Development of Models for Optical Detection of Jaundice

by

Marilyn Hixson
Virgil Anderson

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #78-18

July, 1978

Work supported in part by NIH Grant AM 18871-02.

I. INTRODUCTION

A. Objectives

The long-term goal of this research effort is to develop an advanced noninvasive technique and clinical instrumentation for the optical detection of neonatal jaundice. The principle of the method is to relate spectral reflectance measurements of the skin to the serum bilirubin level (or additional variables) which describe the state of hyperbilirubinemia in an infant. Although the research is a continuing effort, the specific portion of work reported here had a primary objective of establishing a relationship between spectral response and bilirubin level on a large sample population of black and white normal full-term infants and, thus, demonstrating the clinical feasibility of the optical detection of jaundice. The designs of experiments conducted to date and results of statistical analyses are presented, as well as some background information and a brief description of the data acquisition procedures.

B. Background

Bilirubin, in its indirect form, is potentially harmful to the central nervous system of the newborn infant. The severity of damage caused is related to the level of bilirubin in the serum of the blood. In its most severe form, this damage is called kernicterus. Treatment regimens such as exchange transfusions and phototherapy are used to prevent very high levels of bilirubin. Currently, it is felt that lower levels of bilirubin may also be one of the causes for minimal brain dysfunction, a condition thought to be responsible for a large majority of learning disorders in children. If such a relationship is true, early detection and treatment of lower level hyperbilirubinemia becomes even more critical.

The present method for detecting jaundice is visual. The infant's inability to conjugate or excrete sufficient amounts of bilirubin results in excess bilirubin diffusing into the tissues and skin causing a yellowing of the skin. This color change of the skin is due to the characteristic absorption of the bilirubin. A positive diagnosis is verified by a serum bilirubin test using established laboratory techniques. Although these techniques have been shown to be inadequate in certain instances such as in the occasional development of kernicterus in infants with low bilirubin levels (under 10 mg/100 ml), they provide a reasonable indication of a patient's potential for kernicterus in most cases. The disadvantages of the present visual detection and laboratory confirmation process are the following:

- Many lower-level hyperbilirubinemias are missed.
- There is a delay in the initiation of treatment until the laboratory results are known.
- There is discomfort to the infant.
- There are often unnecessary expenses added to the patient's hospital bill.
- The laboratory results may not give an accurate indication of the patient's potential for kernicterus.
- The technique is time consuming and unsuited for mass screening.

In addition to the initial detection process, proper monitoring of bilirubin level during treatment for jaundice is equally important. Improper monitoring can result in excessive or insufficient phototherapy or unintended delay in administering an exchange transfusion. Both initial detection of jaundice and the monitoring of jaundice during therapy are critical in treatment of the disorder.

The process of detecting jaundice in current nursery practice is based upon one vital sign: subtle color change of the infant's skin. The rationale for this study is that the subjective judgment of the pediatrician in recognizing a subtle color change can be replaced by a quantitative method involving measurement of an optical signature which will correlate skin reflectance with serum bilirubin level.

Previous investigations have qualitatively shown that increased serum bilirubin levels cause a decrease of the spectral reflectance in the region near 465 nm. The preliminary research study carried out by Wiechel, DeWitt, and Hannemann in 1975 [3, 5, 6, 7] resulted in a quantitative correlation between serum bilirubin level and selected spectral reflectance values. This type of research needs to be extended in the following ways: (1) a larger population of infants should be sampled for wider inference; (2) the analysis should be expanded to determine an optimum relationship between level of jaundice and skin reflectance; (3) a wider variation of jaundiced infants, particularly those with bilirubin levels above 10 mg/100 ml, should be sampled; (4) the effects due to the reflectance spectra measurement technique (repeatibility and accuracy) need to be evaluated; and (5) other physiologic tests should be performed in addition to the serum bilirubin concentration.

The first two issues have been addressed in this study and statistical analyses have demonstrated that there is a causal, quantitative relationship between serum bilirubin level and reflectance spectra. It is feasible that this noninvasive method can be clinically useful as a mass screening method for initial detection of neonatal jaundice. However, the remaining unresolved points must be studied and scientific evidence developed before this optical method could be recommended for clinical use.

II. DESCRIPTION OF DATA ACQUIRED

A. Clinical Procedures

Clinical observations were made on infants in the nurseries and neonatal intensive care units of Wishard and Riley Hospitals in Indianapolis, Indiana. Spectral reflectance measurements on the infant's skin were made using a reflectometer with the following major components: (1) tungsten light source with power supply; (2) bifurcated fiber optic bundle, custom made to mate with source, probe, and spectroradiometer; (3) spectroradiometer system, consisting of a Wadsworth type spectrometer with flint glass prism using a silicon detector, scan speed range 1 to 32 seconds/scan, and a wavelength range 380 to 800 nm with a spectral resolution typically better than 2%; and (4) data acquisition system consisting of a Hewlett Packard 3960-A four channel FM tape recorder. Two channels of the tape recorder were used for detector level and wavelength position signals; the third channel was used for test/monitor signals (such as indicating whether the detector signal was from a subject, reference standard, etc.); and the fourth channel was used to record audio information such as patient identification, procedures, and other data that might prove useful in later analysis of the clinical observations. A two channel MFE Model M22C strip chart recorder provided the means to monitor proper operation of the system and the sequence of measurement.

An important aspect of the measurement procedure is the data processing of the clinical observations to convert the analog measurements recorded on the tape recorder to digital format suitable for computer analysis. In the first step, the analog data (detector signal) is converted to 10 bit binary data. Wavelength positions of the spectroradiometer are determined from a spectral calibration gauging curve based upon a fourth order polynomial fit

to the calibrating spectral lines. Detector signals from the subject's skin are compared to signals from a barium sulfate reference standard. The resulting relative spectral reflectance is then stored along with identification and clinical observations (such as serum bilirubin and other tests) on magnetic tape.

The procedure used to obtain reflectance spectra on an infant is as follows. The reflectometer system can be set-up, checked out, and calibrated in a nursery environment in 20 minutes. The system is normally located in a separate room and the infants are carefully brought in their mobile cribs to this location. The body site location found to be most reproducible is on the middle part of the back and measurements are taken on both the left and right sides. The reflectometer probe, of diameter 2 cm, contacts the skin causing an area of approximately 1 cm^2 to be irradiated (total radiant power is less than 1 mW). The manner in which the probe contacts the skin is critical. The probe response is dependent upon the pressure applied by the observer. To overcome this, a technique for blanching the skin was used during the clinical observations. Using two fingers and light pressure, the skin was stretched and the probe contacted the region between the fingers. This procedure requires dexterity and experience; whenever practical, only one observer was used to minimize the influence of probe operation on the observed spectra.

Following the reflectance spectra measurement, which generally lasted less than 30 seconds including time to remove the infant's clothing, the infant was returned to the nursery and a blood sample (2 cc) withdrawn. Laboratory tests were performed for total bilirubin (Jendrassik method) and for direct and indirect fractions each time a reflectance observation was made. Total

protein, hematocrit, albumin, and globulin were determined only from the first blood sample of each infant. Other parameters which were entered into the data log of the infant include: time of measurement, time of birth, estimated gestational age, race, sex, length, weight, and blood type of mother and infant.

B. Design of Experiments

Prior to obtaining the major well-baby data set, three experiments were designed and carried out to assess effects which might be important in the major analysis effort. Data was collected to: (1) assess the effect on reflectance measurements of various factors such as probe operator, sex, side of the infant, and blanching and (2) approximately determine the size of the measurement error and repeatibility of the measurement procedure.

In these three initial studies, the spectral reflectance of infants was measured in the 350 to 700 nanometer range with bandwidths of 10 nanometers. Six wavelengths (450, 460, 540, 550, 560 and 600 nanometers) were selected for analysis to be representative of the wavelengths measured. The first two are bilirubin absorportion wavelengths, the next three are hemoglobin absorption wavelengths, and the last was selected for comparison with the first five.

The design of the first experiment was a randomized complete block, Chapter 4 of Anderson and McLean [1], where the babies acted as blocks. The first data set was acquired in April, 1977, and included measurements on both the left and right sides of the back of each infant using both a blanching technique and no blanching. Spectra were gathered on two consecutive days. Three different probe operators were used, permitting comparisons among operators. Other factors examined in this experiment were differences due to side of the infant observed, differences due to the blanching technique,

and differences between days of measurement. The size of the measurement error was also assessed.

The next data set, a 4x4 matrix, was acquired in April and May, 1977. The design of this experiment was a nested factorial, Chapter 6 of Anderson and McLean [1]. For four infants each of white and black, male and female, measurements were acquired on the left and right sides, both blanched and unblanched, at two times during two consecutive days. This data set was acquired before analysis of the previous data set had been completed and, in fact, used some of the same measurements. Two probe operators were used in this experiment so that any possible operator effects are completely confounded with the effect of the infants.

As a result of the analysis of prior experiments, an operator-repeatibility study using a randomized complete block design with babies acting as blocks, was carried out. The purpose of this experiment was to test for any differences in reflectance due to probe operators and to evaluate the changes in measurement values acquired by the same operator at different times during the day. Spectral measurements of four white female infants were taken once in the morning and twice in the afternoon. Measurements were made on both sides of the back of each infant using the blanching technique. Three probe operators measured all infants during each time period. Three readings were acquired on each side of each infant during the first measurement session in the afternoon by all operators. The four infants examined were divided into two groups, fat and thin, to assess effects on measurement of the fleshiness of an individual. This latter classification forced the design to become a nested factorial. This allowed investigation of fat vs. thin.

The major well-baby data set was acquired in early August, 1977. Spectral measurements were acquired in the 400 to 800 nm range. Only one probe operator was used and spectral measurements were acquired on the left

and right sides using the blanching technique. A total of 108 infants were measured: 37 white males, 23 white females, 28 black males, and 20 black females. An average of three observations (on different days) was acquired for each infant.

III. ANALYSIS AND RESULTS

A. Preliminary Analyses

The first analysis was designed to address the effects of various factors on spectral response. Due to three missing observations, the assumption was made that there should be little difference from side to side. In order to check this assumption, the missing cells were filled in by the observation from the other side and an analysis of variance was run. Even after replacing the missing observations with their counterparts from the opposite side, the analysis showed side to be a significant factor. As a result, analyses to determine effects of other factors were run with the incomplete design because the assumption that sides are equal appears to be unfounded.

Several possible explanations for the side differences have been offered. The effect could be due to probe operator as the left side was always measured first and the technique might not be complete repetitive. Secondly, this could be a dermographic effect as a consequence of being contacted by the probe and the skin being blanched. There is limited evidence in support of the second explanation since a large number of the left vs. right inconsistencies showed no particular trends.

Analysis of variance was run pooling the three, four, and five factor interactions for an error term. Most factors were found to have a significant effect on reflectance including side, operator, blanching, and several

interactions. Those interactions which were significant at the 5% level were: baby by side, operator by blanching, baby by blanching, and operator by day.

Additional studies were performed on the probe operator effect and it was determined that only measurements using the blanching technique on the left side of an infant would be used in the major study. The interactions which are significant seem to indicate the presence of some measurement variability due to the probe operator and measurement technique. On one infant, measurements on the left side were consistently greater than on the right side while the opposite occurred on another and the third had equal measurement on both sides when averaged over all operators. One operator got a greater difference between blanched and unblanched measurements. Some babies were more affected by the blanching process than others and one operator measured more consistently from day to day than the other.

The second analysis performed was to evaluate the validity of the pooling process used in the first analysis. The residual in this analysis was similar in magnitude to the first analysis, thus substantiating the approach and results obtained.

An analysis was then performed on the 4x4 matrix study. As this experiment was designed prior to analysis of the earlier studies, two operators were used because no operator effects had been anticipated. Thus, any probe operator differences present are completely confounded with the spectral response of an infant. The error used for testing in the analysis of variance was the pooled error from each of the interactions with baby (the random factor).

Race of the infant and blanching caused a significant difference in reflectance spectra at all the wavelengths analyzed. The side effect was not significant in this analysis. The following interactions were

significant at the 1% level: race by blanching, race by day, and race by sex by blanching. Interactions which were significant only at the 5% level and those which were not significant at 540, 550, or 560 nanometers were considered of minimal importance.

Because of the significance of the interactions, a covariance procedure, Ostle and Mensing [4], using total bilirubin, direct bilirubin, and estimated gestational age was employed. All the blanched observations from the 4x4 matrix were analyzed. The use of any one singly or all covariates together caused all factors (except race) and all interactions to become nonsignificant at the 540, 550, 560, and 600 nanometer wavelengths.

The first analysis carried out on the operator/repeatibility study was an analysis of the repeatibility of measurements. For this analysis, the 72 measurements obtained at the first time in the afternoon were used. The within error was found by pooling all interactions with the random factor (baby). This within error was smaller in magnitude than the errors obtained in previous analyses which indicates a high degree of repeatibility. Side, fleshiness by side, and fleshiness by order were significant effects at some wavelengths. As the fleshiness by side interaction is highly significant ($\alpha = .01$) at five of the six wavelengths, it would be desirable to investigate this effect further. This is not feasible, however, due to lack of an appropriate data set. One interpretation of these results seems to be that the fleshier an infant is the more difficult it becomes to be consistent in the pressure and blanching aspects of the measurement process.

The second analysis which was carried out on the operator/repeatibility data set used the first measurement made by each probe operator at each of

the three times during the day. Again, 72 observations were used in the analysis. The within error was estimated in the same manner, but it was much larger than the previous error as would be expected since more time had elapsed between measurements.

Due to the large error for testing, only the 600 nanometer wavelength band showed any significant effects. There the fleshiness by time and operator by time interactions were significant at the 5% level. However, the fleshiness by side and operator by time interactions may be important at all wavelengths because of the relative size of the mean squares for these effects.

No physical explanation for the fleshiness by side interaction has been identified. Heights and weights of all infants will be recorded for possible use in a covariance procedure. Since one of the operator interactions was significant, more investigation still needs to be done on the effect of operators.

B. Major Well-Baby Study

Analyses carried out on this data set have consisted of analysis of variance and covariance; correlations of reflectance values at 41 wavelengths (from 400 to 800 nanometers) with total bilirubin measurements; regressions predicting total bilirubin by ratios of reflectance values and and correlations of ratios with direct and indirect bilirubin; investigation of a multiple ratio model; and selection of the best subset of wavelengths for a multiple linear regression.

Analysis of variance and covariance. Analyses of variance and covariance, Chatterjee and Price [2], were performed on an equal cell size subset of the data to assess effects and identify possible problem areas. Five

wavelengths (450, 460, 540, 550, and 560 nanometers) were analyzed. Side of the infant did not appear to have an overall effect on reflectance in this data set although the baby by side interaction was significant at the 450 and 460 nanometer wavelengths. Race was a significant effect at all wavelengths as anticipated. A possible source of difficulty was found in the significance of the day effect and the baby by day interaction because these effects were not removed even when using total bilirubin as a covariate.

Since race has been a significant effect at all wavelengths in all the preliminary studies as well as in this data set, the remainder of the analyses were conducted twice, using the data from each race separately. Because of the existence of a possible side effect, only measurements from the left side were used in the analysis. Only blanched measurements were acquired.

Correlation of reflectance values with bilirubin measurements. Plots of bilirubin level as a function of reflectance at each of the 41 wavelengths were generated separately for the white and black babies. Linear (Pearson) correlation coefficients were calculated. Much scatter was present and correlations were very low at all wavelengths measured. At most 21% and 5% of the variation was explained for the white and black infants, respectively.

One initial hypothesis for the lack of correlation was that infants of different ages were represented on the same graph. Plots and correlations, generated for the white infants of the same age in days, did not show a significant improvement over considering all ages together.

Correlations and regression using one ratio of reflectance values. Again, two independent analyses were conducted: one on the white infant data set and one on the black infant data set. The results for the white infants will be

discussed first.

Scattergrams and linear correlations were generated for total bilirubin with a ratio:

$$\frac{X_i}{X_{800}}$$

where X_i is reflectance at the i^{th} wavelength and i = 400, 410, 420,...,790 nm. The 800 nm wavelength was selected to average out melanin effects. The correlations were not significantly improved by this ratio procedure, explaining a maximum of 28% of the variation in bilirubin level. Then correlations were calculated of total bilirubin with a ratio:

$$\frac{X_i}{X_{550}}$$

where i = 400, 410,...,540, 560, 570,...,800 nm. The 550 nm wavelength was selected to average out hemoglobin effects. The R^2 values were .60 for 450 nm and .57 for 460 nm, a large improvement over the use of a single wavelength band. Plots of this ratio were generated for the observations taken at ages two, three, and four days of age. Correlations were somewhat higher for infants of age two days, but substantial scatter and low correlations were the pattern on the other two days. Then, since age in days may not have been a precise enough measurement, a regression for the white infants was tried using age in hours. Linear and quadratic terms improved the regressions only slightly. This approach, as it appeared useless, was not carried out on the black infant data set. One reason that this approach may not have worked is that while bilirubin level may have a generally quadratic shape over time, the maximum achieved by the curve will be different for different individuals although the "endpoints" (beginning and ending bilirubin levels) may be the same.

The ratio using reflectance at 550 nm is not optimum for maximizing R^2 for a single ratioed variable. Four additional wavelengths (470, 500, 510, and 520) were tried as denominators and each achieved a higher R^2 value ranging from .63 to .66 (Table 1, Figure 1). So, if this type of model were to be selected, a study would need to be done to determine the optimum combination.

Analysis for the black data set was conducted in a similar fashion. Ratios with reflectance at 800 nm were calculated; plots and linear correlations were generated. This variable was even less correlated with bilirubin level for the black infants than it was for the whites. At most 7% of the variability in bilirubin level was explained by this ratio. Similarly, the ratio with reflectance at 550 nm did not perform exceedingly well, explaining at most 40% of the variability.

Four additional denominators were tested: 470, 500, 510, and 520 nm. Some ratios with each of these wavelengths (Table 1, Figure 2) provided substantially higher R^2 values, exceeding the best achieved by the white data set. These may not be the optimum ratios, however.

Although the R^2 values were not very high, the relationship between bilirubin measurements and ratios of reflectance values might still prove useful. Prediction intervals were calculated for one ratio model for each race. The 470/500 nm model gave 95% prediction intervals to be $\hat{Y} \pm 2.96$ for whites and $\hat{Y} \pm 3.52$ for blacks. The error appears large, but if it is close to or better than the error in laboratory measurements, this approach would be preferable.

Investigation of a multiple ratio model. Inclusion of more than one ratio in a regression model was considered for the white infants only.

Table 1. R^2 values for prediction of bilirubin level by ratio variables.

Race	Numerator	Denominator	R^2
White	450	550	.60
	460	550	.57
	450	520	.66
	470	510	.63
	470	500	.65
Black	470	500	.67
	480	500	.69
	490	500	.68
	470, 480	510	.67

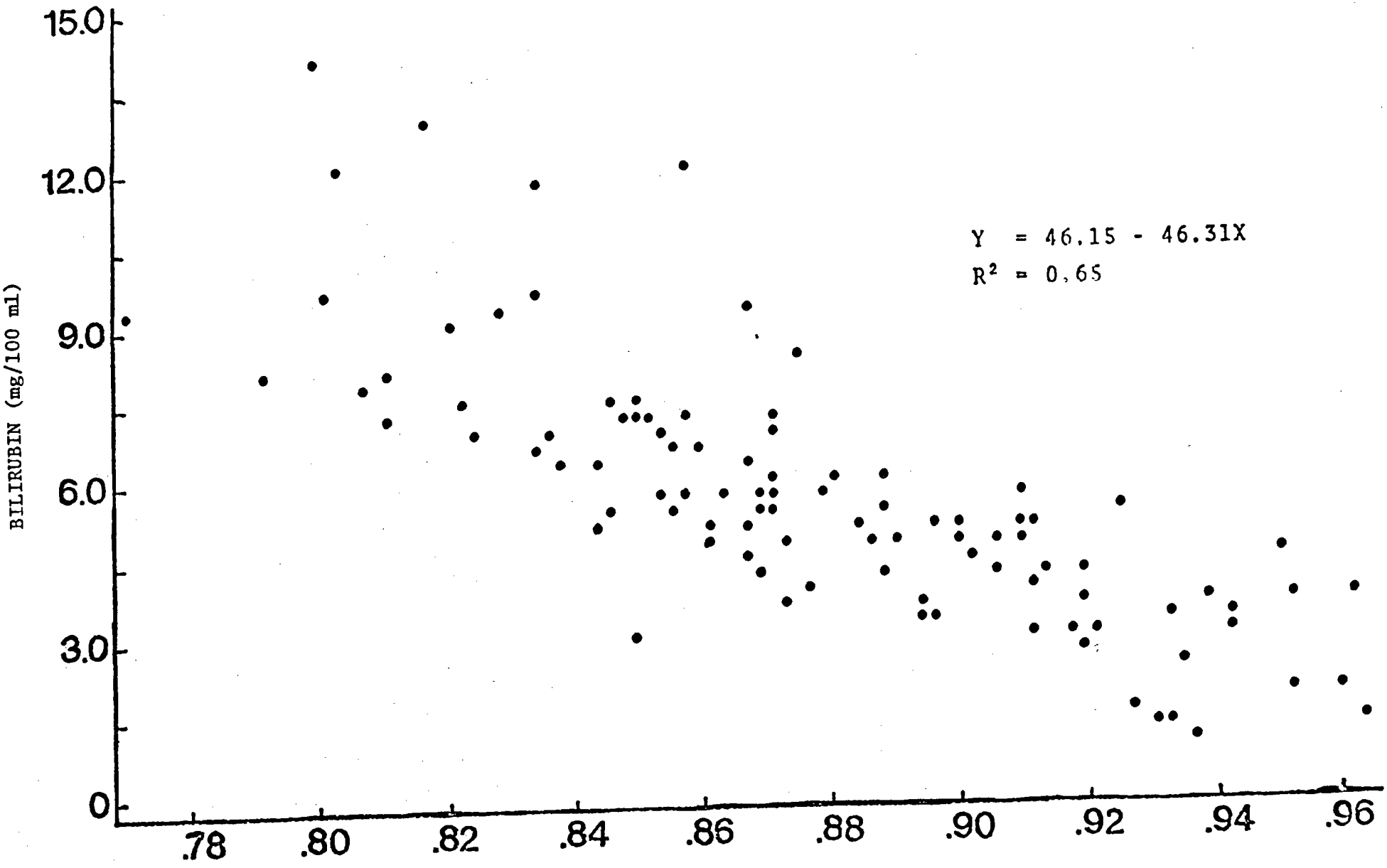

Figure 1. Plot of bilirubin vs. the reflectance ratio $\frac{X_{470}}{X_{500}}$ for a sample population of 60 white infants.

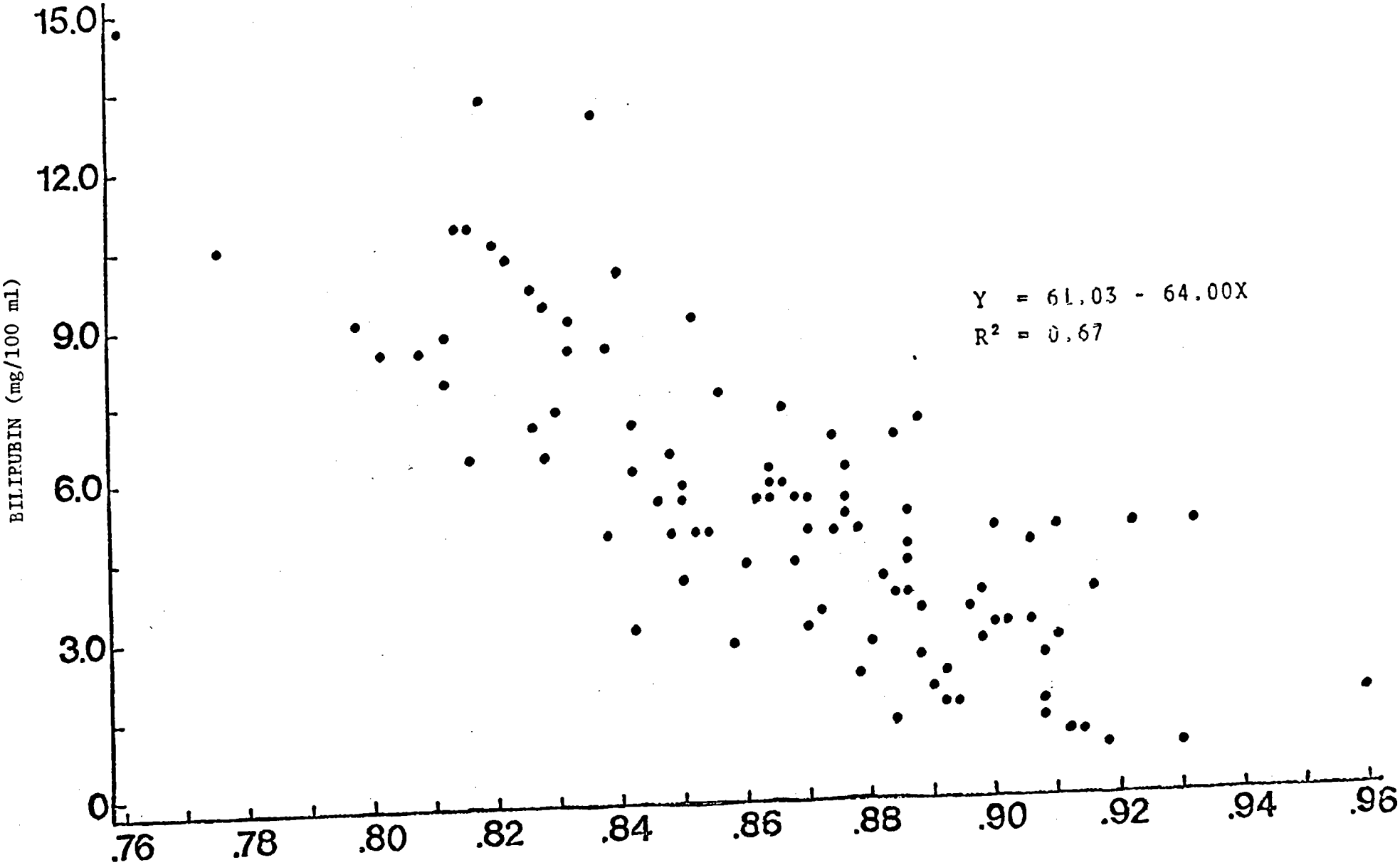

Figure 2. Plot of bilirubin vs. the reflectance ratio $\frac{X_{470}}{X_{550}}$ for a sample population of 48 black infants.

Terms using the reflectance at 400 to 600 nm as numerators and 420 and 510 nm as denominators were tried (a total of 42 variables). Using the forward selection method, no significant increase in R^2 occurred after the thirteenth variable had entered. The model achieved an R^2 value of .79 and is given by:

$$Y = -105.23 + 164.13 \frac{X_{460}}{X_{510}} - 27.82 \frac{X_{430}}{X_{420}} + 10.53 \frac{X_{600}}{X_{420}} - 3.07 \frac{X_{440}}{X_{420}}$$

$$+ 122.79 \frac{X_{520}}{X_{420}} - 70.78 \frac{X_{470}}{X_{420}} - 28.67 \frac{X_{530}}{X_{420}} - 18.46 \frac{X_{590}}{X_{420}}$$

$$+ 58.30 \frac{X_{580}}{X_{510}} - 69.53 \frac{X_{570}}{X_{420}} + 59.68 \frac{X_{560}}{X_{510}} - 109.88 \frac{X_{450}}{X_{510}} + 28.00 \frac{X_{490}}{X_{420}}$$

The optimization of this model and reduction of terms was not pursued.

Investigation of a multiple linear model. A multiple linear model, Chatterjee and Price [2], was investigated for both white and black infants separately. It was desired to find a relatively small subset of wavelengths containing most of the spectral information for prediction of bilirubin level. The methods of all possible regressions and C_p values were used for selection of the subset. All results for white infants will be discussed first, followed by the results obtained for black infants.

Forty-one reflectance measurements had been acquired for each infant. All possible regressions would not run on such a large subset, so separate jobs were run for the 400-600 nm and 610-800 nm wavelengths. The computer was incapable of handling even these reduced subsets. It was suspected that its inability to handle this many variables was that, since many of the reflectance values were very nearly identical relative to the magnitude of

the observations, the (X'X) matrix was probably nearly singular and the method used for matrix inversion in the program could not handle the precision requirements. The approach taken to this problem was to omit variables from consideration which were highly correlated with other variables ($r \geq .98$) and those which would not enter under the default parameters in stepwise regression program.

In the first phase of the analysis, 11 wavelengths were entered into all possible regressions: 400, 410, 420, 440, 470, 490, 520, 550, 580, 590, and 600 nm. Using the C_p criterion, the best subset size was chosen to be six variables. The R^2 value did not increase significantly after this point. The best subset of this size was: 410, 420, 440, 490, 520, and 550 nanometers.

In the second phase of the analysis, the same regression procedure was used to eliminate some variables from the subset to be considered. Nine variables between 610 and 800 nanometers (620, 630, 660, 670, 710, 730, 750, 760, and 780) and the six variables previously selected were entered into all possible regressions. The best subset size was again selected to be six and the same wavelengths were selected as in the first phase: 410, 420, 440, 490, 520, and 550 nm. This regression model gave an R^2 value of .69 and a 95% prediction interval of $\hat{Y} \pm 2.9$. The model is given by:

$$\hat{Y} = 6.43 - 0.59\ X_{490} + 1.53\ X_{520} - 0.74\ X_{550} - 1.18\ X_{440}$$
$$+ 2.06\ X_{420} - 1.13\ X_{410}$$

Residual plots showed that, in general, the worst prediction was for infants with higher bilirubin levels. Use of covariates (estimated gestational age,

pounds, ounces, length, hematocrit, and weight in grams) did not appear to substantially improve the correlation. Residual plots were also done for these regressions, but very little changed from the plots done without covariates. It appears that the single most important covariate is the estimated gestational age (R^2 increase of about .02) and that the other covariates are of little value.

The same procedures were followed for investigation of a multiple linear model for black infants. The analysis was divided into three phases and the wavelengths entered at each phase differed from those selected for white infants.

In the first phase, 14 wavelengths between 400 and 600 nm (400, 410, 420, 430, 440, 470, 490, 510, 520, 550, 560, 580, 590, and 600) were considered by all possible regressions. Using the C_p criterion, all except 410 and 470 nm were selected for further consideration.

In the second phase, 16 wavelengths between 610 and 800 nm (610, 620, 630, 640, 650, 660, 670, 690, 720, 730, 750, 760, 770, 780, 790, and 800) were considered by all possible regressions. Using the C_p criterion, three wavelengths (690, 720, and 790 nm) were selected.

The third phase consisted of running all possible regressions on this subset of 15 variables, combining those variables selected in the first two phases. The C_p value decreased very slowly because the mean square for error was large. Being unreasonable to include a substantially larger number of variables in the model for blacks than required for whites, the decision was made to ignore C_p values in the choice of an "optimum" subset for blacks. It seems to make sense to select five as the optimum number of wavelength bands for the following reasons:

(1) For blacks, forcing of variables is required for six or more

wavelengths bands to enter since high correlations exist between reflectance values.

(2) For whites, the C_p value for five variables did not indicate a substantial bias.

(3) Symmetry on blacks for five variables is good (i.e., using wavelengths spread throughout the spectrum). For whites, the choice of five wavelength bands seems to be reasonable because five of the six optimum are selected and the R^2 decrease is only .02 from the optimum six.

The subset selected was: 400, 490, 520, 560, and 590 nm. The R^2 value is .77 and a 95% prediction interval is given by $\hat{Y} \pm 2.94$. The regression equation is given by:

$$\hat{Y} = 3.32 - 2.37\ X_{490} + 2.94\ X_{520} - 1.31\ X_{560} + 0.24\ X_{690} + 0.32\ X_{400}$$

IV. SUMMARY AND CONCLUSIONS

The results of statistical analysis of these data sets substantiate that there is a strong relationship between serum bilirubin concentration and skin reflectance. It has been hypothesized that the source of variability and lack of correlation is in measurement error rather than due to individual variation. Furthermore, it has been shown that pigmentation does not obscure the relationship. The possibility that the method could be clinically useful has been clearly demonstrated, although experiments using new probe design concepts are required to reduce measurement variability and thus improve the usefulness of this optical method.

Reflectance in a single wavelength band was inadequate for prediction of bilirubin level: at most 21% and 5% of the variation was explained for

the white and black infants, respectively. Ratios of two reflectance values significantly increased the R^2 values; yet insufficient confidence can be placed in this relationship for predictive purposes.

Multiple linear regressions employed information from more regions of the spectrum, thus R^2 values were increased. Of the 41 reflectance measurements acquired, a "best" subset for prediction of bilirubin was selected for each race using a combination of regression techniques. The R^2 value for blacks was higher than for whites. Some possible reasons for this might be:

(1) The range of wavelengths chosen for blacks is much greater than that selected for whites.

(2) Correlations between variables for whites are higher, thus causing less additional increase in R^2 when more bands are included. If there were a wider range of wavelengths available, a higher R^2 might be achieved for whites.

(3) The degree of pigment variability is not as large in the black infants as the whites.

(4) The melanin pigment, more predominant in the blacks, prevents light penetration to the deeper dermal layers where the light scattering and absorbing mechanisms are more complex.

Investigations of a multiple ratio model with wavelengths 400-600 nm over 420 and 510 nm gave R^2 values of .75 for 6 terms (requiring 8 measurements) and .72 for 4 terms (requiring 6 measurements). This type of model improved the R^2 value for the same data set and the same number of measurements over what could be achieved by a multiple linear model.

The amount of variability in bilirubin which can be explained by reflectance is not sufficient in any of the models explored to determine

a predictive relationship. Although data transformations or alternate models might increase this relationship somewhat, plots of the data seem to indicate the source of the variability is in the data itself. Previous work by Weichel, which found a very strong relationship, was biased upward by use of a small, homogeneous data set. This work has indicated the potential for prediction of bilirubin level by spectral measurements if the source of variability present can be identified and eliminated (or adjusted for).

The analysis on this data set was suspended before several additional models could be explored due to the belief that remaining funds could be more profitably used in acquisition and analysis of a new data set. The square root, natural logarithm, and double log transformation of reflectance values should be considered in a model. Multiple ratio models show promise for further investigation and the use of estimated gestational age for improvement in the model should be considered.

From this analysis, several things were learned which may aid in design of future experiments. There are several sources of measurement variability that need to be explored in pursuit of an operational screening system. Two sources of variability which may exist in this data set are the difference in reflectance due to pressure applied in blanching and differences in technique used by the operator over time. An additional source of variability which would be present in an operational situation is the difference between operators. It is believed that reduction of the variability in measurement by use of a different probe would increase the strength of the relationship sufficiently for predictive purposes. Research using a new probe began in July, 1978.

V. REFERENCES

1. Anderson, V. L. and McLean, R. A., "Design of Experiments: A Realistic Approach", Marcel Dekker Inc., 1974.

2. Chatterjee, S. and Price, B., "Regression Analysis by Example", Wiley and Sons, 1977.

3. Hannemann, R. E., DeWitt, D. P., and Wiechel, J. F., Pediatrics Research (in press).

4. Ostle, B. and Mensing R. W., "Statistics in Research ", 3rd. ed. Iowa State University Press, Ames, Iowa, 1975.

5. Wiechel, J. F., "The Optical Detection of Jaundice", M. S. Thesis, Purdue University, May 1975.

6. Wiechel, J. F., DeWitt, D. P., and Hannemann, R. E., Proceedings of the 28th Annual Conference of Engineering in Medicine and Biology, New Orleans, October 1974, p. 340.

7. Wiechel, J. F., DeWitt, D. P., and Hannemann, R. F., 1975 Advances in Bioengineering (A. C. Bell and R. M. Nerem, Editors), American Society of Mechanical Engineers, Bioengineering Division, Winter Annual Meeting, Atlanta, December 1975, p. 25-27.