulations for the exponential distribution. We have observed reduced bias in the ML method compared to ROS. Additionally, Shumway[1] has observed that the ML methods are biased in small samples, with increased censoring. This is also consistent with our results. With 60% censoring and relatively small sample sizes, power under the ROS methodology is greater, and the difference in the bias is quite small.

These simulations demonstrate that ML methods exhibit better performance in terms of reduced bias and increased power when the underlying distribution is well-approximated by a parametric model. When it is not clear that data can be modeled using a parametric distribution, either because of sample size or distributional assumptions, then ROS (or perhaps KM depending on the sample) is a sensible alternative. Although not ideal, the parametric ML method using the lognormal distribution performs well, with respect to power, even when the underlying distribution is not lognormal; although the ROS method is clearly more powerful. Even as methods for estimation using ROS are implemented intostandard software, test statistics based on this estimation are not as well developed. A limitation to the KM method is that models are limited to one categorical

Table 1: Results from lognormal simulation

Censoring rate (percent)	Mean 1	SD 1	Mean 2	SD 2	Bias	Power
No censoring (complete data)	**0.497**	**0.129**	**0.990**	**0.121**	**0.012**	**0.798**
40 (group1: 49, group2: 30)						
Zero	0.650	0.088	1.037	0.099	0.187	0.741
1/2 DL	0.770	0.098	1.111	0.100	0.381	0.737
DL	0.889	0.114	1.185	0.106	0.575	0.705
KM	0.900	0.115	1.193	0.107	0.593	0.726
ROS	0.462	0.181	0.979	0.136	0.059	0.775
ML	0.503	0.146	0.992	0.126	0.011	0.747
50 (group 1 : 60, group 2 : 40)						
Zero	0.586	0.086	0.976	0.086	0.110	0.709
1/2 DL	0.808	0.093	1.126	0.093	0.435	0.703
DL	1.030	0.112	1.276	0.112	0.807	0.666
KM	1.048	0.114	1.289	0.114	0.837	0.690
ROS	0.439	0.225	0.971	0.225	0.089	0.742
ML	0.500	0.160	0.990	0.160	0.011	0.715
60 (group 1 : 70, group 2 : 51)						
Zero	0.500	0.084	0.885	0.084	0.115	0.672
1/2 DL	0.848	0.089	1.139	0.089	0.488	0.675
DL	1.197	0.111	1.393	0.111	1.090	0.614
KM	1.225	0.114	1.412	0.114	1.137	0.653
ROS	0.402	0.295	0.960	0.295	0.138	0.719
ML	0.492	0.183	0.986	0.183	0.022	0.677

Zero (substitution by zero), 1/2 DL (substitution with half-DL), DL (substitution with DL), KM (Kaplan-Meier), ROS (regression on order statistics), and ML (maximum likelihood, lognormal distribution with left-censoring)

Table 2: Results from exponential simulation

Censoring rate (percent)	Mean 1	SD 1	Mean 2	SD 2	Bias	Power
No censoring (complete data)	**2.997**	**0.384**	**5.000**	**0.605**	**0.003**	**0.819**
40 (group 1 : 48, group 2: 32)						
Zero	2.578	0.379	4.709	0.607	0.713	0.816
1/2 DL	3.045	0.385	5.022	0.605	0.067	0.820
DL	3.512	0.406	5.336	0.608	0.848	0.804
KM	3.533	0.408	5.360	0.610	0.894	0.709
ROS	3.141	0.406	5.137	0.614	0.278	0.830
ML	3.066	0.388	5.049	0.606	0.115	0.823
50 (group 1: 59, group 2 : 41)						
Zero	2.321	0.374	4.500	0.607	1.179	0.802
1/2 DL	3.107	0.389	5.049	0.606	0.156	0.804
DL	3.893	0.438	5.598	0.619	1.491	0.792
KM	3.940	0.447	5.642	0.627	1.582	0.710
ROS	3.207	0.431	5.202	0.623	0.409	0.822
ML	3.148	0.398	5.108	0.609	0.257	0.809
60 (group1: 70, group 2 : 51)						
Zero	1.992	0.370	4.200	0.600	1.808	0.793
1/2 DL	3.225	0.395	5.100	0.608	0.324	0.793
DL	4.457	0.483	6.000	0.644	2.457	0.772
KM	4.557	0.498	6.076	0.654	2.633	0.727
ROS	3.291	0.477	5.283	0.642	0.574	0.814
ML	3.299	0.412	5.219	0.616	0.518	0.800

Zero (substitution by zero), 1/2 DL (substitution with half-DL), DL (substitution with DL), KM (Kaplan-Meier), ROS (regression on order statistics), and ML (maximum likelihood, lognormal distribution with left-censoring)

Table 3: Sample demographics

		Group		
		Normal N = 50	**Adenoma N = 31**	**Cancer N = 96**
Age[a]		63.9 ± 8.8	63.2 ± 12.2	61.8 ± 12.1
Sex	Male	58% (29)	48.3% (15)	62.5% (60)
Cancer	Stage 1			29.2% (28)
	Stage 2			18.8% (18)
	Stage 3			24.0% (23)
	Stage 4			28.1% (27)
Adenoma[b]	Villous		14.2% (4)	
	Tubulovillous		57.1% (16)	
	Tubular		17.9% (5)	
	Serrated		10.7% (3)	
	Large[c]		44.0% (11)	

a Mean ± standard deviation, b N = 28, one FAP patient, one patient with polypoid hyperplasia, and one patient with a missing pathology report removed for these calculations. c N = 27, one patient missing size information.

Table 4: Percent values below DL for each group

Bile Acid	Normal N = 50 (%)	Group Adenoma N = 31 (%)	Cancer N = 96 (%)
CA	46	42	44
CDCA	62	55	47
DCA	18	19	23
LCA	46	29	44
UDCA	70	71	59

covariate. Overall, even with relatively small sample sizes and 60% censoring, ML methods perform adequately, even with a misspecified distribution. Our suggestion would be to perform a sensitivity analysis when distributional assumptions are questionable, to assess the performance of the estimation procedures; when there is a large difference between methods a trade-off will need to be made between power and bias. In all cases, we recommend ceasing the use of substitution methods for data with DL.

EXAMPLE: BILE ACIDS

Background

Nutritional factors such as a diet high in fat have been associated with colon cancer development.[19] A number of studies have suggested that elevated levels of secondary bile acids (BAs) may be in part responsible for this association.[20] In response to fat intake, BAs are secreted to aid in the absorption of cholesterol and fat-soluble vitamins.[21] BAs exist in several forms. The primary BAs such as cholic acid (CA) and chenodeoxycholic acid (CDCA) are derived from cholesterol in the liver and are secreted in the bile, mainly as glycine or taurine conjugates. In the colon, secondary bile acids, deoxycholic acid (DCA) and lithocholic acid (LCA), are formed from conjugated forms of CA and CDCA, respectively, through deconjugation and dehydroxylation, by the anaerobic bacterial flora. Subsequently, the tertiary bile acids, ursodeoxycholic acid (UDCA) and sulpho-LCA (SLCA) are formed through epimerization of CDCA or sulfation of LCA. BAs are absorbed from the intestines into the portal blood system via 'enterohepatic circulation,' resulting in circulating concentrations of 0.01 – 0.02 mmol/l, predominantly bound to albumin. In the circulation, CA and CDCA each comprise of about 30 – 40%, DCA about 20 – 30%, and LCA < 5% of the total amount of detectable BAs. DCA and LCA, the main fecal bile acids, are suspected to be the forms of bile acids that are implicated in colorectal carcinogenesis.

Several epidemiological studies indicate an association between BAs and colorectal cancer. Elevated fecal levels of the secondary BAs[22] with higher fecal LCA / DCA ratios[23] are seen in patients with colorectal cancer, as compared to healthy volunteers. Additionally, male patients with colon adenomas have higher serum levels of DCA than control patients with negative colonoscopies.[24] These studies, however, have measured levels of only a few select BAs. Using a sensitive HPLC-mass spectrometry assay, we simultaneously determined the serum levels of all major BAs in healthy volunteers and patients with colorectal ademonas and cancers of all clinical stages. A recent review article lays out a strong Darwinian argument for the accumulation of damage from high levels of bile acids as a strong risk factor for GI cancer.[25] Given this evidence, we wish to closely examine the difference in serum bile acid concentrations in three groups of patients, those with colorectal cancer, adenomatous polyps, and normal colorectal mucosa.

MATERIALS AND METHODS

Patient samples

Serum samples from patients with normal colonoscopies,

Table 5: Means and *P*-values for overall tests and tests for trend

Bile Acid[a]	Means in nM (95% CI)			*P*-value (Adjusted)		*P*-value (Zero)	
	Normal N = 50	Adenoma N = 31	Cancer N = 96	Overall	Trend	Overall	Trend
CA	9.06 (5.22, 15.73)	14.87 (10.03, 22.05)	14.87 (10.03, 22.05)	0.3302	0.1477	0.4531	0.2094
CDCA	9.56 (5.05, 18.11)	19.37 (9.17, 40.93)	24.39 (15.95, 37.37)	0.0410	0.0137	0.0708	0.0227
DCA	35.80 (21.95, 58.39)	83.46 (45.03, 154.68)	63.91 (44.85, 91.07)	0.0683	0.0952	0.2673	0.2914
LCA	9.11 (5.05, 13.17)	12.74 (9.64, 15.83)	13.26 (10.53, 15.97)	0.0271	0.0519	0.0700	0.1439
UDCA	3.86 (1.81, 8.25)	3.67 (1.44, 9.40)	7.84 (4.65, 13.20)	0.1354	0.0452	0.1531	0.0778

a DLs for each BA: CA (11.1 nM), CDCA (20 nM), DCA (12.9 nM), LCA (5 nM), and UDCA (11.8 nM)

Table 6: Kendall's Tau correlation coefficients

	CA	CDCA	DCA	LCA	UDCA
CA		0.47	0.25	0.11	.026
CDCA			0.40	0.23	0.36
DCA				0.53	0.28
LCA					0.21
UDCA					

adenomatous polyps, and colorectal cancers were collected from 2001 – 2007 at Vanderbilt University Medical Center. In addition, tumor specimens of a majority of colon cancer patients were also collected at the time of operation. In total, serum from 177 patients was collected. This included, 50 normal patients, 31 with adenoma, and 96 with colorectal cancer. This was a cross-sectional sample, there was no matching or sampling procedure performed.

Sample processing

Using a sensitive HPLC-mass spectrometry assay, we analyzed a portion of our collected samples for their serum levels of major BAs in their unconjugated, as well as glycine- and taurine-conjugated forms (15 individual BAs). For the purpose of illustration of the methods described here, we will focus only on the unconjugated BAs in this manuscript. The serum samples (100 μL) were processed by mixing them with an equal volume of D4-CDCA (1 μM) as an internal standard and subsequently with 1 mL of ice-cold acetonitrile (CH3 CN). The supernatant fraction was separated by centrifugation at 13,000 g for 20 minutes and evaporated to dryness using Speed-Vac. The samples were reconstituted using 150 μl of the mobile phase. Calibration samples containing known concentrations of five major bile acids were processed simultaneously with the patient serum samples.

Liquid chromatographic conditions and mass spectrometric detection

Analyses of serum BAs were carried out using a Surveyor HPLC system (Thermo-Electron) and mass spectrometric detection was performed using an LTQ mass spectrometer (Thermo-Electron) equipped with a standard API-1 electrospray source. To maintain maximum sensitivity, the selected ion monitoring (SIM) parameters for the detection of each bile acid were optimized. An Acquity BEH C18 column (1.0 mm x 10 mm, 1.7 μm, source) was used for all chromatographic separations. The column and autosampler tray temperatures were maintained at 60°C and 10°C, respectively. The mobile phases were made up of 10 mM tributyl ammonium acetate in (A) H2 O / CH3 CN(5:95) and in (B) H2O/CH3CN (95:5). Gradient conditions were as follows: 0 – 0.5 minutes, B = 32.5%; 0.5V – 5 minutes, B = 32.5 – 57.5%; 5 – 5.25 minutes, B = 57.5 – 100%; 5.25 – 8 minutes, B = 100%; 8 – 8.25 minutes, B = 100 – 32.5%; 8.25 – 13 minutes, B = 32.5%. The flow rate was maintained at 200 μL / minute. The total chromatographic run time was 13 minutes. The sample injection volume was 5 μL. The autosampler injection valve and syringe needle were flushed and rinsed with H2 O / CH3 CN (1 : 1) between each injection. The mass spectrometer was operated in the negative ion mode. Quantitation was based on selected reaction monitoring for the corresponding m / z ratios of each bile acid. Data acquisition and quantitative spectral analysis were conducted using Thermo-Finnigan Xcaliber version 2.0 and Thermo-Finnigan LCQuan version 2.5, respectively. The calibration curves were constructed by plotting the peak area ratios of each bile acid over the internal standard (D4-CDCA) against the analyte concentrations for a series of BA standards.

Statistical analysis

Our primary interest is to test whether there are statistically significant differences in the mean BA concentrations between normal, adenoma, and cancer patients. We are interested both in pair-wise differences as well as a trend from normal to cancer. We used ML (censored lognormal or mixture distribution) methods to estimate the means and their 95% confidence intervals, unless there was reason to believe that the data were markedly not lognormally distributed. We used the F-statistic to test the overall and trend when ML methods based on the lognormal distribution were used; the Peto-Prentice test was used for any BA that was estimated by KM methods and a likelihood ratio $\chi 2$ test statistic was applied for mixture models. We performed two types of hypothesis tests, pair-wise tests

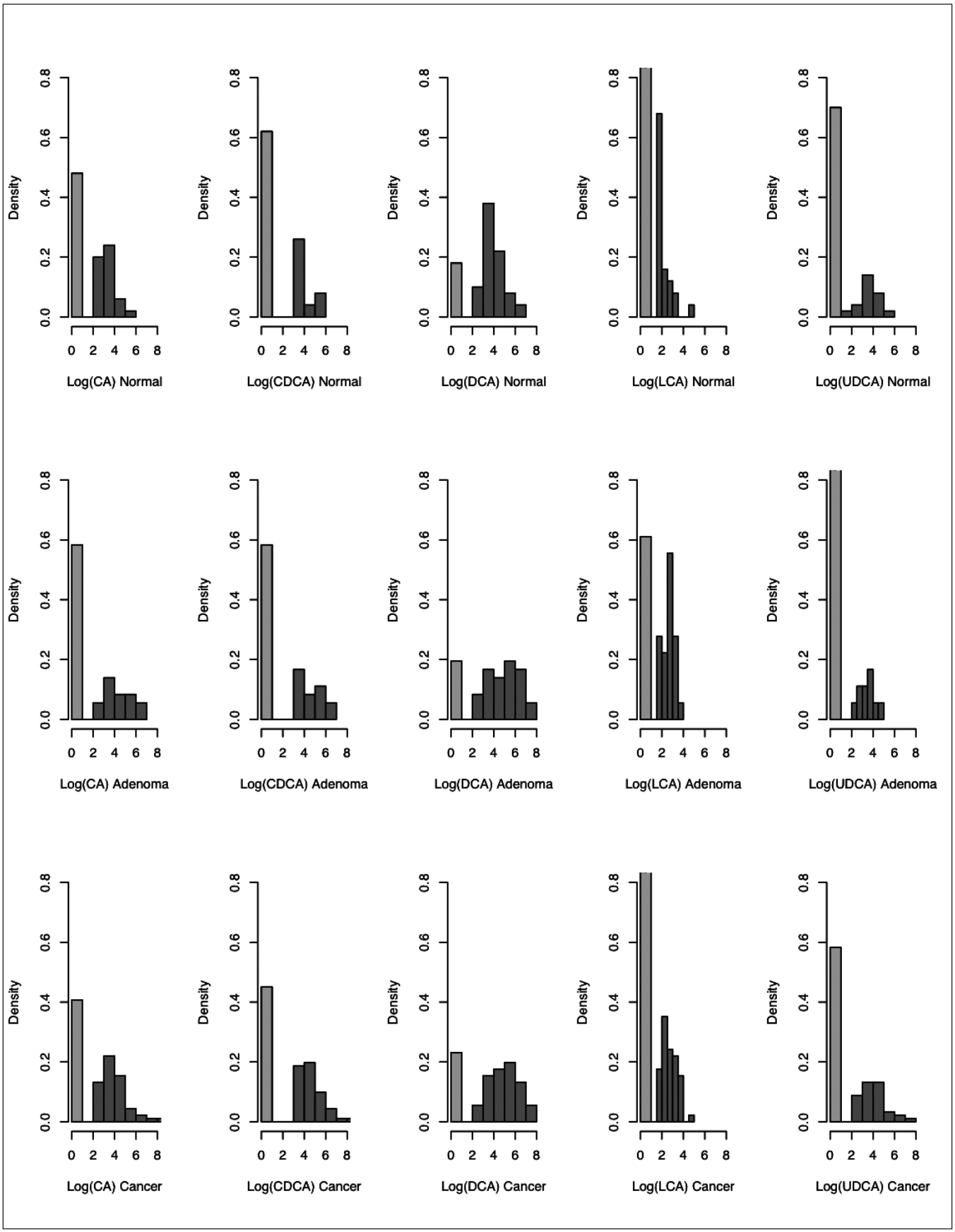

Figure 1: Histograms for log transformed bile acid concentrations (nM) for all patient groups

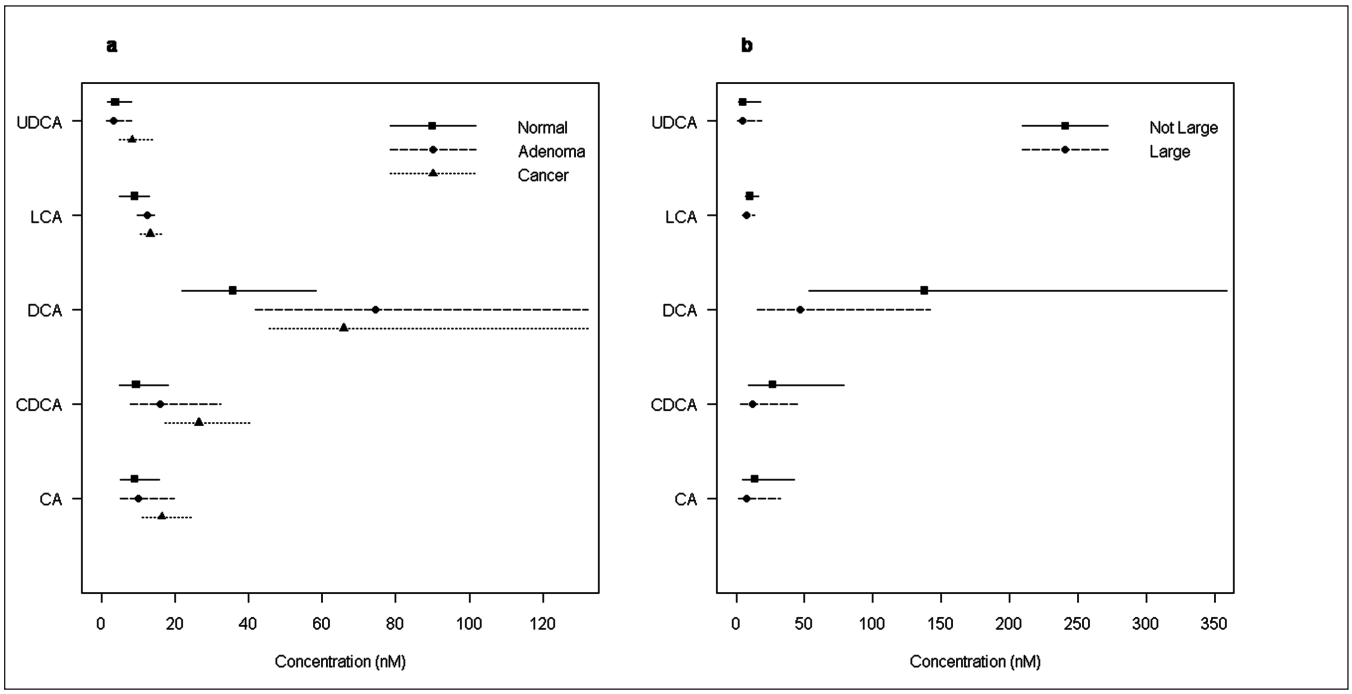

Figure 2: Mean concentration levels for each bile acid and 95% confidence intervals. These were obtained using censored lognormal models, the confidence intervals were not symmetric because the estimates were back-transformed from the log scale. (a) Sample sizes were 26, 18, 21, and 24 for Stages 1, 2, 3, and 4, respectively. *P*-values were 0.7413, 0.5963, 0.6168, 0.7371, and 0.2098 for CA, CDCA, DCA, LCA, and UDCA, respectively. (b) Sample sizes were 14 and 11 for Not Large and Large, respectively. The *P*-values were 0.5209, 0.3102, 0.1475, 0.4341, and 0.9001, respectively.

to assess differences between any two patient types (e.g., normal versus cancer) and a test for linear trend going from normal to cancer. To examine the distribution of the bile acids (as well as demonstrate the proportion of values below the DL) we examined histograms for each bile acid by patient group. Lognormal probability plots, along with 95% confidence intervals were used to judge the fit of the lognormal distribution to each BA. The difference between the observed proportion of BDL and the expected probability of BDL under a censored lognormal distribution was calculated for each bile acid. A difference greater than 0.1 indicated that a Bernoulli / lognormal mixture distribution should be used to estimate test group-wise differences.[3] We also examined the correlation between each unconjugated BA and the other BAs using a Kendall's tau correlation coefficient, per Akritas.[26] We examined BA concentration differences in adenomas less than 3 cm versus those greater than 3 cm, as a proxy for advanced stage, and between stages in cancer samples. Analyses were done in R version 2.8.0 (www.r-project.org) and SAS version 9.2 (SAS Institute, Carey NC).

RESULTS

Demographic and sample information can be found in Table 3. The pathological stage and adenoma subtype were based on a medial record search and was not obtained as part of the serum collection. Large adenomas were defined as any adenoma greater than 3 cm. Within the adenoma group there was one FAP patient, one patient with polypoid hyperplasia, and one patient missing a pathological record.

Figure 1 shows the histograms for each of the bile acids for each patient type (Normal, Adenoma, Cancer); the light gray bar between 0 and 1 denotes the proportion of samples that fall below the DL, the actual percentages are listed in Table 4.

The probability plots indicated that the distributions for LCA and UDCA concentrations depart from the lognormal distribution (not shown), as demonstrated by the lack of fit of the points to the 45° line. Furthermore, for UDCA, the differences in the observed and expected probability was 0.144, which indicated a larger-than-expected proportion of censoring, necessitating the need for mixture distribution modeling. We compared the estimates (and test statistics) from the ROS, KM, and ML methods for LCA, and the estimates were identical across the methods. Therefore, we decided to use the ML methods, despite the indication of a slight departure from the lognormal distribution.

We investigated by using mixture models to estimate the test differences in the UDCA concentrations. We used the Akaike Information Criterion (AIC) as a metric to choose between

models (Akaike, 1974). The AIC combines a model goodness-of-fit measure (the log likelihood) with a penalty for the number of parameters. Models can be ranked according to AIC, and the smallest AIC is the 'best' fitting model. From these comparisons there is no evidence to suggest that adding a Bernoulli component is necessary. Therefore, we have used ML estimation and hypothesis tests for this BA.

Table 5 shows the overall means for each BA, as estimated by modeling using the censored lognormal distrubution along with 95% confidence intervals for each patient group. There were higher mean DCA levels in adenoma and cancer patients compared to normal subjects, however, this difference was not statistically significant. It is likely that this was due to a high variability in this BA (demonstrated by very large confidence limits). LCA and CDCA showed statistically significant differences between the groups, with pairwise differences between normal versus cancer for CDCA (P-value = 0.0100) and between normal and both cancer and adenoma (P values = 0.0326 and 0.0059, respectively) for LCA. All p-values shown in Table 4 are based on test statistics from ML method (assuming a lognormal censored distribution) as well as 'straw man' P-values when substitution by zero was used. Note that if the values below the DL had been replaced by zeros, the statistically significant difference in LCA would not have been detected. CDCA, LCA, and UDCA, all showed statistically significant increasing trends in the mean concentrations going from normal to adenoma to cancer. Again, DCA also exhibited this trend, but with the high degree of variability present, it was not statistically significant.

We also tested the mean differences in BA concentrations for each of the four cancer stages and between them. However, the sample sizes for these subgroups were small (28, 18, 23, and 27 for stages 1 – 4, respectively) and we were not able to detect any statistically significant differences. Graphical representation of these means and 95% confidence intervals can be found in Figure 2. Mean LCA, CDCA, and UDCA concentrations for stage 4 are higher than for stages 1 – 3. Although CA concentrations are lower in stage 1 than all other stages, DCA concentrations show a general trend for increasing levels for stages 2 – 4, while stage 1 has mean concentration levels similar to stage 4. Similar to the patient group analyses, this BA has exhibited a large degree of variability, resulting in wide confidence limits.

The sample size was not large enough to examine adenoma subtypes, however, we used the adenoma size as a proxy for advanced adenoma (large adenomas were those larger than 3 cm). There were no statistically significant differences between large and not large adenomas, the mean concentrations and 95% confidence intervals for each bile acid can be found in Figure 2. The mean levels in DCA were lower for large adenomas, 46.92 nM (95% confidence interval: 15.51, 142.00), compared to 137.81 nM (95% confidence interval: 52.97, 358.54), but there was a lot of variability in concentrations, in both groups. The analysis did not include two patients with three and four polyps, the FAP patient, or the patient with polypoid hyperplasia, and there was one patient with missing pathology data.

Finally, we wanted to test the correlation between all five unconjugated BAs. We used Kendell's tau, which is a measure of concordance between two measures. Kendell's tau is based on ranks, and it adjusts for ties when many observations have the same value. Table 6 shows the correlation matrix for the BAs. The highest correlation, and only correlation above 0.5, is between LCA and DCA, the two most common BAs associated with colon cancer.

DISCUSSION

We have described the most commonly used methods for summarizing and analyzing data with BDL values. We used a simulation study to demonstrate the properties of these methods, particularly compared to the frequently used substitution methods. For the most part, when the underlying distribution could be described using a parametric distribution (normal, lognormal, Weibull, etc.) the ML methods worked best for censored data; meaning they had the lowest bias and highest power. When there was a large amount of skewness, the ML methods were still a viable alternative, although in special cases their advantage over nonparametric, or even substitution methods, was unclear.

As ML methods are easy to implement in most commercial statistical software, there is little reason for the continued use of naive substitution methods for data with observed DLs. There are cases when more sophisticated models must be used; and determination of when their use is required is both conceptual (depends on what is being measured and whether true zeros are expected) and statistical (whether there appear to be a greater than expected number of BDL values). For most practical problems, especially with a large percentage of BDL values, we do not advocate the use of MI methods with these types of data.

As an example, we examined the serum levels of five common unconjugated primary and secondary BAs from a population of patients with cancer and adenomous polyp disease, and compared them with serum from patients with normal colonoscopy. We used a combination of graphical methods (histograms and quantile–quantile (QQ) plots) for testing the probability of higher than expected proportion of BLD

values, to choose between ML, mixture distributions, or nonparametric methods of estimation and testing for group differences. We found statistically significant differences in CDCA concentration levels between normal and cancer groups, and between normal and both adenoma and cancer in LCA. We found a statistically significant increasing trend in concentrations (from normal to adenoma to cancer) for CDCA, LCA, and UDCA BAs. Of note, and pertinent to the actual focus of this manuscript, differences in LCA and the significant trend in CDCA would not have been detected if substitution by zero was utilized as the analytic method. We expected to see differences in the DCA concentration between the normal and cancer, and possibly between the normal and adenoma groups, but because of the high degree of variability in these measures we were unable to detect a statistically significant difference.

We did not find a strong pairwise correlation between these five BAs. A possible explanation is that the multivariate relationship between these BAs is 'compositional' in nature; compositional data arise when the total amount is fixed and variability is observed in the relative amounts of the different forms. Therefore, as one BA concentration increases the others would need to decrease, and large positive correlations are tempered by the fixed total.

This particular example, although pertinent to the discussion of prevention of colon cancer, was used primarily for illustrative purposes. Our future work will be to investigate all 15 BAs, the five unconjugated forms along with their glycine- and taurine-conjugated forms. Primary hypotheses include the examination of how these conjugated forms differ between patient groups, as well as their relationship with the unconjugated forms. These investigations will necessitate the development of new statistical methods, or an extension of the existing methods. Furthermore, we are interested in investigating the utility of using serum BA concentration as a predictor, or classifier, for cancer. Characterizing data when the predictor has DL remains an open statistical research topic. Addtionally, we are currently working on evaluating the predictive ability of BA concentrations for adenomous polyp disease and cancer. All of these future considerations will necessitate the implementation of current statistical methodologies, as well as extentions and novel strategies. Knowledge of appropriate statistical methods is an important aspect of all cancer research.

ACKNOWLEDGMENTS

This study was supported in part by funding from the National Cancer Institute (NCI): P50 CA95060 [Arizona Cancer Center SPORE in GI Cancer (LaFleur)] and P50 CA09513 [Vanderbilt GI SPORE Program (Merchant, LaFleur)].

REFERENCES

1. Shumway RH, Azari AS, Johnson P. Estimating mean concentrations under transfor- mation for environmental data with detection limits. Technometrics 1989;31:347-56.
2. Slymen DJ, de Peyster A, Donohoe RR. Hypothesis testing with values below detection limit in environmental studies. Environ sci technol 1994;28:898-902.
3. Moulton LH, Halsey NA. A mixture model with detection limits for regression analyses of antibody response to vaccine. Biometrics 1995;51:1570-8.
4. Chu H, Moulton LH, Mack WJ, Passaro DJ, Barroso PF, Muoz A. Correlating two continuous variables subject to detection limits in the context of mixture distributions. J R Stat Soc Ser C Appl Stat 2005;54:831-45.
5. Schisterman EF, Vexler A, Whitcomb BW, Liu A. The limitations due to exposure detection limits for regression models. Am J Epidemiol 15 February 2006;163:374-83.
6. Cole SR, Chu H, Nie L, Schisterman EF, Estimating the odds ratio when exposure has a limit of detection. International journal of epidemiology 2009;38:1674-80.
7. Nie L, Chu H, Liu C, Cole SR, Vexler A, Schisterman EF. Linear regression with an independent variable subject to a detection limit. Epidemiology 2010;21:S17-24.
8. Gilliom RJ, Helsel DR. Estimation of distributional parameters for censored trace level water quality data: 1. Estimation techniques. Water Resour Res 1986;22:135-46.
9. Gleit A. Estimation for small normal data sets with detection limits. Environ sci technol 1985;19:1201-6.
10. Newman MC, Dixon PM, Looney BB, Pinder JE. Estimating mean and variance for environmental samples with below detection limit observations. J Am Water Resour Assoc 1989;25:905-16.
11. Helsel DR. More than obvious: Better methods for interpreting nondetect data. Environ sci technol 2005;39:419A-23.
12. Helsel DR. Less than obvious - statistical treatment of data below the detection limit. Environ sci technol 1990;24:1766-74.
13. Lee L, Helsel D. Statistical analysis of water-quality data containing multiple detec- tion limits: S-language software for regression on order statistics. Comput Geosci 2005;31:1241-8.
14. She N. Analyzing censored water quality data using a non-parametric approach1. J Am Water Resour Assoc 1997;33:615-24.
15. Fleming TR, Harrington DP. Counting Processes and Survival Analysis. USA: John Wiley and Sons; 1991.
16. Lynn HS. Maximum likelihood inference for left-censored HIV RNA data. Stat Med 2001;20:33-45.
17. Lachenbruch PA. Analysis of data with excess zeros. Stat Methods Med Res 2002;11:297-302.
18. Schafer JL. Multiple imputation: A primer. Stat Methods Med Res 1999;8:3-15.
19. Willett W. The search for the causes of breast and colon cancer. Nature 1989;338:389-94.
20. Nagengast FM, Grubben MJ, van Munster IP. Role of bile acids in colorectal carcino- genesis. Eur J Cancer (Oxford, England: 1990) 1995;31A:1067-70.
21. Reddy BS. Diet and excretion of bile acids. Cancer Res 1981;41:3766-8.
22. Kishida T, Taguchi F, Feng L, Tatsuguchi A, Sato J, Fujimori S, *et al*. Analysis of bile acids in colon residual liquid or fecal material in patients with colorectal neoplasia and control subjects. J Gastroenterol 1997;32:306-11.
23. Owen RW. Faecal steroids and colorectal carcinogenesis. Scand J Gastroenterol Suppl 1997;222:76-82.
24. Bayerdorffer E, Mannes GA, Richter WO, Ochsenkuhn T, Wiebecke B, Kopcke W, *et al*. Increased serum deoxycholic acid levels in men with colorectal ade- nomas. Gastroenterology 1993;104:145-51.
25. Bernstein H, Bernstein C, Payne CM, Dvorak K. Bile acids as endogenous etiologic agents in gastrointestinal cancer. World J Gastroenterol 2009;15:3329-40.
26. Akritas MG, Murphy SA, LaValley MP. The theil-sen estimator with doubly censored data and applications to astronomy. J Am Stat Assoc 1995;90:170-7.

AUTHOR'S PROFILE

Dr. Bonnie LaFleur, is an Associate Professor of Biostatistics at the University of Arizona.

Dr. Wooin Lee, is an Assistant Professor in the Department of Pharmacology at the University of Kentucky.

Dr. Dean Billheimer, is an Associate Professor of Biometry and Director of the Statistics Consulting Laboratory at the University of Arizona.

Dr. Junmei Liu, is a statistician in the Statistics Consulting Laboratory at the University of Arizona.

Dr. Craig Lockhart, is an Associate Professor of Medicine in the Division of Oncology at Washington University.

Dr. Nipun Merchant, is an Associate Professor of Surgery at Vanderbilt University.