Cleveland Clinic
Section of Biostatistics
Quantitative Health Sciences/ JJN-3
9500 Euclid Ave, Cleveland, OH 44195
QHSrequest@ccf.org

Date: 7 April 2017

To: Sandra Siedlecki
Senior Nurse Scientist, Office of Research & Innovation CCF

From: Robert S. Butler
Senior Biostatistician, Quantitative Health Sciences CCF

Re: Assessment of Accuracy, Precision, Equivalence, and Agreement of TempTraq® compared to Pulmonary Artery Catheter for Monitoring Temperature in Adults in ICU

Summary:

The TOST equivalence test indicates TempTraq® is not equivalent to Pulmonary Artery Catheter. The Bland-Altman test for agreement indicates TempTraq® is in agreement with Pulmonary Artery Catheter and there is an overall bias of .3 between the two measurement systems. There is no significant difference in the precision of the two methods. There is a significant difference in accuracy. The accuracy analysis confirms the significance of the bias observed in the Bland-Altman test. If the bias value of .3 is added to each TempTraq® measure a re-analysis indicates the two measures are equivalent, in agreement, and exhibit no significant difference in precision or accuracy.

Data and Analytical Methods for Agreement and Equivalence:

The test for equivalence for TempTraq® and Pulmonary Artery Catheter is the Schuirmann's TOST (Two One Sided T-Test) method. Since this is a two sided test alpha was set to .025. The upper and lower bounds for equivalence could not exceed ±.2.

The Bland-Altman test was used to test for agreement. The test consisted of two parts: a check for bias (constant offset) between the two methods and a test for significant trending in the differences between the results for the two tests using a kernel smoother (LOWESS regression). The line of the kernel smoother remains close to the mean of the differences (-.3) indicating there are no systemic differences between TempTraq® and PA Catheter measurement methods.

Paired temperature measurements for both methods were taken simultaneously every 2 hours for an 8 hour period for each of 60 patients. The demographic data gathered for each patient was gender, race, and measurement side. Before testing for equivalence the data was checked for possible significant correlation between these demographic measures and time and the differences between temperature measurements (TempTraq® – PA Catheter) within each patient using repeated measures mixed models. The only statistically significant correlation ($P < .05$) was time and that significance was driven by the mean of the differences for the first 2 hour period when contrasted with the other time intervals.

Analysis:

Both the equivalence and the agreement tests were run using all of the data, all of the data except data from the first time point, and each time point separately. In every instance the result was the same: the two

methods are in agreement but they are not equivalent.

An examination of the Bland-Altman plot for all of the data (Figure 1) suggests the cause of the lack of equivalence might be those points below the lower control limit. However, if these points are removed from the analysis the lack of equivalence remains. Thus, the lack of equivalence is due to the overall bias between the two measures and not to the existence of a few data points outside of the lower 95% confidence interval of the mean of the differences. If the absolute value of the bias calculated using all of the data (.3) is added to each of the TempTraq® measures the Bland-Altman bias between TempTraq® and PA catheter measurements is reduced to near zero and the derived upper and lower bounds for equivalence are within the specified ±.2 equivalence bounds. An examination of the histograms of TempTraq® and PA catheter data for actual data and adjusted data (Figures 2 and 3) illustrates the shift induced in the TempTraq® data and the overlay of the distribution curves.

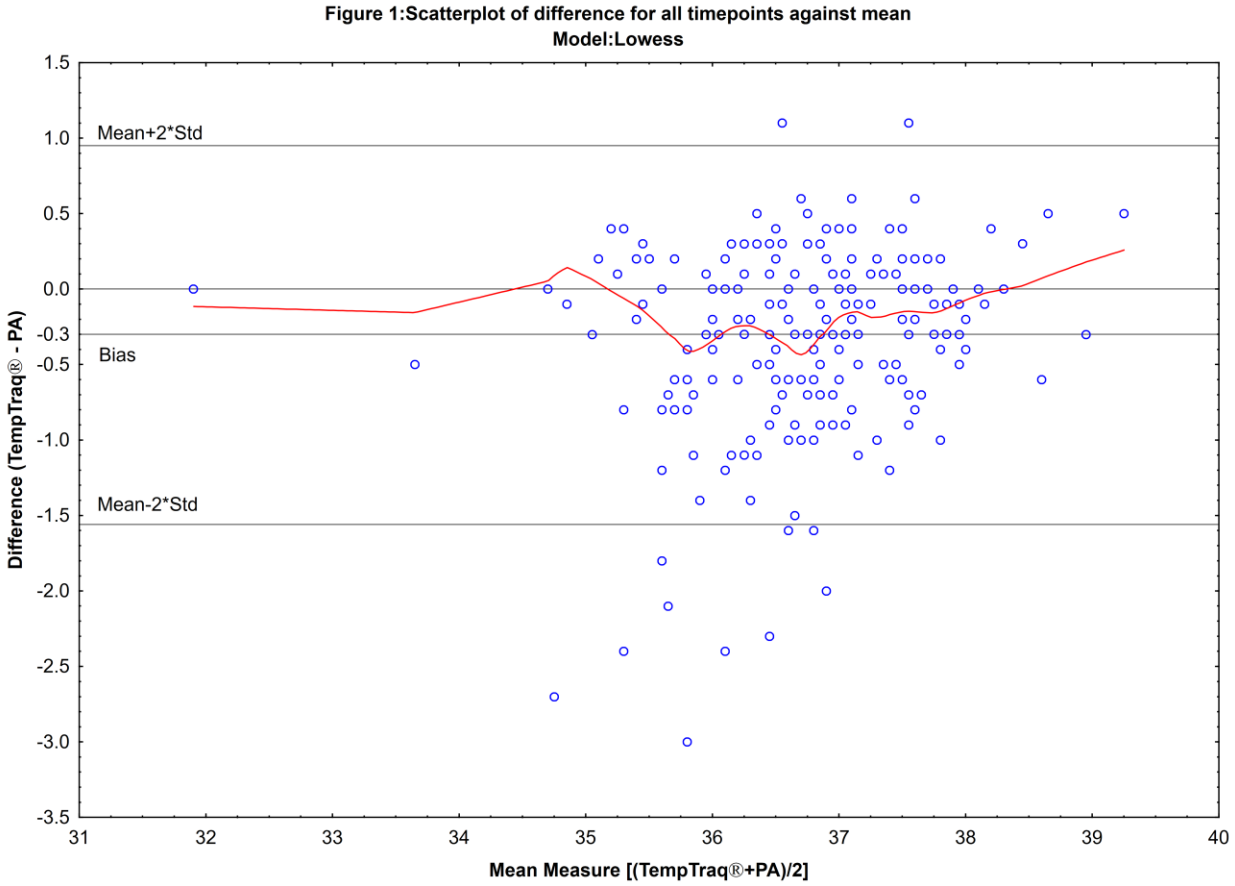

Figure 1 Summary:

The Bland-Altman test indicates there is an offset (bias) in temperature measurements between TempTraq® and Pulmonary Artery Catheter of -.31. An evaluation of the LOWESS smoothed curve (red line) suggests an essentially random path through the bulk of the data points. This indicates there are no systemic differences between the two measurement methods.

Figure 2: Histogram of Actual TempTraq® and Actual Pulmonary Artery Catheter Temperature Distributions

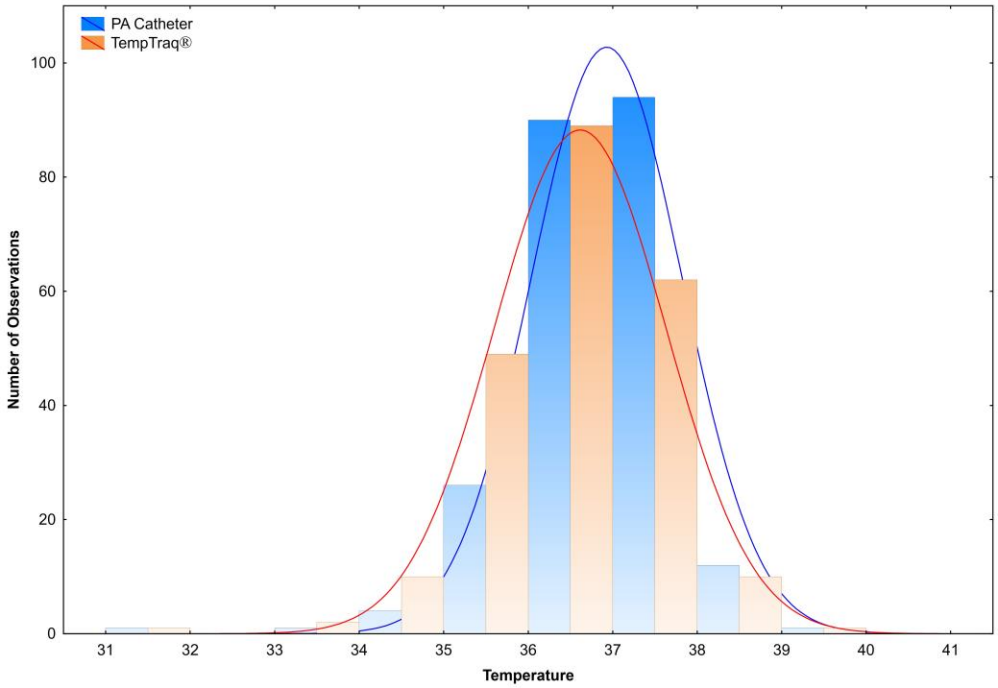

Figure 3: Histogram of Bias Adjusted TempTraq® and Actual Pulmonary Artery Catheter Temperature Distributions

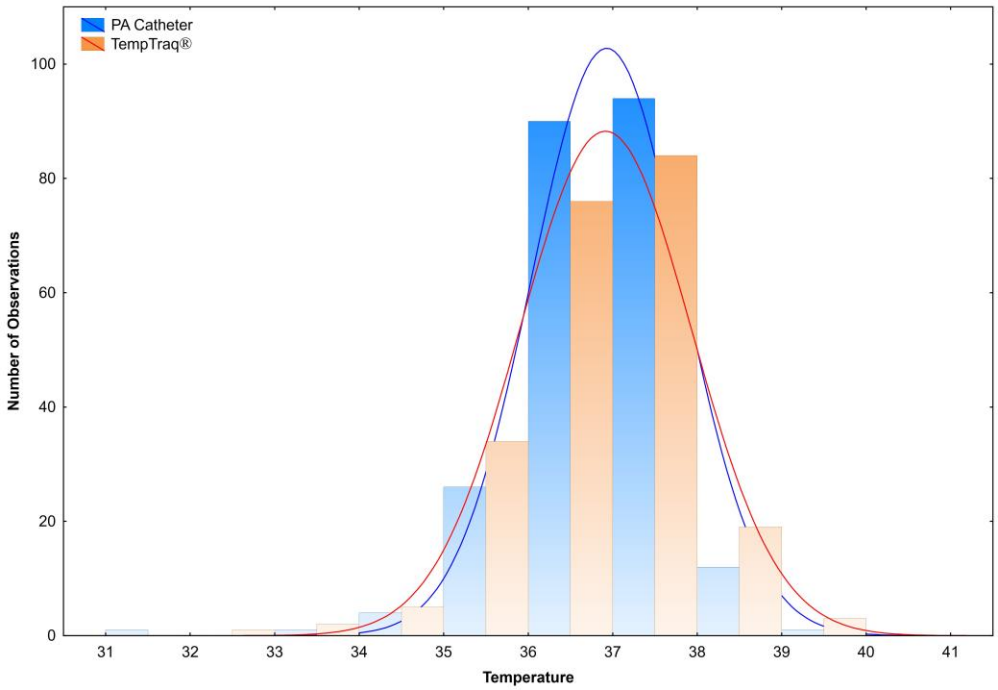

Table 1 is a summary of the bias values from the Bland-Altman analysis and the derived upper and lower bounds from the equivalence test for each of the data set analysis.

Table 1: Agreement and Equivalence Summary			
Data Set	Bland-Altman Bias	Equivalence Bounds – Reference (-.2 , .2)	Equivalent
All Data	-.30	(-.39, -.22)	No
All Data less Time 1	-.20	(-.29, -.11)	No
Time 1	-.60	(-.78, -.42)	No
Time 2	-.24	(-.39, -.097)	No
Time 3	-.16	(-.30, -.03)	No
Time 4	-.19	(-.37, -.02)	No
Bias Adjusted all Data	-.005	(-.09, .08)	Yes

Data and Analytical Methods for Accuracy and Precision:

The temperature measurements using the Pulmonary Artery Catheter are the gold standard against which TempTraq® measures were assessed for accuracy and precision. Accuracy is the measure of the degree of closeness of a test measurement to a gold standard and precision is the degree of variation or consistency about a given measure. Tests of accuracy are tests of differences in means and tests of precision are tests of differences in variation about means.

Accuracy Assessment

Simultaneous TempTraq® and Pulmonary Artery Catheter measurements were taken on 60 patients at four separate time points. Differences between the simultaneous measures of the two methods were computed. If there are no significant differences between the measurements of the two methods the mean of the distribution of the point-by-point differences of two methods should not be significantly different from zero. The distribution of the differences was examined using a paired t-test. The mean of the differences (TempTraq® – Pulmonary Artery Catheter) was -.31. This difference is significantly different from zero ($P<.0001$) and is in accordance with the bias reported in the Bland-Altman summary for agreement.

The measurements were taken at four separate time points on the same individual. Thus the measurements at different points in time, even after differencing, are not independent. To insure that this lack of independence was not impacting the finding reported above the data was split into four groups corresponding to the four time points and each sub-group was analyzed separately. While the magnitudes of the differences between TempTraq® and Pulmonary Artery Catheter measurements variable according to the time grouping the means of the differences remained significantly different from zero and the pooled difference is equal to the difference reported previously.

Precision Assessment

The data for both measures were separately analyzed using repeated measures mixed models. The models controlled for time point measure, patient race, gender, and the location of the measurement site. The residual from these two models represent measures which have been adjusted to eliminate variation due to the four model variables. Thus the variability of the residuals is a measure of the precision of the associated measurement system. An F-test was used to check for significant differences in the precision (variability) of the two measures. The P-value = .20 which indicates the difference in the precision of the two methods is not statistically significant. The variance for the Pulmonary Artery Catheter is .77 and the variance for the TempTraq® measurements is .91.

Table 2: Accuracy and Precision Summary

Accuracy	Mean of Difference	P-value – mean of difference ≠0	Conclusion	Notes
Difference (TempTraq® – PA) Catheter Measurements	-.31	<.0001	Significant bias between measurement and Gold Standard	Difference in accordance with Bland-Altman test for agreement
Precision	Variability (Precision Estimate)	Precision P-value TempTraq® ≠ Pa Catheter	No significant differences between precision estimates	Precision estimates made after controlling for patient demographic measures
TempTraq®	.91	.20		
PA Catheter	.77			

Sensitivity and Specificity

Using the PA Catheter measurements as the gold standard and a cut point of 37.8 for fever/no fever threshold results in TempTraq® sensitivity and specificity estimates of 54.5% and 97.9% respectively.

Conclusions

The analysis indicates TempTraq® measurements are in agreement with Pulmonary Artery Catheter temperature measurements and the precision of the two methods are not significantly different. The analysis also indicates there is a bias of .3 degrees between the two measurement systems. It is because of this bias that the two measurements are not equivalent and do not exhibit the same degree of accuracy. It may be possible to employ engineering methods to modify the TempTraq® system to reduce the bias thus insuring no statistically significant differences in instrument accuracy as well as guaranteeing their equivalence.