# Investigation of near infrared instrument performance during a clinical study of human sera

**James R. Long, Marilyn R. Gatin, Paul W. Schmitt, Paul J. Galley and John F. Price**

*Boehringer Mannheim Corporation, 9115 Hague Road, Indianapolis, IN 46250, USA.*

## Introduction

Near infrared (NIR) spectroscopy has proven to be a highly valuable analysis tool for a broad range of applications. Coupled with the power of chemometric techniques, NIR spectroscopy has flourished in many fields including the food, agricultural, petroleum and pharmaceutical industries. Recently, the use of NIR spectroscopy and chemometrics for making clinical laboratory measurements has attracted a great deal of attention. As a clinical laboratory tool, NIR spectroscopy has several attractive features. The technique is fast, nondestructive, requires minimal sample preparation and requires low sample volumes. Additionally, a single scan can potentially be applied to any number of analyte models, eliminating the need for acquiring a new measurement if additional analyte predictions were desired.

Given the complex nature of biological samples (serum), the use of primary standards as a means for calibration is not feasible. Without primary standards, serum samples must adequately represent the variations and matrix effects one would expect to find in a normal population. Since it is impractical to collect a calibration set for every new set of samples, robust models are needed which can stand the test of time. This means that prediction models must be insensitive to matrix variations and instrument changes.

The purpose of the work described here was to examine instrument stability and model robustness over time. This paper will focus on our findings concerning instrumental changes encountered during two studies. Even though several analytes were examined, only glucose will be discussed in this paper. Descriptions of the results for the other analytes are given in the paper by Gatin *et al.*[1]

## Experimental

Two studies were held six months apart to investigate the potential impact of instrument variations with time. The identical nature of the two studies allowed for intra- and inter-study comparisons of model performance. Cross-validated prediction results for each study were generated. A predictive model was created by each study and used to predict the other study.

A total of 259 serum samples (139 samples from Study A and 120 samples from Study B) were measured on an NIRSystems Model 6500 (Perstorp Analytical Inc.) which has an available range of 400–2500 nm. For this study, the detector gain was optimized for the upper wavelength regions causing saturation in the lower region of the spectrum measured by the lead sulfide detector (1100–2500 nm). Thus, the usable data collected ranged from approximately 1300–2500 nm

From Near Infrared Spectroscopy: The Future Waves

at 2 nm resolution (10 nm bandpass). Spectra were collected in the transmittance mode using a temperature-controlled (37ºC) 1 mm path length flow cell. A solution of 0.9% saline was used for a reference. Each sample measurement was the result of 100 co-added scans. Clinical reference data was collected on a BM/Hitachi 717 Analyzer (Boehringer Mannheim Corporation). Data preprocessing included *n*th order derivatives, smoothing, and autoscaling. Preprocessing, PLS and MD were implemented using Matlab (The Math Works Inc.).

## Results

Data from each study were initially treated independently. Each set was subjected to outlier detection with resulting subsets yielding reasonable PLS cross-validation results (Table 1) for glucose. Study A had a standard error of validation (*SEV*) of 26 mg $dL^{-1}$. *SEV* for Study B was slightly higher at 35 mg $dL^{-1}$. A model was then created from Study B data and used to predict the Study A set. The Study B model did a fair job of predicting Study A with an *SEP* equal to 32 mg $dL^{-1}$. Study A was then used to create a model to predict Study B. From Table 1 it can be seen that the prediction performance degraded (*SEP* = 44 mg $dL^{-1}$).

Figure 1 contains the unprocessed serum spectra from Study B. Visual inspection of the data did not reveal any obvious clues as to the cause of the poor prediction performance. The Mahalanobis distances for each data set were then calculated with respect to the opposite set. When Study A samples were screened for outliers based on a Study B model, 115 of the original 139 samples were retained. Figure 2 indicates which samples were retained. Rejected samples are indicated by "holes" in the bar graph.

Cross-validation of the new Study A subset yielded an *SEV* of 30 mg $dL^{-1}$ (Table 2). This reflected a slight decrease in performance over initial Study A results (Table 1). When Study B samples were screened for outliers based on a Study A model, a time-dependent trend became obvious (Figure 3). It can be seen that after about sample number 66, nearly all of the remaining samples were classified as outliers. Inspection of difference calculations from sequential reference scans during the second half of the study revealed a gradual increase in measurement noise. Further diagnostics revealed the problem to be a failing lamp. Sample number 69 was chosen as a cutoff point since it was the last measurement for that day of the study, which left a total of 57 samples for Study B. Cross-validation of the new Study B subset resulted in an improvement over prior results. *SEV* for Study B decreased from 35 mg $dL^{-1}$ to 25 mg $dL^{-1}$ (Table 2).

The prediction of Study A data by the Study B (truncated) model improved slightly (*SEV* = 27 mg $dL^{-1}$). The scatter plot is shown in Figure 4. Likewise, the prediction of Study B (truncated) data by the Study A model also improved decreasing from 44 mg $dL^{-1}$ to 27 mg $dL^{-1}$ (Figure 5).

**Table 1. Statistics for Studies A and B and their cross-predictions for glucose from spectral data (2030–2398 nm).**

| Calibration set | Test set | *SEV* (mg $dL^{-1}$) | *SEP* (mg $dL^{-1}$) | # factors |
|---|---|---|---|---|
| A | A | 26 | — | 13 |
| A | B | — | 44 | 13 |
| B | B | 35 | – | 8 |
| B | A | — | 32 | 8 |

From Near Infrared Spectroscopy: The Future Waves

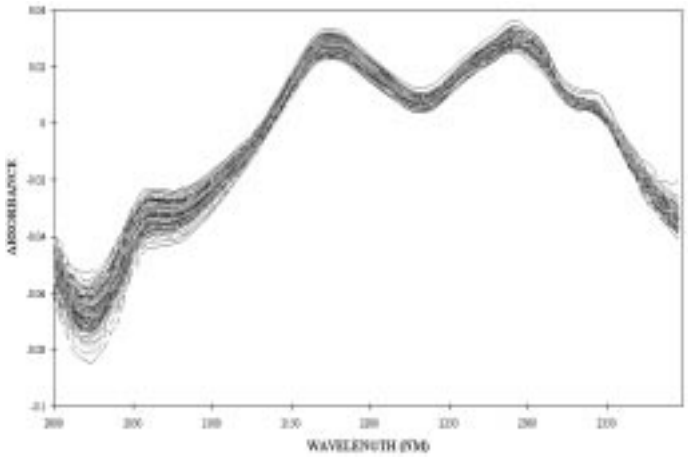


**Figure 1. NIR spectra of serum samples (0.9% saline reference).**

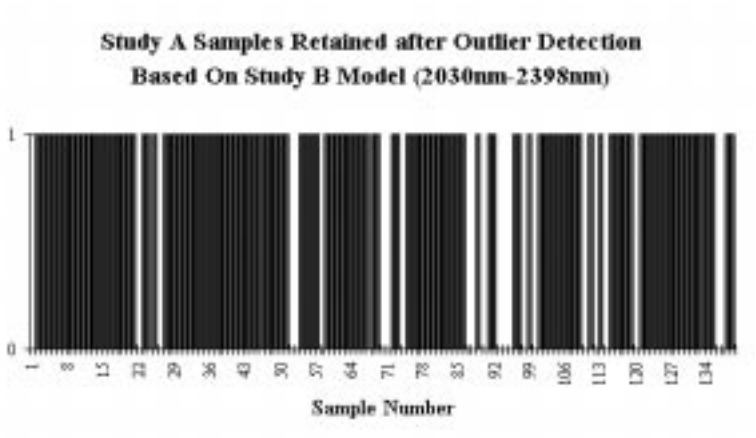


**Figure 2. Mahalanobis distance results for Study A data. Sample inclusion indicated by value equaly to 1 and exclusion indicated by value equal to 0.**

**Table 2. Statistics for Studies A and B and their cross-predictions for glucose for truncated Study B spectral data (2030–2398 nm).**

| Calibration set | Test set | *SEV* (mg $dL^{-1}$) | *SEP* (mg $dL^{-1}$) | # factors |
|---|---|---|---|---|
| A | A | 30 | — | 12 |
| A | B truncated | — | 27 | 9 |
| B truncated | B truncated | 25 | — | 12 |
| B truncated | A | — | 27 | 12 |

From Near Infrared Spectroscopy: The Future Waves

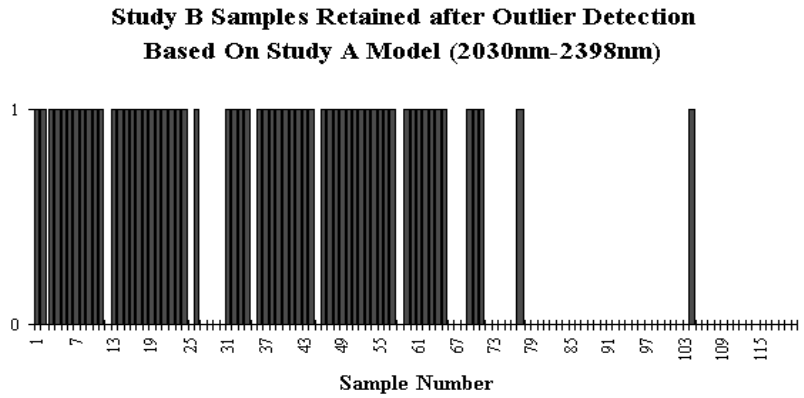


**Figure 3. Mahalanobis distance results for Study B data. Sample inclusion indicated by value equal to 1 and exclusion indicated by value equal to 0.**

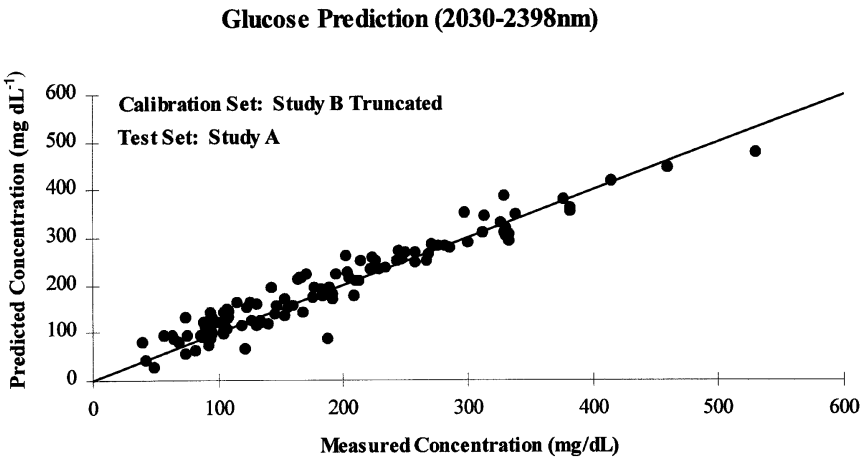


**Figure 4. Scatter plot of PLS prediction of Study A by Study B (truncated) model. *SEP* = 27 mg dL$^{-1}$.**

## Conclusion

NIR spectroscopy has been shown to be a viable tool for making clinical laboratory measurements. Glucose predictions have been presented which have a reasonable level of error. Due to the difficulties associated with collecting a calibration set, it becomes apparent why it is essential to create the most robust prediction model possible. Careful monitoring is needed to insure that analyte ranges within the calibration samples adequately span the population and that instrument reproducibility is maintained. PLS has been shown to be a useful method for making multivariate predictions of glucose from NIR spectral information. Outlier detection has been shown to be

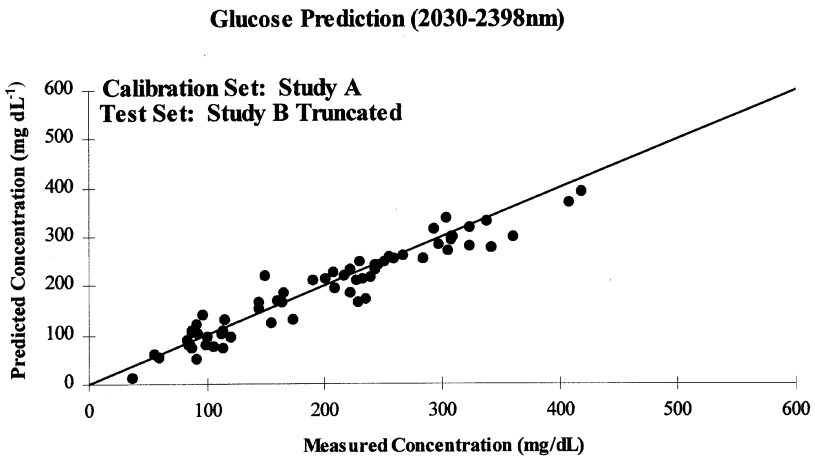


**Figure 5. Scatter plot of PLS prediction of Study B (truncated) by Study A model. *SEP* = 27 mg dL$^{-1}$.**

valuable not only for screening suspect or corrupt samples but as a diagnostic tool for investigating instrument performance as well.

## References

1. Marilyn R. Gatin, James R. Long, Paul W. Schmitt, Paul J. Galley and John F. Price, "Comparison of Clinical Studies: Near-Infrared Predictions of Multiple Analytes in Human Sera", in *Near Infrared Spectroscopy: The Future Waves*, Ed by A.M.C. Davies and P.C. Williams. NIR Publications, Chichester, pp. 347–352 (1996).