QUALITY ENGINEERING®
Vol. 16, No. 4, pp. 541–549, 2004

Using Logistic Regression to Evaluate New Lid Designs

Tony Gojanovic and Leslie Sidor*

Coors Brewing Company, Golden, Colorado, USA

ABSTRACT

New product designs from competing vendors must often be evaluated in order to make the best decision on which design to use in manufacturing. In most cases, the designs must be challenged over a range of environmental factors, process tolerances, and incoming material variations. At times, the performance characteristic may take on the form of a passing or failing response. Due to the dichotomous nature of a pass/fail outcome, ordinary regression techniques based on a normal distribution of error terms with constant variance are not appropriate. In the case of a pass/fail response, a natural distribution to consider is the binomial distribution. The authors have exploited the field of generalized linear models, specifically logistic regression, to select a new lid design for a beverage container based on the correct distributional assumptions.

Key Words: Applied logistic regression; Beverage can lid design; Pass/fail data.

INTRODUCTION

In manufacturing, performance characteristics measured on a continuous scale are common and desired. For modeling experiments, the traditional statistical tool that is used is the classical linear model. The assumptions required for the linear model are that the errors are normally distributed, the variance is constant and independent of the mean, and the model parameters are linear. However, there are cases when the best possible evaluation of performance has a binary outcome such as the defect is present/absent or the color matches a target/does not match a target. When the performance characteristic is dichotomous, the traditional assumptions related to the linear model are seldom met (Hamada and Nelder, 1997; Myers and Montgomery, 1997).

To coerce the data into a form more appropriate for linear regression, a transformation of the response could be considered (Freeman and Tukey, 1950). Assuming that the response is the sum of individual Bernoulli trials, the response could be transformed using the $a\sin^{-1}\sqrt{\pi}$, where π is the probability of an event (Bisgaard and Fuller, 1994–1995). Selecting a

*Correspondence: Leslie Sidor, Coors Brewing Company, Golden, CO 80401, USA; E-mail: leslie.sidor@coors.com.

DOI: 10.1081/QEN-120038016
0898-2112 (Print); 1532-4222 (Online)

transformation to simultaneously satisfy the assumptions of normality and constant variance is tenuous in light of models that would take advantage of the observed distribution.

A better solution to model a pass/fail response is to select a generalized linear model (GLM). This family of models, the GLMs, takes advantage of the distribution of the response and yields appropriate parameter estimates. The specific GLM used for modeling depends on the selection of the link function. For a more detailed discussion of GLMs, the authors refer the reader to Refs. (Hamada and Nelder, 1997; Hosmer and Lemeshow, 1989; McCullagh and Nelder, 1996; Myers and Montgomery, 1997).

The performance characteristic, a score panel eversion, measured for the experiment was pass/fail (pass: the score panel opened normally or fail: the score panel blows outward). The performance characteristic consisted of independent Bernoulli trials with parameter π, and the sum of the Bernoullis has a binomial distribution with parameters n and π. This makes the binomial distribution a logical choice for the response distribution. For a response with a binomial distribution, a logit link is selected. By selecting the logit link, the function used for modeling the probability of a nonfunctional lid is

$$E(y_i) = \pi_i = \frac{e^{\mathbf{X\beta}}}{(1 + e^{\mathbf{X\beta}})},$$

where $\mathbf{X\beta}$ is $\beta_o + \beta_1 x_1 + \cdots + \beta_k x_k$.

The predicted response, $\hat{\pi}$, is the probability of an event occurring and $1 - \hat{\pi}$ is the probability of an event not occurring. This is the logistic model. The logistic function is sigmoidal with left and right asymptotes at 0 and 1, respectively. The parameter estimates are developed through maximum likelihood estimation based on an iteratively weighted least squares algorithm.

Once the logistic model has been fit adequately, interpretation of the coefficients or parameters can be made. Unlike the least squares model, the interpretation of the coefficients in the logistic model is somewhat different. For a dichotomous independent variable, the estimated coefficients represent the odds of an outcome being present among subjects coded as 1 relative to subjects coded as 0 or how much more likely an outcome is to be present among those with $x = 1$ than among those with $x = 0$. The interpretation of the dichotomous independent variable can be easily generalized to a polytomous independent variable. For a continuous independent variable, the parameter estimate is the change in the log odds for an increase of "1" unit in x. The odds ratio for a given coefficient, β_i, and the $1–\alpha/2$ confidence interval is

$$\text{odds ratio} = e^{\left[\hat{\beta}_i \pm z_{1-\alpha/2} SE(\hat{\beta}_i)\right]},$$

where SE is the standard error for the ith coefficient.

LID MANUFACTURE

Coors Brewing Company produces malt beverage products that are shipped in aluminum cans, bottles, and kegs. The integrity of the container must be of the highest quality and integrity so that the consumer receives the product in pristine condition. A great majority of our product is shipped in aluminum cans, which are composed of a can body and a lid. Coors introduced the all-aluminum can in 1959, and today the aluminum can is one of the most widely used containers in the world. Coors produces its own lids at the Rocky Mountain Metal Container End Manufacturing facility in Golden, Colorado. Annual production of lids for Coors exceeds 2 billion, and a small increase in an otherwise small defect rate can result in many dissatisfied consumers.

The beverage lid forming process begins with rolled aluminum "sheet," which typically comes in 4000-pound coils. The 0.0090-inch thick aluminum metal is fed into a press where the basic form of the lid, the shell, is stamped out. The shell is subsequently "curled" in order to allow the lid to fit over a beverage can and be seamed. The curled shell then goes to a lining operation in which a thin gasket of latex is applied under the curl. The purpose of the latex is to provide a hermetic seal with the can body. Once the latex is applied and cured, the next major process is conversion. During conversion, the score panel is formed along with a "rivet" by passing through a series of forming tools in a conversion press. A blade "cuts" the score residual line, which allows the score panel to tear under pressure from the tab and fold downward so that the liquid contents will pour out. At the same time the shell is going through the conversion press, another press working simultaneously within the conversion press produces a tab. The tab is then staked onto the rivet to produce a finished lid. The converted lid (Fig. 1) then is "bagged" into long paper sacks and sent to a filling operation for use on cans.

Typical lid manufacturing tolerances are measured to the ten-thousandths of an inch. Quality inspection of the lid includes dimensional as well as

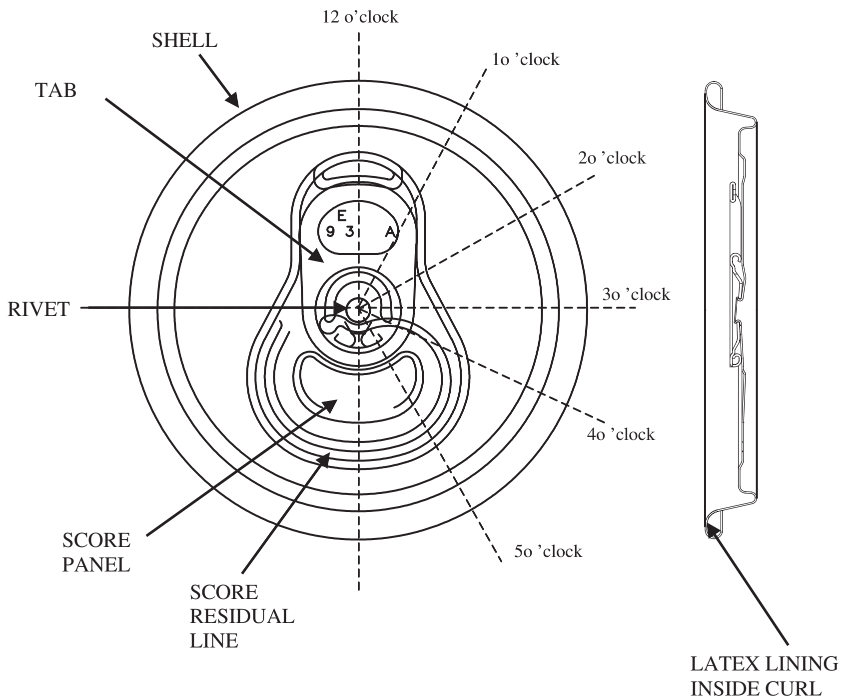

Figure 1. Beverage can lid with metal grain orientations (dashed lines).

functional checks. A lid must be able to withstand the high pressures of carbonation, elevated temperatures, and shocks incurred from rail and truck shipping of filled cans and ultimately provide a leak-free container that is easily opened by the consumer. One critical feature occurs in the conversion operation with the formation of the score panel and the amount of metal left in the score residual line. A shallow cut, or a high score residual, will not allow the tab to open the lid, whereas a deep cut, or a low score residual, can result in a lid that opens during shipping, resulting in a partial or empty container for the customer. The typical tolerance for the score residual is 0.0008 inch. However, failures do occur.

One failure that prompted this study was incorrect opening of the score panel. Under elevated temperatures, pressure inside the can increases. The failure in question occurred when a warm container was opened. Instead of the score panel folding into the can, it opened outward. This type of failure is called an eversion and is not desired. Eversions can cause product to spew and splash from the can onto the consumer, as well as provide a sharp edge for injury. Although eversions were infrequent, the root cause of the phenomenon required complete understanding to minimize their long-term occurrence.

EXPERIMENTAL DESIGN

From expert knowledge obtained through several brainstorming sessions, it was hypothesized that eversions were caused by a combination of internal can pressure, score residual depth, grain orientation of the metal, and the general design type from differing manufacturers of the tooling. The exact magnitude of each effect was not known, and a designed experiment was developed to ascertain the effects and any interactions. Metal grain could not be controlled practically in the process on a day-to-day basis but was of interest for questions of robustness, e.g., do different grain orientations produce different likelihoods of an eversion. Consequently, grain orientation was controlled for experimentation purposes. Score residual, pressure, and different lid designs could be controlled in an experimental setting. Score residual and design are controlled in manufacturing. Pressure is highly dependent on the consumer and could not be controlled but was of

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 1. Design matrix summary.

Factor	Level					
A: Lid design	Design 1		Design 2		Design 3[a]	
B: Score residual	−2 Low	Low	Nominal		High[a]	
C: Grain orientation	12 O'clock[a]	1 O'clock	2 O'clock	3 O'clock	4 O'clock	5 O'clock
D: Pressure	54 psig to 81 psig in 3-psig increments					

[a]Reference level in the logistic model.

interest for robustness questions. Other features of the lid that may have contributed to eversions were left at nominal settings between all lid designs tested as determined by best engineering knowledge.

The primary objective was to select the most robust lid design (factor A). Three tooling designs were submitted for evaluation: the current design and two designs submitted from outside manufacturers.

The levels for score residual (factor B) were chosen relative to the specification as high, nominal, low, and under specification by 0.0002 inch. The choice of the four different score residual levels was based on the common-sense assumption that as the score residual drops; i.e., as the tooling produces a deeper cut into the lid, the higher the likelihood of an eversion under elevated can pressures. The goal was to characterize the entire specification range, hence the inclusion of higher score residuals in the test.

Grain orientation of the metal (factor C) was the machine direction of the metal coil relative to the tab. Six grain orientations were selected and were described using the face of a clock with the tab in the 12 O'clock position (Fig. 1). Since grain orientation cannot be controlled and since the goal was only to understand design robustness, the team agreed to use only six categories for describing grain orientation.

Pressures (factor D) for the experiment varied from 54 psig to 81 psig in 3-psig increments. The pressure range was selected based on a theoretical range of the consumer's worst-case opening conditions; e.g., opening a beverage container with a given carbonation level after it had been sitting in a car trunk in the hot sun. Although the 3-psig increments resulted in 10 levels for this factor, this type of designed experiment for eversion failures had not been run before in our industry, to the best of our knowledge. Therefore, the characterization of the design space with adequate detail was critical. In retrospect, the same results could have been obtained with a reduced number of pressure levels.

A full factorial was used to select the best lid design to be used for manufacturing. A total of four factors were evaluated with 20 replications at each level (Table 1). The full factorial design resulted in 720 runs and with replication 14,400 trials. There was significant debate on fractionating the design, eliminating factors, and limiting the sample size. Logistic regression was ultimately to be used for the analysis tool. Therefore, there was a concern that alias patterns would not be discernable since logistic regression parameters do not necessarily produce orthogonal parameter estimates. Additionally, a sample size of 20 lids for each run was deemed a minimum due to the pass/fail response and sensitivity issues surrounding a dichotomous response variable. The factors chosen were determined to be necessary, and it was the internal testing group that pushed to add factor C, grain orientation of the metal, even though it would create six times the work. Later on, the addition of factor C would eliminate some important controversies and ultimately provide a better model fit by reducing the overall error component. Additionally, previous experience on pilot studies prior to conducting the full experiment indicated that 20 lids per run produced realistic parameter estimates with reasonable standard errors. Real-time constraints and financial considerations were critical but were tempered with the need to perform an experiment that was done right once and produced clear results, even at the expense of extending the time line and budget slightly.

Test product was produced according to the matrix. The experiment was randomized in order to assure that subsequent results would generalize and statistical optimality properties were achieved.

Once the different lid factors were produced; i.e., score residuals, grain orientations, and design type, they were marked and physically randomized. The randomized ends were then divided into 10 groups of 1440 lids, representing an equal quantity of test levels and samples that were then subjected to a randomized pressure between 54 and 81 psig. Changing individual pressures randomly for each level of the other factors would have been prohibitive. In order to track the logistics of the testing operation, we developed worksheets and a large board with cups to hold lids for testing to assure that each pressure block saw the same quantities of randomized ends for the test factors in order to produce a balanced design. Data sheets were input daily and checked for errors as testing progressed. Because all design types, score residuals, and grain orientations were randomized, the statistical inference criteria were met since treatments were equally represented across the randomized pressure blocks. Additionally, random verification samples were frequently run through the testing equipment to assure that the machine was working properly and not changing over the weeks of testing. Verification also included that all machine settings were set at the correct levels prior to testing. The pressure equipment is automated, which reduced operator error.

DATA ANALYSIS

Exploratory analysis of the data revealed the sigmoidal shape of the logistic regression model. The data were graphed by design, factor A, and by pressure, factor D, pooled over score residual, factor B (the amount of metal remaining in the score portion of a beverage lids), and grain orientation, factor C (Figs. 2–4). The height of each bar represents the total number of failures observed at each pressure tested, and the solid line represents the predicted probability of failure using the logistic model. It was noted that with factor A, design 1, two different failure modes were observed. Although logistic regression can be used to model a polytomous response, discussions with the engineers revealed that the second failure mode was mechanically the same as the first failure mode. Therefore, the data were recoded to reflect the reclassification of the two failure modes into a single failure mode.

It is apparent from the exploratory plots that design 3 (Fig. 4) had the lowest proportion of failures and would be the most logical choice to adopt for manufacturing lids. The effect of score residual and grain orientation on the number of failures observed was not as obvious. Therefore, to model the data completely, a logistic model was necessary.

The data were analyzed using SPSS version 8.0 and SYSTAT version 8.0. Since the study focused on choosing the design with the greatest likelihood of success, the main effects were of primary concern. Concentrating on the main effects was considered reasonable since the proposed main effects model did adequately characterize the data.

RESULTS

First, the overall fit of the model was assessed. The log-likelihood statistic tested the hypothesis that all of the coefficients, less the constant, were equal to zero. The likelihood ratio statistic was analogous to the F-test for linear regression models. The p-value for the likelihood ratio statistic was <0.0001. This was significant at the 95% level. Another statistic used to assess overall fit was McFadden's Rho-Squared value. McFadden's Rho-Squared mimics

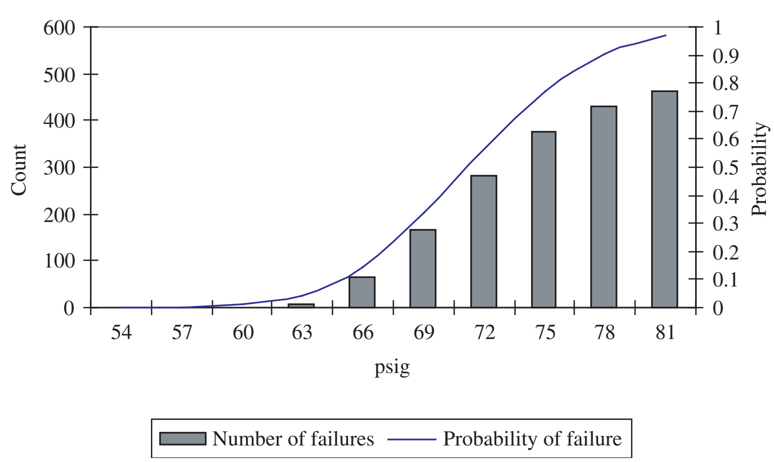

Figure 2. Design 1 by pressure (grains and scores combined). *(View this art in color at www.dekker.com.)*

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

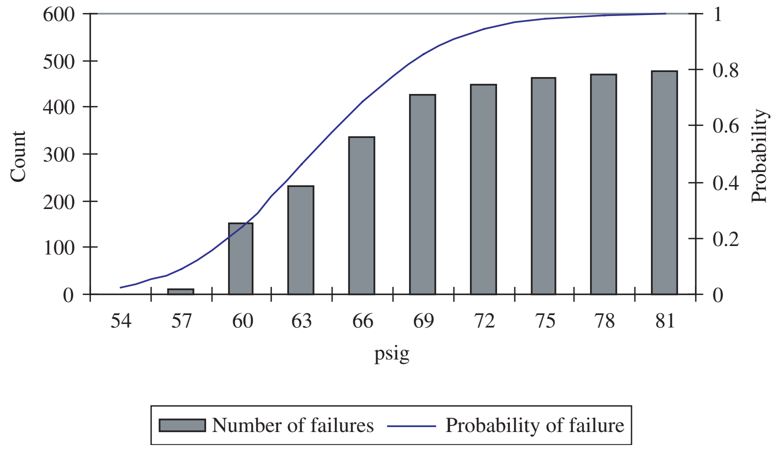

Figure 3. Design 2 by pressure (grains and scores combined). *(View this art in color at www.dekker.com.)*

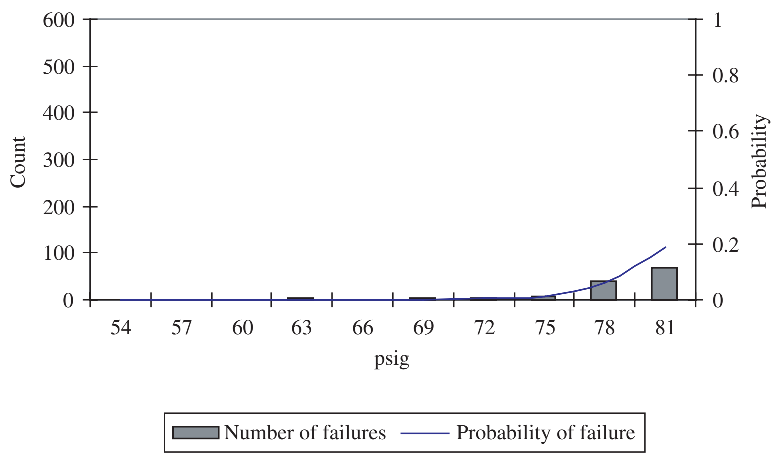

Figure 4. Design 3 by pressure (grains and scores combined). *(View this art in color at www.dekker.com.)*

the R-square found in linear regression. This value varies between 0 and 1 but is typically much lower than the traditional R-squared value. Values between 0.2 and 0.4 are considered satisfactory (Steinberg and Collo, 1998). The McFadden's Rho-Squared for this logistic model was 0.756, which indicated a respectable fit. Now that the adequacy of the model had been determined, the coefficients could be interpreted.

The independent variables that were categorical were "dummy" coded, with the last value becoming the reference group. In the case of factor A, design, design 3 became the reference group. The parameter estimates and odds ratios with 95% confidence intervals were provided in Table 2. Note that all of the main factors were significant except for C1, which was considered to be marginal but important.

Since the reference category for factor A was design 3, the odds ratios tell us that design 2 was 99,266 times more likely to fail than design 3. Design 1 was 1720 times more likely to fail than design 3. These results were commensurate with Figs. 2 through 4 and clearly indicated that design 3 provide superior performance. The pressure response was continuous and represented the odds of a failure given a 1-psig increase in pressure. The pressure variable reflected possible pressures in beverage containers experienced by the consumer. The odds ratio for pressure indicated that for every 1 psig increase, the odds of a lid failure would increase 1.7 times.

Finally, additional diagnostics were also available to assess the goodness of fit of the logistic model.

Table 2. Logistic model coefficients.

Coefficient	Estimate	S.E.	*t*-Ratio	*p*-Value	Odds ratio	95% Bounds Upper	95% Bounds Lower
Constant	−45.563	0.993	−45.88	<0.0001			
A1: design 3 vs. design 1	7.4560	0.183	40.75	<0.0001	1720.26	2461.65	1202.17
A2: design 3 vs. design 2	11.506	0.254	45.32	<0.0001	99266.33	163274.73	60351.07
B1: high score vs. −2 score	3.498	0.132	26.51	<0.0001	33.06	42.82	25.52
B2: high score vs. low score	2.539	0.123	20.56	<0.0001	12.66	16.13	9.94
B3: high score vs. nominal	1.596	0.118	13.57	<0.0001	4.93	6.21	3.92
C1: 12 O'clock vs. 1 O'clock	0.209	0.127	1.65	0.100	1.23	1.58	0.96
C2: 12 O'clock vs. 2 O'clock	−0.679	0.130	−5.24	<0.0001	0.51	0.65	0.39
C3: 12 O'clock vs. 3 O'clock	−3.922	0.159	−24.65	<0.0001	0.020	0.027	0.014
C4: 12 O'clock vs. 4 O'clock	−2.474	0.143	−17.30	<0.0001	0.084	0.11	0.064
C5: 12 O'clock vs. 5 O'clock	−0.954	0.131	−7.28	<0.0001	0.39	0.50	0.30
D: pressure	0.526	0.012	44.81	<0.0001	1.69	1.73	1.65

Notes: Log likelihood of constants only model = LL(0) = −9259.797.
2× [LL(N)–LL(0)] = 14007.832 with 11 df Chi-sq *p*-value = 0.000.
McFadden's Rho-Squared = 0.756.

Table 3. Model prediction success.

Observed classification	Predicted classification: Response (failures)	Predicted classification: Reference (passes)	Actual totals
Response (failures)	4269.7	670.3	4,940
Reference (passes)	670.3	8789.7	9,460
Predicted total	4,940	9,460	14,400
Correct	0.864	0.929	
Success ind	0.521	0.272	
Total correct	0.907		
Sensitivity: 0.864		Specificity: 0.929	
False reference: 0.136		False response: 0.071	

These diagnostics were provided by SYSTAT. The diagnostics are related to the classification power of the model (Table 3). The classification power describes the model's ability to correctly and incorrectly predict eversions (ends with the failure) and noneversions (ends with normal opening functionality).

DISCUSSION

When an experiment is conducted in a manufacturing environment, there are a number of obstacles to overcome and decisions to be made that have the potential for compromising the validity of the experiment. This section outlines the specific challenges in running this experiment and the approaches taken to overcome the issues.

Factor/Level Selection

The factors grain orientation and score residual could have been measured using a continuous scale (grain orientation: degrees, and score residual: inches). However, to reduce experimentation time and cost, the actual metal grain orientation and score residuals were placed into logical groupings of values as described in the Experimental Design Section. By taking an analysis of variance (ANOVA) like approach to the factor/level assignment, modeling precision was compromised. Given that the primary objective of the experiment was design selection in terms of product ruggedness, the trade-off was deemed acceptable.

Restricted Randomization

Another compromise made in the modeling process was relative to the restricted randomization of the experiment. As noted in the Experimental Design Section, the experiment was blocked on pressure to control the cost and experimentation time. The restricted randomization produced a split-plot design-type scenario. A split-plot design would require the interpretation of both factors in the whole plots, the pressure block, and the subplots. However, the interpretation and the estimation of the standard errors were treated as if the experiment were completely randomized. At the time of the analysis, the compromise to the model was considered to be low risk. Since the factor design type was assigned to a subplot, design type was not confounded with the pressure blocks. This provided a clean interpretation of each of the three designs relative to each other, which was the primary objective of the experiment. This risk may not have been appropriate if a design type by pressure interaction was of interest or suspected.

Sample Size Considerations

In reference to the sample size of 20 lids per run, further improvements to the original designed experiment could have been made that would have reduced the cost and time devoted to the experiment. After conducting the experiment, a guideline found in van Belle's *Statistical Rules of Thumb* (van Belle, 2002) states that 10 events per study provide stable estimates in a logistic regression experiment. Given the data observed, the original sample size of 20 lids at some of the lower pressures was probably too low (especially when evaluated at high score residuals). Conversely, a smaller sample size would have been adequate at the higher pressures. Another option to reducing the sample size would have been to run a smaller screening study to provide better focus to the scope of the experiment.

Analysis Tips

Graphing the data first provided the least complex, yet easy, method to interpret the results. The plots offered an opportunity to discuss issues such as the characterization of multiple failure modes with the design engineers and insight to possible modeling concerns. Upon inspection of the plot for design 3 (Fig. 4), it was obvious that the low number of observed eversions could yield mathematical instability to the logistic model. In fact, the low number of eversions for design 3 generated very wide confidence intervals for the odds ratios labeled A1 and A2 in Table 2. Had there been problems with convergence or exceptionally high standard errors associated with the A1 and A2 coefficients, alternative modeling techniques would have been explored.

CONCLUSIONS

This article presents a method to model pass/fail data under the financial constraints of an experiment conducted in a manufacturing environment. The logistic model provided an objective tool for selecting a lid design and making statements in reference to lid design robustness under different process and consumer conditions. Using the results of the model, design 3 was successfully implemented; to date, no failures have been reported.

In addition to answering the specific research question related to the selection of the most robust lid design, the knowledge gained regarding design robustness given specific process parameters such as the low score residual and the 12 O'clock grain orientation provided focus for future lid evaluations. Since the concern for manufacturing is to characterize and evaluate worst-case situations, testing for subsequent material qualifications has been dramatically reduced. The model also aided in the development of an eversion specification. The development of this new specification is key to monitoring the process and assuring product quality.

ABOUT THE AUTHORS

Tony Gojanovic is the Quality, EH&S Manager at Coors Rocky Mountain Metal Container End Division. He has worked for Coors for 15 years. He obtained both undergraduate and graduate degrees in Mathematics and Statistics from the University of Colorado with a Master's Thesis on Failure Mode Effects Analysis. He has worked with a partnership between the University of Colorado and Lockheed Martin developing computer simulations to assess the risk of a successful landing of the next generation of Mars landers given varying surface terrains. His interests are in risk management, statistical design of experiments, and robust statistics. He is a member of the American Statistical Association.

Leslie Sidor is a Principal Quality Engineer at Coors Brewing Company and has worked at Coors for 18 years. She has been a certified ASQ Quality Engineer since 1986. Undergraduate and graduate degrees in the Biological Sciences were received from Wayne State University, Detroit, MI, and her master's degree in Biometrics was received from the University of Colorado, Health Sciences Center. She has guest lectured at the Colorado School of Mines and the Daniel's College of Business, University of Denver on data quality in statistics and strategic test design. In addition, she has presented papers on the application of Taguchi methodology for process optimization at the annual ISTA/IoPP technical conferences and the Rocky Mountain Chapter of the IBPA.

REFERENCES

Bisgaard, Soren, Fuller, Howard T., (1994–1995). Analysis of factorial experiments with defects or defectives as the response. *Quality Engineering* 7:429–443.

Freeman, Murray F., Tukey, John W., (December 1950). Transformations related to the angular and the square root. *The Annals of Mathematical Statistics* 21(4):607–611.

Hamada, Micheal, Nelder, John A., (July 1997). Generalized linear models for quality improvement experiments. *Journal of Quality Technology* 29(3):292–304.

Hosmer, David W., Lemeshow, Stanley. (1989). *Applied Logistic Regression*. New York: John Wiley and Sons, Inc.

McCullagh, P., Nelder, J. A. (1996). *Generalized Linear Models*. 2nd ed. London: Chapman and Hall.

Myers, Raymond H., Montgomery, Douglas, C., (July 1997). A tutorial on generalized linear modes. *Journal of Quality Technology* 29(3):274–291.

Steinberg, Dan, Colla, Phillip. (1998). Logistic regression. *Systat 8.0 Manual*. Chapter 17, Chicago: SPSS Inc., 517–583.

van Belle, Gerald. (2002). *Statistical Rules of Thumb*. John Wiley & Sons, 87–88.