ENDOCRINE PRACTICE Rapid Electronic Article in Press
Rapid Electronic Articles in Press are preprinted manuscripts that have been reviewed and accepted for publication, but have yet to be edited, typeset and finalized. This version of the manuscript will be replaced with the final, published version after it has been published in the print edition of the journal. The final, published version may differ from this proof.
DOI:10.4158/EP13186.OR

Original Article EP13186.OR

STATISTICAL TRANSFORMATION AND THE INTERPRETATION OF INPATIENT GLUCOSE CONTROL DATA

Running title: Inpatient Glucose Control

George E. Saulnier, MS[1]*; Janna C. Castro, BS*[1]*; Curtiss B. Cook, MD*[2,3]

From the [1]Department of Information Technology, Mayo Clinic Hospital, Phoenix, Arizona, [2]Division of Endocrinology, and [3]Division of Preventive, Occupational, and Aerospace Medicine, Mayo Clinic, Scottsdale, Arizona.
Address correspondence to Curtiss B. Cook, MD, Division of Endocrinology, Mayo Clinic, 13400 E Shea Blvd, Scottsdale, AZ 85259
E-mail: cook.curtiss@mayo.edu
Publisher: To expedite proof approval, send proof via e-mail to scipubs@mayo.edu.

DOI:10.4158/EP13186.OR
© 2013 AACE.

ABSTRACT

Objective: To introduce a statistical method of assessing hospital-based non–intensive care unit (non-ICU) inpatient glucose control.

Methods: Point-of-care blood glucose (POC-BG) data from non-ICU hospital units was extracted for January 1 through December 31, 2011. Glucose data distribution was examined before and after Box-Cox transformations and compared to normality. Different subsets of data were used to establish upper and lower control limits, and exponentially weighted moving average (EWMA) control charts were constructed from June, July, and October data as examples to determine if out-of-control events were identified differently in nontransformed vs transformed data.

***Results*:** A total of 36,381 POC-BG values were analyzed. In all 3 monthly test samples, glucose distributions in nontransformed data were skewed but approached normal distribution once transformed. Interpretation of out-of-control events from EWMA control chart analysis also revealed differences. In the June test data, an out-of-control process was identified at sample 53 with nontransformed data, while the transformed data remained in control for the duration of the observed period. Analysis of July data demonstrated an out-of-control process sooner in the transformed data (sample 55) than the nontransformed (sample 111), while for October, transformed data remained in control longer than nontransformed data.

Conclusion: Statistical transformations increase the normal behavior of inpatient non-ICU glycemic data sets. The decision to transform glucose data could influence the interpretation and conclusions about the status of inpatient glycemic control. Further study is required to determine whether transformed vs nontransformed data influence point-of-care decisions or evaluation of interventions.

Keywords: diabetes; hospital; hyperglycemia; inpatient; statistical process control

Abbreviations

EWMA = exponentially weighted moving average; **ICU** = intensive care unit; **POC-BG** = point-of-care blood glucose

INTRODUCTION

Data on glucose levels in a large sample of US hospitals are now available that can serve as a benchmarl achieved, there is lack of consensus concerning how data should be collected, analyzed, and reported (2). Previous publications have typically focused on evaluating and comparing the relationship of different inpatient glucose measures (eg, mean glucose, patient-day weighted mean glucose, hyperglycemic index) with mortality or some other specific outcome (2).

What is lacking in the literature in the discussion of inpatient hyperglycemia is how to assess the impact of interventions on glucose control or how to monitor if glucose levels remain in control over time. Specifically, given the desire to measure the effectiveness of interventions targeting management of hyperglycemia in the hospital, how do we make better decisions on care from a statistical standpoint? Is it possible, with the introduction of statistical techniques, to accept or reject the effectiveness of interventions sooner? To evaluate changes over time, mean glucose levels would be the simplest metric to track, but given the extreme values that can occur in the hospital, data may be skewed. Moreover, given the large number of measurements that occur in the hospital, small changes may be statistically but not necessarily clinically relevant (3).

Control charting has been suggested as 1 method of assessing the impact of interventions targeting inpatient hyperglycemia, but little data exist on the use of this approach when it is applied to hospital-based glucose measurements (4). Statistical process control charts can help to determine whether changes in care processes are having a real impact on outcomes (5). In the instance of inpatient diabetes management, variations in care practices that have an undesired impact on glucose control could be identified and rectified by examining events that exceed

upper or lower limits of statistical control. A fundamental assumption of process control, however, is that the data being charted reasonably resemble a normal distribution (6,7). Inappropriately assuming the nature of the distribution can lead to false alarms and incorrect conclusions (8). In health care, these may impede patient care and risk safety.

This paper introduces a statistical method of identifying changes in inpatient glucose control via control charts. The method first transforms large glucose data sets to maximize normal distribution–based behavior and then compares control charts before and after transformation. The impact of the transformation on conclusions that may be derived from analyzing control chart data is discussed.

METHODS

Data Extraction

Inpatient point-of-care blood glucose (POC-BG) data on patients not requiring a stay in the intensive care unit (non-ICU) were extracted from the laboratory information system for the period spanning January 1 through December 31, 2011. POC-BG measurements were conducted on a standardized point-of-care testing device (Roche Accu-Chek Inform) that allows scanning of patient identification with direct download into the laboratory information system. All repeat measurements that might have been performed for the same patient within 60 minutes of each other to confirm a high or low value were removed. A total of 36,381 glucose values were extracted for the period. This analysis did not involve any patient identifiers and was part of overall quality improvement/quality assurance efforts on inpatient glycemic control at the authors' institution, and the Mayo Clinic Institutional Review Board determined that formal review and approval were not needed.

Data Transformation

Statistical procedures for process control commonly assume data sets are approximated by a normal distribution (9). Mathematical transformations enable skewed data sets to approach the behavior of normalized data (10). Research indicates that distributions of outpatient glucose data

exhibit bimodal characteristics (11). It is unclear, however, whether the incorporation of multimodal distribution characteristics into the study of inpatient glucose management improves decision making. To facilitate the potential adoption of an already complex model into the clinical setting, the authors elected to strike a balance between accounting for further distribution characteristics and ease of use. In all scenarios presented, therefore, unimodal distributions were assumed. Expanding this assumption to include bimodality may be an area of future study.

The Box-Cox transformation, applied in our analysis of non-ICU POC-BG data, is a family of transformations that optimally normalizes a variable of interest and eliminates the need to randomly apply different approaches. It represents a potential "best practice" for normalizing data (12,13). Transforming the data enables the application of common analytical techniques that are designed to leverage the characteristics of normal distributions. These techniques include statistical process control (6,14). In this effort, control charting was applied to both raw (nontransformed) and transformed POC-BG data, with comparisons in observations then made between the 2 analyses.

An iterative mathematical process (Figure 1) was used to identify the Box-Cox transformation parameter lambda (λ) that represented the closest fit of a raw data set to a normal distribution, based on a χ^2 goodness-of-fit test (15).

Mathematically, the transformation, T(y), is represented by the following equation:

$$T(y) = (y^{\lambda} - 1)/\lambda, \text{ for } \lambda <> 0$$

$$T(y) = \ln(y), \text{ for } \lambda = 0$$

where y = response variable, and λ = transformation parameter (13).

Goodness-of-fit tests generally involve calculating test statistics and rejecting a null hypothesis that the data conform to a specified distribution if the statistic exceeds a predetermined critical value (15,16). As the goodness-of-fit test required the collection of data points into samples of 5 or more, individual observations were grouped (17). For the purposes of this study, we arbitrarily

chose a sample size of 10. The data were grouped without knowledge of the clinical status of the patient.

Control Charts

An exponentially weighted moving average (EWMA) control chart was constructed to assess the performance of glucose values over time. Unlike traditional control charts that require sophisticated rules to identify significant events, EWMA charts highlight out-of-control processes as soon as plotted points exceed a control limit (18,19). Different 3-month subsets of data were used to calculate and establish upper and lower control limits before proceeding with the EWMA chart. From baseline data, upper and lower control limits were established to maintain glucose levels in a band that extended above and below a baseline average by 3 times the standard deviation, commonly referred to as the 3-sigma limit (6). Once the baseline was established, example charts were constructed in arbitrarily chosen 1-month intervals. Graphs included in this analysis specifically highlight portions of the 1-month intervals where differences were observed. For this report, we highlight 3 examples from different months illustrating how data transformations may impact the recognition of out-of-control events identified by control charts.

RESULTS

Example 1

In the first example, taken from analysis of June 2011 data, nontransformed glucose data yielded exceptionally high χ^2 results, suggesting poor normal distribution characteristics (Figure 2, top panel). The high χ^2 values and the left-skewed probability plots of the original (nontransformed) data supported a poor fit to a normal distribution. Lambda values between –2.0 and +2.0 were iteratively applied to the data, with the goal of minimizing the test statistic. After transformation, the same data yielded a significantly lower χ^2 result. The greater symmetry evident in the probability plot also supports a closer normal fit. While the transformations with

the lowest χ^2 results failed to meet a 95% confidence critical value, the transformations nonetheless improved the normal behavior characteristics of the various data sets.

An EWMA control chart (Figure 2, bottom panel) was then constructed. Before and after transformations were superimposed to highlight differences. With nontransformed data, an out-of-control process was identified at sample 53. The transformed data set, however, while approaching the lower 3-sigma limit at sample 53, remained in control for the duration of the observed period. Thus, in this example, the extra step of transformation of the glucose data indicated that the care processes in place resulted in glucose remaining within control limits throughout the entire period.

Example 2

Figure 3 illustrates another example of the impact of a statistical transformation on the analysis of inpatient glucose data from July 2011. As with the first example, nontransformed data (Figure 3, top panel) were skewed, yielding an excessively high χ^2 statistic. After transformation, the χ^2 statistic improved, and the resulting distribution was closer to normal.

In this EWMA control chart example, for the purpose of graphic representation and since no out-of-control hypoglycemic events were noted, the lower statistical limit was removed. When assessing nontransformed vs transformed glucose data, the transformed data hit an out-of-control limit at sample 55, while the nontransformed data did not (Figure 3, bottom panel). While both the nontransformed and transformed data exceeded the upper control limit by sample 111, only the nontransformed data reached the upper clinical limit of 180 mg/dL (20). This was over 50 samples (and with groupings of 10 observations per sample, ie, 500 observations) after the transformed data at sample 55. Thus, the transformed data detected an out-of-control glucose process sooner than did the nontransformed data.

Example 3

In example 3, once again, the skewed nature of glucose values is observed, brought closer to normal after transformation (Figure 4, top panel). The EWMA control chart (Figure 4, bottom

panel) was constructed to assess the performance of October glucose values. The EWMA control chart revealed differences between nontransformed and transformed results. In this example, the nontransformed process exceeded the lower statistical control limit at sample 118. However, the transformed data remained in control longer—until sample 127. Unlike the nontransformed data, when the statistical control limit was eventually exceeded, it surpassed the upper level rather than the lower.

DISCUSSION

Development of statistical methods to assess inpatient hyperglycemia are needed to provide researchers, clinicians, hospital administrators, and quality improvement organizations meaningful data on the impact of care processes designed to monitor glucose control. In our study, we applied the Box-Cox family of transformations to establish our statistical methods. Nonetheless**,** a variety of statistical approaches should ultimately be explored and compared. This would establish methods that provide the best representation of either longitudinal changes in glucose control or the magnitude to which glucose control is influenced by changes in care. Control charts are one quality improvement tool that can be used to evaluate inpatient hyperglycemia care, and this analysis introduced one method of identifying out-of-control inpatient glucose events in a statistical manner for data derived from inpatients who were not critically ill.

The analysis did confirm the nonnormal distribution of non-ICU inpatient POC-BG data. Box-Cox transformations did result in a closer approximation of distributions toward normality. The skewness of nontransformed glucose data will need to be taken into account when choosing the best statistical method to evaluate differences in glycemic control across time or when comparing results of quality improvement efforts targeting inpatient hyperglycemia.

Clinically recommended glucose limits of 70 to 180 mg/dL are much broader than the relatively tight bounds established through the statistical analysis of these data (20). The purpose of a control chart is to identify an out-of-control process, preferably as soon as possible, and

represents another statistical tool to evaluate inpatient glucose control. It will continue to be of interest to institutions to report absolute frequencies of hypoglycemic and hyperglycemic events compared to national guidelines. However, hospitals may find it more important to establish if, or when, a previously controlled glucose management process became out-of-control, rather than just knowing whether a set of values meet a national standard. In this analysis, control charting utilizing transformed data identified out-of-control events on the upper rather than the lower control limits, which suggests that hyperglycemia, rather than hypoglycemia, was the out-of-control process in these data. This observation is consistent with previous analyses both from this institution and nationally suggesting that hypoglycemic events were uncommon with respect to hyperglycemic events (1,21).

As illustrated in example 1, transformed data enabled a seemingly out-of-control process to remain in control throughout the observation period. In this example, nontransformed data could have triggered a false alarm, resulting in an unnecessary change in glucose control processes of care. Differences based on data transformations were also demonstrated in example 3, when the nontransformed data revealed a process that led to glucose levels exceeding the lower statistical control limit, but transformed data showed that the process maintained glucose in control longer—until sample 127. Both examples 1 and 3 suggest that transformations, by enabling processes to remain in statistical control longer, may reduce false alarms associated with naturally skewed data. Example 3 also illustrates how the alarms may change—from out-of-control at a lower limit to a higher limit—with transformation. Example 2 illustrates how transformed data can identify an out-of-control process sooner. Earlier identification of an out-of-control process can lead to less delay in making corrective changes in care.

The main limitation of this analysis is that it is based on retrospective data. The significance of identifying out-of-control events at different times in nontransformed vs transformed data—in terms of whether care would have been altered sooner rather than later—cannot be determined here. What would be of interest is whether a similar transformation in real

time would alter the conclusions of an actual intervention in the hospital targeting inpatient hyperglycemia. The data were limited to non-ICU glucose measurements, and a comparable analysis is needed for data derived from critically ill patients. The impact on care management of glucose data sets exhibiting potentially bimodal distribution characteristics might be of future interest. Finally, this analysis comprised pooled data from all the inpatient areas (excluding critical care). Once refined, this analytic approach could be developed to assess process control based on specific geographic areas within the hospital or by the specialty service managing the patient.

The analysis did demonstrate that statistical transformations increase the normal behavior of inpatient glycemic data sets, at least the ones derived from non-ICU patients. Statistically, results with transformed data were more robust as the tools used to obtain these results are based on an underlying presupposition of normality. The decision to analyze transformed or nontransformed glucose data could influence the interpretation and conclusions about the status of inpatient glycemic control or the impact of interventions. Further study is required to determine whether transformed vs nontransformed data influence point-of-care decisions or evaluation of interventions.

REFERENCES

1. **Swanson CM, Potter DJ, Kongable GL, Cook CB.** Update on inpatient glycemic control in hospitals in the United States. *Endocr Pract*. 2011;17:853-861.

2**. Cook CB, Wellik KE, Kongable GL, Shu J.** Assessing inpatient glycemic control: what are the next steps? *J Diabetes Sci Technol.* 2012;6:421-427.

3. **Bersoux S, Cook CB, Kongable GL, Shu J.** Trends in glycemic control over a 2-year period in 126 US hospitals. *J Hosp Med.* 2013;8:121-125.

4.**Schnipper JL, Magee M, Larsen K, Inzucchi SE, Maynard G; Society of Hospital Medicine Glycemic Control Task Force.** Society of Hospital Medicine Glycemic Control Task Force summary: practical recommendations for assessing the impact of glycemic control efforts. *J Hosp Med.* 2008;3(5 Suppl):66-75.

5. **Benneyan JC, Lloyd RC, Plsek PE**. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care*. 2003;12:458-464.

6. **Montgomery DC.** *Introduction to Statistical Quality Control*. 2nd ed. New York, NY: Wiley; 1991.

7. **Buthmann A.** Dealing with non-normal data: strategies and tools. 2010 Feb 26.Available at: http://www.isixsigma.com/tools-templates/normality/dealing-non-normal-data-strategies-and-tools/ Accessed for verification August 13, 2013.

8. **Breyfogle FW 3rd**. Non-normal data needs alternate control chart approach. Available at: http://www.isixsigma.com/tools-templates/control-charts/non-normal-data-needs-alternate-control-chart-approach/ Accessed for verification August 13, 2013.

9. **Siegel AF**. *Practical Business Statistics with STATPAD.* Homewood, IL: Irwin; 1990.

10.**Osborne JW**. Notes on the use of data transformations. *Pract Assess Res Eval.* 2002;8. Available at: http://pareonline.net/getvn.asp?v=8&n=6. Accessed for verification August 13, 2013.

11. **Fan J, May SJ, Zhou Y, Barrett-Connor E**. Bimodality of 2-h plasma glucose distributions in whites: the Rancho Bernardo study. *Diabetes Care.* 2005;28:1451-1456.

12. **Osborne JW.** Improving your data transformations: applying the Box-Cox transformation. *Pract Assess Res Eval.* 2010;15. Available at: http://pareonline.net/pdf/v15n12.pdf Accessed for verification August 13, 2013.

13. **National Institute of Standards and Technology/Semiconductor Manufacturing Technology (NIST/SEMATECH).** E-Handbook of statistical methods: Box-Cox normality plot Available at: http://www.itl.nist.gov/div898/handbook/eda/section3/eda336.htm. Accessed for verification August 13, 2013.

14.Sherman PJ. Tips for recognizing and transforming non-normal data [Internet]. 2010 Feb 26 [cited 2013 Jul 11]. Available from: http://www.isixsigma .com/tools-templates/normality/tips-recognizing-and-transforming-non-normal-data/.

15.**National Institute of Standards and Technology/Semiconductor Manufacturing Technology (NIST/SEMATECH).** E-Handbook of statistical methods: chi-square goodness-to-fit test. Available at: http://www.itl.nist.gov/div898/handbook/eda/section3/eda35f.htm. Accessed for verification August 13, 2013.

16. **National Institute of Standards and Technology/Semiconductor Manufacturing Technology (NIST/SEMATECH).** E-Handbook of statistical methods: confidence limits for the mean. Available at: http://www.itl.nist.gov/div898/handbook/eda/section3/eda352.htm. Accessed for verification August 13, 2013.

17. **National Institute of Standards and Technology/Semiconductor Manufacturing Technology (NIST/SEMATECH).** E-Handbook of statistical methods: chi-square goodness-of-fit test. Available at: http://www.itl.nist.gov/div898/handbook/prc/section2/prc211.htm. Accessed for verification August 13, 2013.

18. **National Institute of Standards and Technology/Semiconductor Manufacturing Technology (NIST/SEMATECH).** E-Handbook of statistical methods: example of EWMA control chart for mass calibrations. Available at: http://www.itl.nist.gov/div898/handbook/mpc/section3/mpc3522.htm.

19. *SAS/QC User's Guide*, version 8. Cary, NC: SAS Publishing; 1999.

20. **Umpierrez GE, Hellman R, Korytkowski MT, et al.** Management of hyperglycemia in hospitalized patients in non-critical care setting: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2012;97:16-38.

21.**Cook CB, Castro JC, Schmidt RE, et al.** Diabetes care in hospitalized noncritically ill patients: more evidence for clinical inertia and negative therapeutic momentum. *J Hosp Med.* 2007;2:203-211.

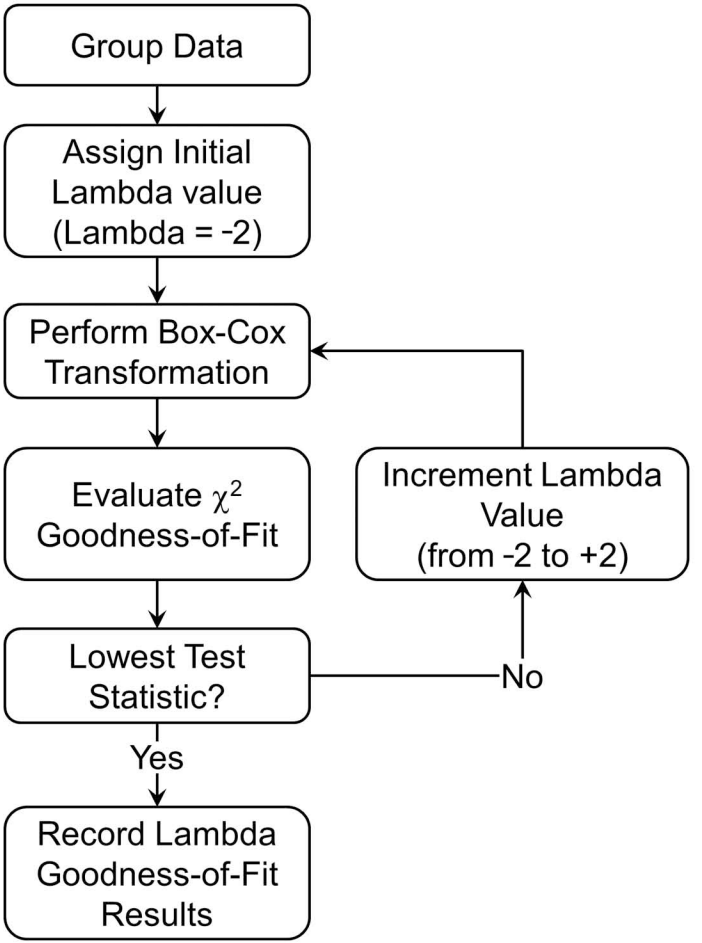
Legends

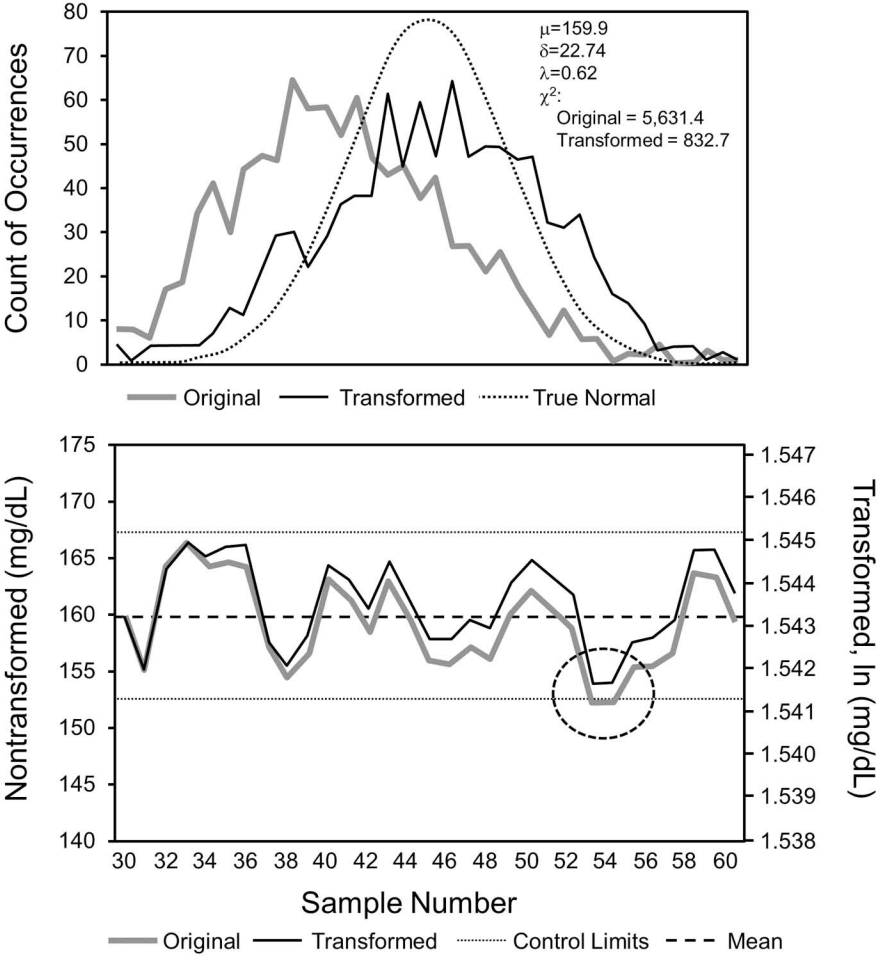
Figure 1. Analytical approach to iteratively performing a Box-Cox transformation. This method resulted in the selection of the transformation parameter lambda that maximized the normal distribution characteristics of the underlying glucose data set.

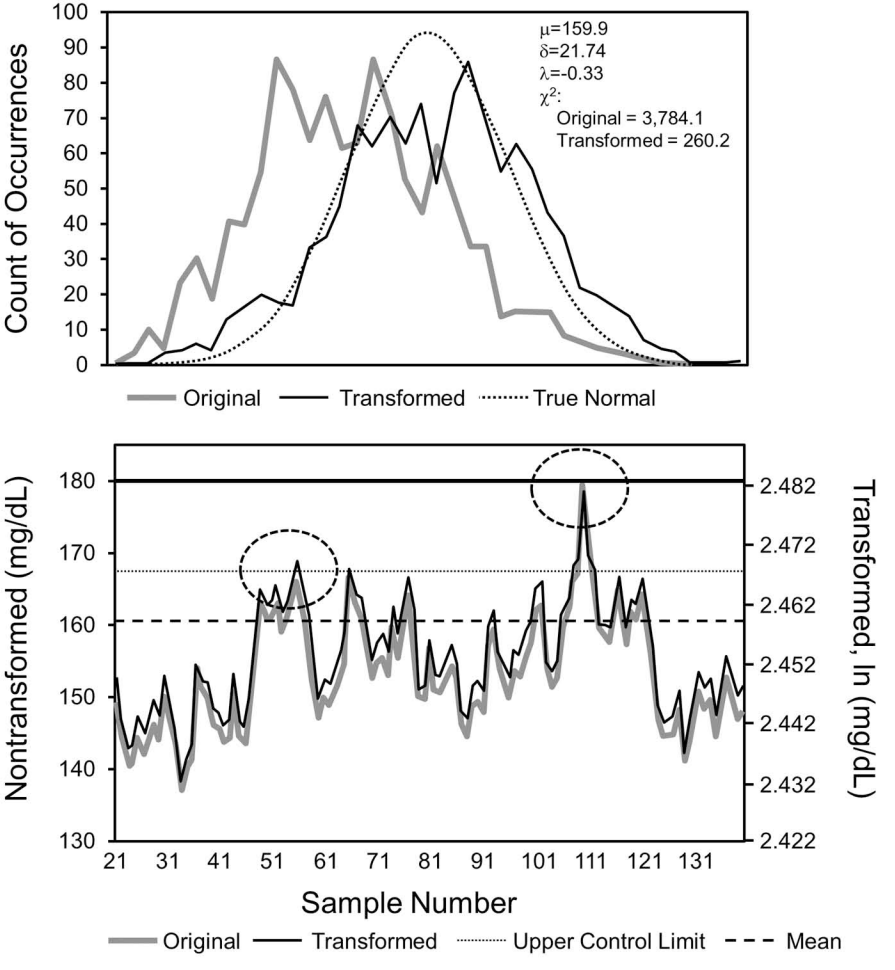
Figure 2. Probability plot (top panel) and control chart (bottom panel) of original (nontransformed) and transformed glucose data sets, from June 2011, used in example 1. The population mean of the nontransformed data set is denoted by μ, and standard deviation, by σ. The decrease in the χ^2 test statistic from 5,631.4 to 832.7 demonstrates an improvement in the normal distribution characteristics of the transformed data. See text for interpretation of control chart findings.

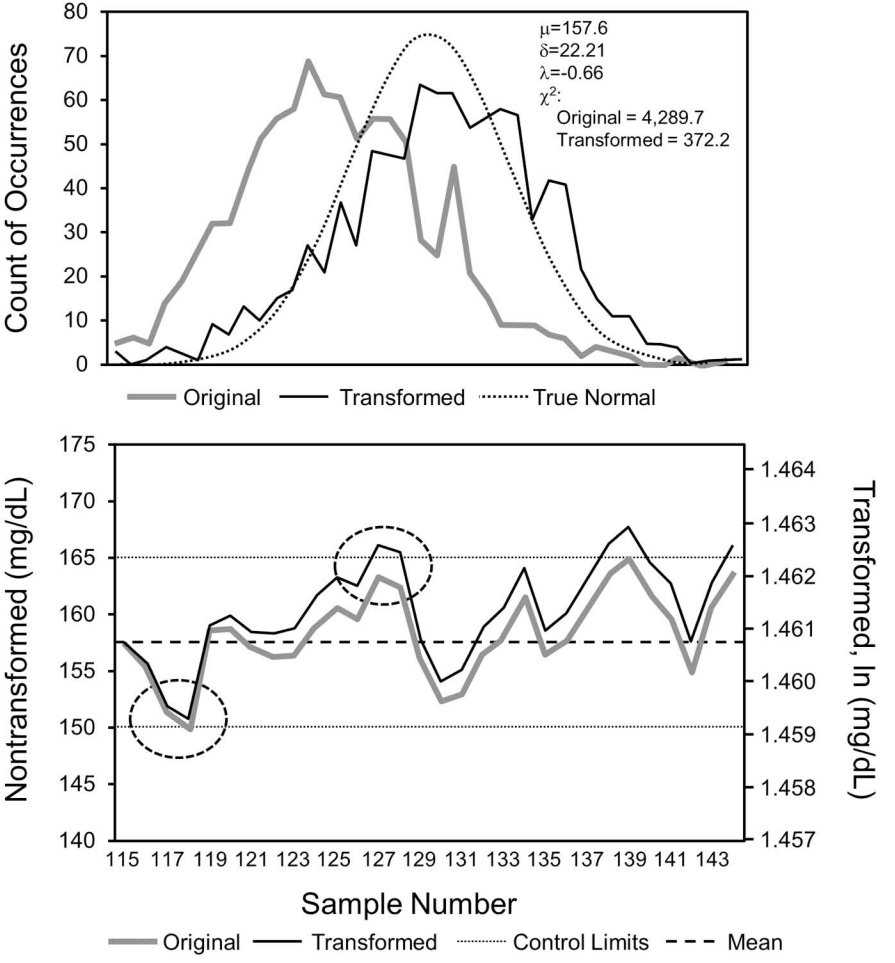
Figure 3. Probability plot (top panel) and control chart (bottom panel) of original (nontransformed) and transformed glucose data sets, from July 2011, used in example 2. The population mean of the nontransformed data set is denoted by μ, and standard deviation, by σ. The decrease in the χ^2 test statistic from 3,784.1 to 260.2 demonstrates an improvement in the normal distribution characteristics of the transformed data. See text for interpretation of control chart findings.

Figure 4. Probability plot (top panel) and control chart (bottom panel) of original (nontransformed) and transformed glucose data sets, from October 2011, used in example 3. The decrease in the χ^2 test statistic from 4,286.7 to 372.2 demonstrates an improvement in the normal distribution characteristics of the transformed data. See text for interpretation of control chart findings.


Group Data
Assign Initial Lambda value (Lambda = -2)
Perform Box-Cox Transformation
Evaluate χ^2 Goodness-of-Fit
Lowest Test Statistic?
No
Increment Lambda Value (from -2 to +2)
Yes
Record Lambda Goodness-of-Fit Results


Count of Occurrences
80
70
60
50
40
30
20
10
0
μ=159.9
δ=22.74
λ=0.62
χ2:
Original = 5,631.4
Transformed = 832.7
Original
Transformed
True Normal
Nontransformed (mg/dL)
175
170
165
160
155
150
145
140
Transformed, ln (mg/dL)
1.547
1.546
1.545
1.544
1.543
1.542
1.541
1.540
1.539
1.538
30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60
Sample Number
Original
Transformed
Control Limits
Mean


Count of Occurrences
100
90
80
70
60
50
40
30
20
10
0
μ=159.9
δ=21.74
λ=-0.33
χ²:
Original = 3,784.1
Transformed = 260.2
Original
Transformed
True Normal
Nontransformed (mg/dL)
180
170
160
150
140
130
Transformed, ln (mg/dL)
2.482
2.472
2.462
2.452
2.442
2.432
2.422
21
31
41
51
61
71
81
91
101
111
121
131
Sample Number
Original
Transformed
Upper Control Limit
Mean


μ=157.6
δ=22.21
λ=-0.66
χ²:
Original = 4,289.7
Transformed = 372.2
Count of Occurrences
Original
Transformed
True Normal
Nontransformed (mg/dL)
Transformed, ln (mg/dL)
Sample Number
Original
Transformed
Control Limits
Mean