

## Comparing Reliabilities of the Strength of Two Container Designs: A Case Study

Esteban Walker<sup>1</sup> and Frank Guess<sup>1</sup>

<sup>1</sup>*University of Tennessee, Knoxville*

*Abstract:* Two designs for PET (polyethylene terephthalate) beverage bottles were compared. These bottles are used for carbonated beverages; and thus, a very critical property is their burst strength. The burst strengths of bottles from each design across 24 cavities were measured. Standard nonparametric methods suggested a highly significant difference in the reliability of the two designs. Using simple graphical techniques, it was found that the reliability data of the new design appeared to be a mixture of distributions caused by the presence of “arly mortality,” due possibly to different failure modes. Even though they were clearly different, neither design was uniformly more reliable than the other. Standard parametric methods showed inadequate fit due to the bimodality of the strength data of the new design. The paper stresses (1) the need of operational clear definitions for “reliability,” (2) the need of graphical exploratory analysis to discover anomalies in the data, and (3) the value of nonparametric methods, and (4) the problems of using parametric techniques when the assumptions are violated. To justify work on improvement of the new design, the potential effect of the removal of the early mortality on the new design was analyzed.

*Key words:* Bursting strength, early mortality, lack of fit, mixture, Weibull.

### 1. Introduction

A very critical characteristic of PET (polyethylene terephthalate) bottles for carbonated beverages is the bursting strength of the bottles. In

a certain facility, these bottles are produced using 24 cavity stretch blow machines that produce 440 bottles per minute. Bottles are made from injection molded pre-forms that are manufactured elsewhere. The bursting strength is measured using a plastic pressure tester device. The bottle is filled with water and pressurized until it bursts or a pressure of 250 psi is reached (which could produce a potential censoring).

A new mold design was introduced recently. The advantage of this design is that it reduces the time to change and thus the downtime of the blow machines. However, the bottles made with the new design appeared to have problems with low bursting strength. To investigate and assess this problem, one bottle per cavity per day over a period of 32 days were tested from each design. The data appear in the Appendix. Section 2 includes the sequence of analyses performed and discusses the statistical aspects and validity of the results. All the techniques used are standard and available in many commercial statistical programs. JMP (SAS, 2001), which is widely available and fairly “user-friendly,” was utilized to generate all the graphs and results (Sall, Lehman, Creighton, 2001). Section 3 discusses the practical consequences of the results and examines potential gains. A summary and conclusions are included in Section 4. For more information on reliability or strength data see, for example, Barlow and Proschan (1981) or Meeker and Escobar (1998).

## 2. Analysis and Results

The first step in investigating the causes of the low bursting strength was to determine if there was a time effect. That is, whether the strength of the bottles changed over time. A scatterplot of the strengths against time revealed no significant trends on either design.

Another issue in comparing the reliabilities for the two designs was to determine if there were any “cavity effects.” In other words, whether there were differences among the 24 cavities within each design. This is crucial because if differences exist, the concept of “reliability of a design” is not well defined and comparisons between designs are meaningless. Figure 1 shows dot plots of the psi at failure of each cavity for the two designs. At first glance, it appears that most points in the new design are below 200

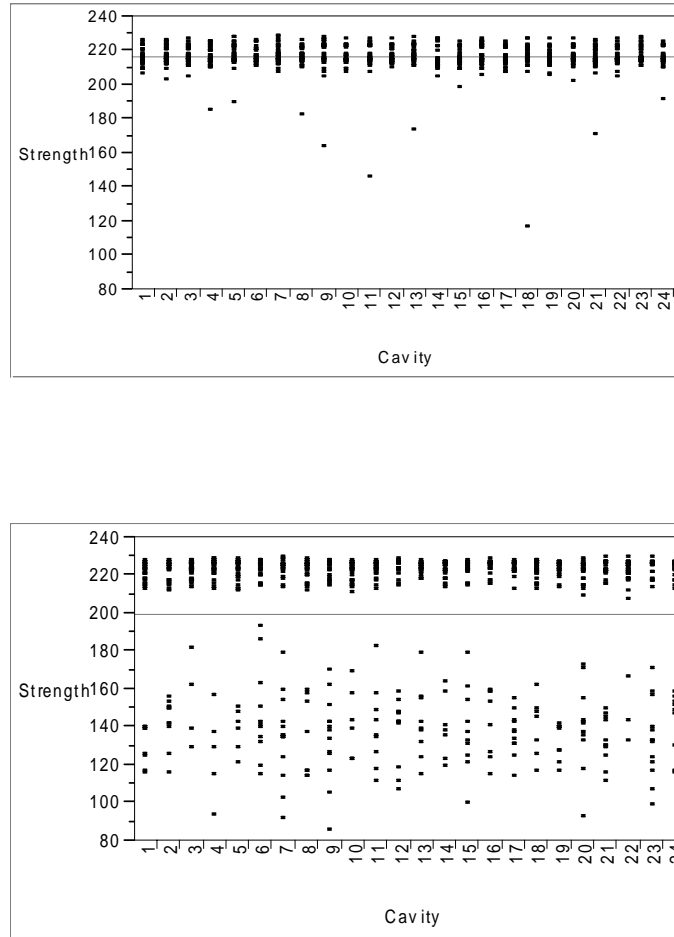


Figure 1. Dot plots of the bursting strengths of 32 bottles for each cavity for the old (top) and the new design (bottom). The horizontal line is the average strength.

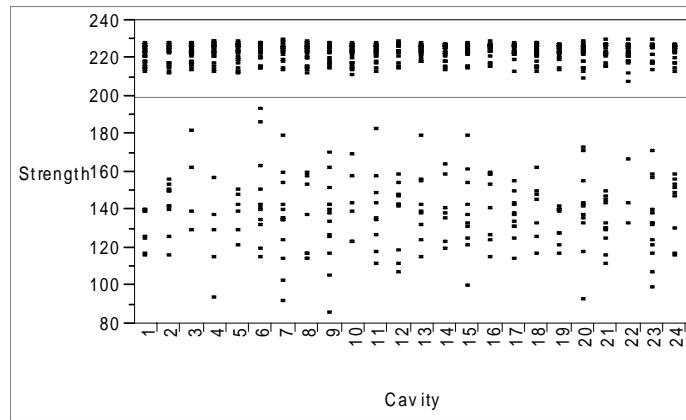
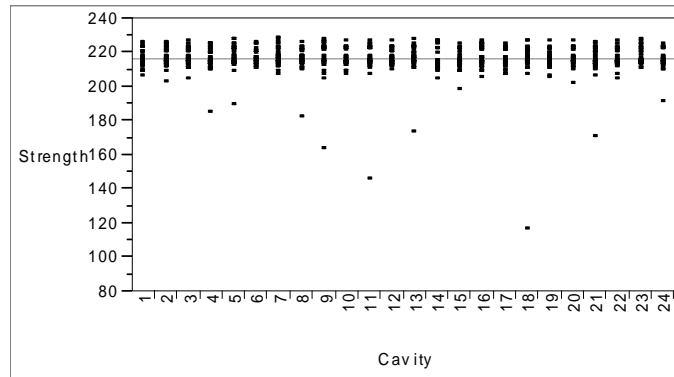


Figure 2. Kaplan-Meier estimators for each cavity for the old (top) and the new design (bottom).

psi. This effect is due to overlapping points, since about 73% of the observations of the new design are above 200 psi. This fact will be more evident later.

The Kaplan-Meier estimators of each cavity for the two designs are shown in Fig. 2. Since the distributions are fairly similar within designs, shift differences between cavities were tested using the Kruskal-Wallis test. The  $p$ -value for the old design was 0.30 and 0.67 for the new design, suggesting that the differences between cavities were not significant. These results justify pooling the cavities to compare the designs.

Fig. 3 shows side-by-side box plots based on the quartiles of the bursting strength of each design across all cavities (sample size for each design:  $n = 32 \times 24 = 768$ ). Given the sample size, a two-sample  $z$ -test was used to compare the means. The  $p$ -value was  $< 0.0001$  suggesting a significantly higher mean for the old design. On the other hand, the median test ( $p < 0.0001$ ) indicated that the median was larger for the new design. (Aside: the Wilcoxon rank sum is not appropriate to test differences in the designs because the distributions differ in shape. See Hollander and Wolfe, 1999.) The two-sample Kolmogorov-Smirnov test also indicated that the distributions were different ( $p < 0.0001$ ). Table 1 shows that the old design is superior with respect to the 25th percentile and the mean, whereas the new seems better with respect to the median and the 75th percentile. It is clear, that without an operational definition of reliability no design can be deemed superior to the other.

Table 1: Percentiles and mean for each design

Design	25th percentile	50th percentile (median)	75th percentile	Mean
NEW	160	223	226	199.40
OLD	214	216	221	216.42

A graph of the Kaplan-Meier estimators (Fig. 4) reveals the difference between the reliabilities of the two designs and helps to explain the results

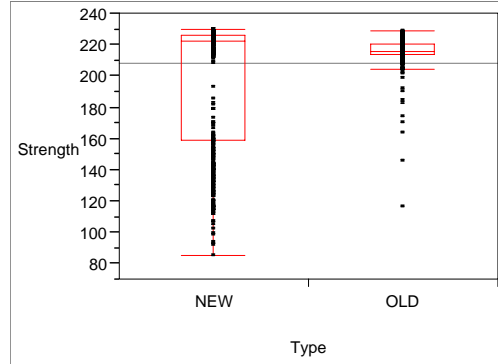


Figure 3. Box plots and quartiles of the bursting strength of the two designs.

obtained previously. The curve for the old design is above the one for the new design for pressures below about 215 psi. This means that the reliability for the new design is lower than the old design for pressures below 215 psi. At higher pressures; however, the curve for the new is above the one for the old, suggesting that the reliability for the new design is higher than the old for pressures above 215 psi.

The drop in the reliability of the new design around 150 psi is caused by a phenomenon called "early or infant mortality." (see, for example, Barlow and Proschan, 1981 or Guess and Proschan, 1988). These are bottles that, for some reason, burst at an unusually low pressure. Early mortality occurs sometimes when not all the reasons for failure have the same origin, in other words, there might be different failure modes present (see, for example, Meeker and Escobar, 1998, or Jensen and Petersen, 1982). As will be seen later, if this earlier mortality could be removed, the new design would have an overall better reliability than the old (i.e., the reliability function for the new design would be higher than that for the old design everywhere).

The histograms (Fig. 5) further confirm the presence of early mortality mentioned before. This early mortality is the reason why the mean reliabi-

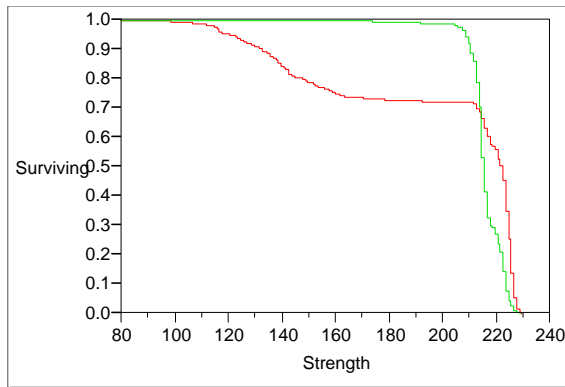


Figure 4. Product-limit survival fit (Kaplan-Meier) to the two designs.

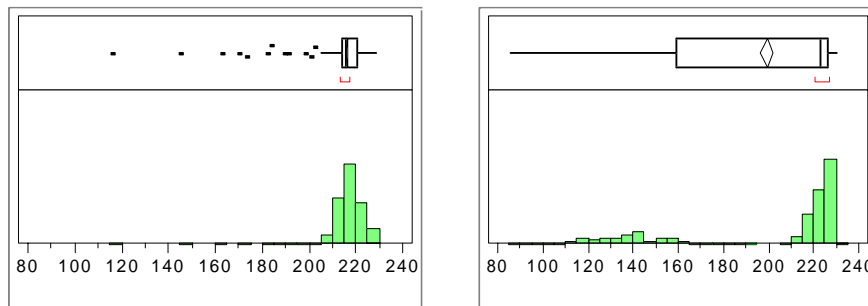


Figure 5. Histograms of the bursting strengths for the old (top) and the new design (bottom).

lity of the new design is lower than that for the old, but the medians behave in the opposite way. The other interesting fact that is visible in the histogram (and also Fig. 3) is that for the new design there is a very well defined gap between the bottles that fail early and those that do not. The reliability of the new design appears to be a mixture of different distributions, suggesting the presence of two or more failure modes.

Parametric analysis of the data was also carried out by fitting commonly used distributions. Weibull and lognormal plots showed that these distributions had significant lack of fit. This is primarily due to the bimodal nature of the data of the new design. Consequently, the many techniques that are based on common distributional assumptions would not be appropriate to analyze these data. Actually, a parametric regression analysis based on the Weibull distribution yielded a  $p$ -value of 0.46, suggesting that there were no differences between designs.

### 3. Improving the new design

A bottle with carbonated content at a high enough pressure can explode and cause serious injury. It is believed that the maximum pressure to which these bottles are subject in the field is usually 120 psi. Therefore, one of the key concerns of the bottle manufacturer is to avoid any failures below 120 psi (it should be noted that this could be the basis of an operational definition of reliability). As is evident in Fig. 3, the new design does not meet specifications. In this sense, the old design is better since it only has one bottle failing below 120 psi (from cavity 18).

The conclusion from the previous analyses is that the reliability problem of the new design is caused by early mortality. Recall that the manufacturer wants to adopt the new design due to the potential savings in downtime of the machines. The key question from the business point of view is whether work on this design should be pursued further. As mentioned before, the early mortality seems to be caused by a distinct failure mode (or modes). Possibilities include unfamiliarity with the new mold or inadequacies in the blow machines' air flow to the molds (since there were no differences between cavities, the likelihood of individual cavities causing the problem is remote). In any case, discovering and removing the causes of the problem requires



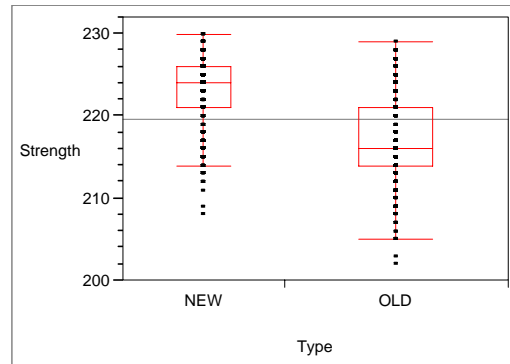


Figure 6. Dot plots and quartiles of the bursting strengths of the two designs removing the observations below 200 psi.

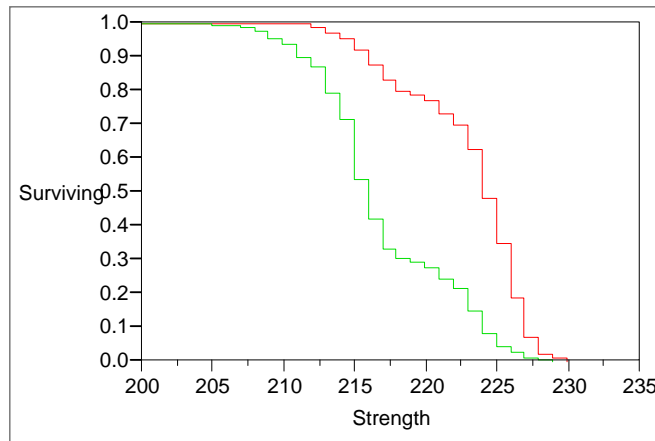


Figure 7. Product-limit survival fit (Kaplan-Meier) to the two designs after removing the observations below 200 psi. New is above the old.

a concerted effort from the “owners” of the process.

In today’s competitive business environment, requests for resources to improve a process must be supported with assessments of the potential gains. To investigate the potential impact of eliminating early mortality in the new design, the analysis was carried out removing the observations below 200 psi. This pressure is where the new design has a clear gap. Setting higher standards of quality can prevent consumer injuries; and thus, needless lawsuits.

After 200 psi the failure mode causing the early mortality appears to have vanished (i.e. the weaker mixture is removed). Fig. 6 displays box plots and Fig. 7 the Kaplan-Meier estimators for the two designs after the removal of the observations. All the tests to compare the designs were highly significant ( $p < 0.0001$ ) indicating the superiority of the new design. Fig. 7 and Table 2 clearly show that now the new design dominates the old design with respect to all definitions of reliability. Such improvement goals would result in the new design being clearly superior with a mean of 223 psi versus 217 psi for the old design.

Table 2: Percentiles and means for each design after removal of observations below 200 psi.

Design	25th percentile	50th percentile (median)	75th percentile	Mean
NEW	221	224	226	223.10
OLD	214	216	221	216.42

#### 4. Summary and conclusions

Before the two designs could be compared in a meaningful manner, the data were analyzed to determine if there was a consistent effect for each design across the different cavities. Graphical and analytical techniques found no significant differences in the cavities within designs. Thus the data

for the cavities were pooled to compare the reliability of the two designs.

Using simple graphical techniques, it was found that the reliability of the new design was likely a mixture of distributions possibly due to different failure modes. This prevented the use of standard parametric techniques. The use of parametric methods when the assumptions, in which they are based, do not hold can easily lead to misleading conclusions. In the pursuit of parsimony, parametric techniques to analyze mixtures were not used in this paper (see, for example, Meeker and Escobar, 1998). Our emphasis was on the nonparametric analysis.

The fact that the reliability functions “crossed” produced seemingly contradictory results (means and medians reversing order). The need for an operational definition of reliability was emphasized. It is easy to find examples of designs that can be more or less reliable than others depending of the definition of reliability considered (Guess, Walker, and Gallant, 1992, and Berger, Boos, and Guess, 1988). The new design is clearly inadequate using 120 psi as the lower specification limit. However, with the early mortality eliminated, its reliability will be uniformly better than that of the old design. The most important improvement would be on the 25th percentile, which increases from 160 psi to 221 psi. This could be the crucial motivation to commit resources on improving the performance of the new design and for prevention of consumer injuries.

### **Acknowledgements**

The authors are grateful for research funding from the University of Tennessee College of Business Administration Scholarly Research Grant Program and the Department of Statistics. Dr. Walker, also, appreciates support from the University of Tennessee’s Center for Executive Education for writing this paper. We both thank retired Oak Ridge National Laboratory Senior Engineer Ned Clapp and Dr. Dennis Lin of Pennsylvania State University for thoughtful reading and providing helpful discussions/comments on this paper.

### **Appendix**

Data for the old and new design by cavity number are available from the following URL: <http://mpd.pagras.net/mtchao/JDataScience/JDS113-data.doc>

## References

- Barlow, R. E. and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing: Probability Models*. To Begin With, Silver Spring, MD.
- Berger, R., Boos, D. and Guess, F. (1988). Tests and confidence sets for comparing two mean residual life functions. *Biometrics* **44**, 103-115.
- Guess, F. and Proschan, F. (1988). Mean residual life: theory and applications. In *Handbook of Statistics: Quality Control and Reliability*. (Edited by P. R. Krishnaiah and C. R. Rao) **7**, 215-224. North Holland, Amsterdam.
- Guess, F., Walker, E., and Gallant, D. (1992). Burn-in to improve which measure of reliability? *Microelectronics and Reliability* **32**, 759-762.
- Hollander, M. and Wolfe, D. (1999). *Nonparametric Statistical Methods*, 2nd Ed. Wiley & Sons, New York, NY.
- Jensen, F. and Petersen, N. E. (1982). *Burn-In: An Engineering Approach to the Design and Analysis of Burn-in Procedures*. Wiley & Sons, New York, NY.
- Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley & Sons, New York, NY.
- Sall, J., Lehman, A. and Creighton, L. (2001). *JMP Start Statistics*, 2nd. Ed. Duxbury, Pacific Grove, CA.

Received November 15, 2001; accepted March 13, 2002

Esteban Walker  
Department of Statistics  
University of Tennessee  
Knoxville, TN 37996-0532, USA  
ewalker@utk.edu

Frank Guess  
Department of Statistics  
University of Tennessee  
Knoxville, TN 37996-0532, USA  
fguess@utk.edu