

ASSESSMENT OF THE FEASIBILITY OF DETERMINATION OF CHOLESTEROL AND OTHER BLOOD CONSTITUENTS BY NEAR-INFRARED REFLECTANCE ANALYSIS

ROBERT A. LODDER* and GARY M. HIEFTJET†

Department of Chemistry, Indiana University, Bloomington, IN 47405-4001, U.S.A.

WELLS MOOREHEAD

Department of Pathology, Indiana University Medical Center, Indianapolis, IN, U.S.A.

STEVEN P. ROBERTSON and PHILLIP RAND

Miles Laboratories, Inc., Elkhart, IN 46514, U.S.A.

(Received 6 May 1988. Accepted 8 September 1988)

Summary—Near-infrared reflectance spectrometry of blood serum can yield values for serum cholesterol that correlate reasonably well ($r = 0.96$) with those from common reference analytical methods. However, the variability of serum can cause ostensibly validated calibrations to fail on new samples. The determination of blood components such as cholesterol and triglycerides by near-infrared reflectance is complicated by their low concentrations, the variety of forms in which they appear, and by the natural variability of the blood matrix. These difficulties, when combined with the problems encountered in obtaining a representative sample from a given individual, can make it almost impossible to select, by a regression procedure, a wavelength combination that is characteristic of the complete blood matrix. The failure of the regression process to find characteristic wavelengths generates a false-sample problem in which even small changes at the analytical wavelengths produce a grossly unreliable cholesterol or triglyceride determination.

Epidemiological studies performed over a period of years have indicated that reduction of blood cholesterol levels significantly reduces the risk of atherosclerosis, ischemia, myocardial infarction and death. For some time these data have been cited in experimental attempts to prevent arterial disease. A 1% reduction in plasma cholesterol concentration in individuals at risk for cardiovascular disease has been shown to reduce the risk of cardiac events in these individuals by approximately 2%.¹ More recent data² indicate that lowering the cholesterol level improves the condition of coronary arteries partially blocked by atherosclerotic lesions, and can actually effect regression of the disease. These data have been used to make a case for creating a target level for total blood cholesterol of 185–200 mg/dl, a level below the average for the U.S. population. Another report has indicated the discovery of a new mechanism by which atherosclerosis may initiate high blood pressure, cardiac disease, and transient ischemic attacks in the brain.³ In this report, the accumulation of deposits in arterial walls is described as interfering with the supply of endothelium-derived relaxing factor (EDRF) to muscle fibers, resulting in the onset of

vasospasm. Animals fed high-fat diets developed atherosclerotic changes in blood vessels, accompanied by impaired EDRF secretion. Switching to a normal diet restored endothelial production of EDRF, indicating that a low-cholesterol diet may actually act to reduce vasospasms in addition to bringing about the reduction of atherosclerotic lesions.

Plasma cholesterol itself has been the object of considerable research. Cholesterol has been shown to be almost totally carried in lipoprotein particles of different size, density, and lipid and apolipoprotein composition.⁴ Cholesterol and triglycerides in different lipoprotein particles take different metabolic pathways, and have different effects on arterial disease. High-density lipoproteins (HDLs) have been identified as removing cholesterol from tissues (and exhibiting a protective effect against arterial disease) by a process known as reverse cholesterol transport.⁴ Apolipoprotein A-I, the principal apoprotein of HDLs, activates the plasma enzyme lecithin cholesterol acyltransferase, forming nonpolar cholesteryl ester and shifting cholesterol from the surface of the HDL particle to the hydrophobic core for transport to the liver. In the liver, cholesterol can be eliminated from the body either as bile acids or as cholesterol, but predominantly as the former. Low-density lipoprotein (LDL), composed of esterified and

*Present address: College of Pharmacy, University of Kentucky, Lexington, KY 40536-0082, U.S.A.

†Author to whom correspondence should be addressed.

nonesterified cholesterol as well as phospholipid, transports cholesterol to peripheral cells for membrane synthesis. Elevated levels of LDL have been shown to increase the risk of cardiovascular disease.

The increase in knowledge of lipoprotein metabolism and atherogenesis has been accompanied by an abundance of experiments designed to test a variety of treatments for atherosclerosis and hypercholesterolemia. Dietary studies have suggested that foods containing pectin,³ such as grapefruit, and calcium pectate,⁶ such as carrots, onions, and cabbage, may be useful in reducing cholesterol levels. Calcium pectate appears to lower cholesterol by sequestration of bile acids.⁶ Though the effectiveness of fish oils in modifying lipoprotein levels has been debated, ethanol in small daily doses has been shown to increase levels of HDL₂ and HDL₃ (the major blood HDL components), leading to a 46% reduction in the risk of myocardial infarction.⁷ Of course, ethanol consumption can lead to another set of health problems, and use of ethanol to increase HDL levels is never recommended.⁴ Experiments to modify cholesterol levels with drugs are continually being conducted. Niacin, which in low doses acts as a vitamin, when administered in doses of more than 300 mg/day reduces LDL cholesterol by interfering with the initial secretion of atherogenic particles. Cholestyramine resin and colestipol hydrochloride reduce cholesterol by sequestering bile acids.⁴ New drugs, based on the inhibition of HMG (hydroxymethylglutaryl) coenzyme A reductase, break the chain of cholesterol synthesis at the mevalonic acid stage and appear even more effective in reducing cholesterol levels than colestipol, cholestyramine, and niacin.⁸ Finally, surgical remedies for cardiovascular disease such as balloon angioplasty (150,000/yr in the U.S.) and coronary bypasses (220,000/yr in the U.S.) are being augmented by a number of methods intended to make certain that newly opened arteries remain open. Stents (tubes left inside blood vessels to hold them open), sound waves from modified lithotrippers, and laser catheters made of sapphire-tipped optical fibers were all described at the recent 60th American Heart Association meeting in Anaheim, CA.⁹ In short, the role of lipoprotein cholesterol in cardiovascular disease is becoming increasingly clear, and methods for interdiction of the disease are becoming increasingly numerous and effective.

The natural question for the analytical chemist, then, is how to identify individuals with hyperlipoproteinemia. On the surface this appears to be a question that was answered long ago: clinical blood-analysis instruments from Technicon and Dupont,¹⁰ among others, have been available for some time. However, concern about the accuracy of the present determinations as well as calls for the mass-screening of individuals to detect incipient hyperlipoproteinemia have indicated the need for a new, rapid, low-cost method of blood analysis. On 5 October 1987 the National Cholesterol Education Program

of the National Heart, Lung, and Blood Institute released a widely publicized report^{11,12} calling for the testing of all adults, aged over 20, every five years. If the test result is in excess of 200 mg/dl (total serum cholesterol) or other risk factors exist (such as being male, a smoker, obese, diabetic, hypertensive, or having a family history of premature coronary disease), then more frequent (as often as several times each year) and more complex (simultaneous LDL cholesterol determination) testing may be indicated. These tests can cost from \$11 to \$40 apiece, and even then do not ensure an accurate reading of an individual's cholesterol levels. A 1985 College of American Pathologists Comprehensive Chemistry Survey found that 47% of the 5000 testing laboratories volunteering for the survey could not get a result within 5% of the true cholesterol value.¹² This error is significant because clinical risk brackets are generally less than 10% wide.¹⁰ Laboratory inaccuracies and natural biological variation (which can be¹⁰ as high as 10% even in individuals maintained under metabolic-ward conditions) combine "to render a single test virtually meaningless." Further,¹² in order "to be sure [of the actual value] within 5%, the test . . . need[s] to be repeated five to ten times." Still, a single test remains a better screening method than no test at all.

The large number of tests generated by a mass-screening program is made much larger by the number of repeat tests necessary to achieve a clinically meaningful result, creating an analytical problem with a solution that appears expensive in terms of both time and money. The development of a rapid, low-cost, and completely spectroscopic method of analyzing blood with good precision and accuracy would be a major step toward achieving the mass-screening goal. Work has been done on completely spectroscopic cholesterol determinations in both the infrared and near-infrared. In the infrared,¹³ simultaneous determinations of relatively pure tripalmitin, dipalmitoyl-DL- α -phosphatidylcholine, and cholesteryl palmitate in reagent-grade chloroform solution were performed by using 15 wavelengths and a multiple linear regression procedure. A training set of 85 mixtures of these reagents was required for the regression even with these carefully prepared samples. Near-infrared determination of serum cholesterol in 30 human sera samples (ranging from 3 to 12mM in 0.3mM increments) by a similar multiple linear regression procedure has been reported.¹⁴ The report described the careful construction of the training set and regression at 5, 6, and 7 wavelengths to give correlations with total cholesterol, with *r* values ranging from 0.92 to 0.93.

The present report describes not only the near-infrared determination of cholesterol and triglycerides, but also lists a number of factors that often thwart such determinations. The pattern-recognition spectroscopic analysis of samples such as blood serum is complicated by (1) the large number of very

similar components present at about the same low concentration, (2) the lack of a satisfactorily accurate reference method, (3) the natural variability of the matrix relative to the components of interest (e.g., different diets, fitness levels, and diseases affect the lipid distribution in serum), and (4) the difficulty in obtaining a representative sample from a given individual (e.g., it is known that both the time the tourniquet is left on and the standing or sitting position of an individual during sample-drawing can affect the cholesterol measurement¹¹). Instrumental factors such as drift, inadequate signal-to-noise ratio, and lack of availability of sufficient independent wavelengths can also increase the difficulty of spectroscopic pattern-recognition. Finally, the large number of samples required to represent adequately the variability of the population can also become an obstacle to developing effective pattern-recognition methods.

EXPERIMENTAL

Apparatus

A Technicon InfraAlyzer 400 filter spectrophotometer was used to collect near-infrared reflectance data at 18 different wavelengths. The spectral data were analyzed on a VAX 11/780 computer (Digital Equipment Corp., Maynard, MA). The data-analysis programs were written in Speakeasy IV Epsilon (VMS version, Speakeasy Computing Corp., Chicago, IL). Serum samples were analyzed in a 70- μ l disposable microcell¹⁵ designed for use in the Technicon spectrophotometer.

Materials

The near-infrared method of serum analysis requires no reagents. The serum samples were acquired from participants in a "Fitness Fair" conducted by the Indiana University School of Medicine in Indianapolis, IN. Blood samples were permitted to clot and were then centrifuged. The resulting serum was transferred to the microcell with a precision pipet (Rainin Instrument Co., Woburn, MA). Four scans were taken of each sample and these scans were averaged prior to data analysis.

Reference cholesterol values for the samples were obtained by using a Dupont aca instrument that was provided by the Department of Pathology. The reproducibility of measurements taken with such a device is good¹⁰ and the bias is unlikely to be very large as long as reference samples are regularly employed to control it. Additional biochemical screening tests were performed with an Eastman Kodak Ektachem 700 instrument.

RESULTS AND DISCUSSION

An initial experiment was performed with a small training set (30 samples) that was carefully assembled to cover the entire range of human cholesterol values. This experiment is similar to the one described by Peuchant *et al.*,¹⁴ where the cholesterol range from 4 to 12mM was covered by 30 samples in 0.3mM increments. To enhance the realism of the test, this experiment was performed with unaltered clinical samples (no cholesterol added), so total coverage of the range from 4 to 12mM had to be approximated by samples drawn from the Fitness Fair pool. These samples ranged from about 4 to 10mM

(156–373 mg/dl) total cholesterol (no samples with higher cholesterol concentration were available). A validation set of 30 samples (*i.e.*, samples not used to develop a calibration equation) was also drawn.

The calibration equation was developed by principal-component regression (Speakeasy Computing Corp.) on the 18-wavelength data for the 30 samples in the training set. The correlation coefficient (r) between the experimental values and the fitted values for the training set was 0.966, and the corresponding standard error of estimate (SEE) was 12.4 mg/dl (*i.e.*, RSD = 6%). The calibration equation was then applied to determination of the total cholesterol values of the validation set. The correlation coefficient between the cholesterol values determined by the reference method mentioned above and the values obtained by application of the calibration equation (to the reflectance data for the validation samples) was 0.960, somewhat better than had been obtained by Peuchant *et al.*¹⁴

Unfortunately, such a small number of carefully selected samples can easily fail to cover the range of variations that can exist in a complex sample such as blood serum. Clinical samples are often affected by hemolysis or turbidity, for instance. The presence of large numbers of chylomicrons in some samples must be expected. Furthermore, the determination of total serum cholesterol is really the determination of cholesterol contained in a number of particles that are quite different. About two-thirds of the total cholesterol in plasma is carried in LDL and IDL (intermediate-density lipoprotein) particles 21–35 nm in diameter. The LDL particles have a surface layer that contains about 8% cholesterol, while about 42% of the core is formed of cholesteryl esters. The remainder of the LDL particles comprises 6% triglycerides (found in the core), 22% phospholipids (found at the surface), and 22% protein (also found at the surface). The principal LDL apolipoprotein is apoB (95% of the apolipoprotein content), but traces of at least 7 other apoproteins can be found. IDL has a surface composition similar to LDL and a core that contains more triglycerides (23%) but less cholesteryl ester (29%).

The second-largest carrier of plasma cholesterol is HDL, which contains mostly protein (about 50% located at the surface), phospholipids (also at the surface, about 30%), and esterified cholesterol (in the core, about 15%). The major HDL apolipoprotein is apoA-I (64% of apolipoprotein content), but significant amounts of A-II, C-I, C-III, ARP, and D can also be found in HDL particles. This may seem already complex enough, but plasma HDL can actually be separated into two distinct components, HDL₂ (particles about 10 nm in diameter) and HDL₃ (particles about 7.5 nm in diameter). HDL₂ has more cholesteryl ester (in the core) and phospholipid (on the surface) and less protein (on the surface) than HDL₃.

Cholesterol and cholesteryl ester are also found

in smaller quantities in VLDL (very low-density lipoprotein) and chylomicrons as well as in certain abnormal lipoproteins.

If the total cholesterol concentration for an individual were to be partitioned into the six or so major particle-environments for cholesterol, the average cholesterol concentration in each environment would be roughly 30 mg per dl of the total serum. This concentration is low by near-infrared standards, suggesting that a good, general calibration equation for cholesterol may be difficult to achieve. Kisner *et al.*¹³ discussed the problem of infrared determination of cholesterol, and needed 85 three-component training samples (prepared as solutions in chloroform, in the laboratory) monitored at 15 wavelengths, to develop a useful calibration equation. Indeed, larger training-set sizes were proposed as a means of generating more accurate calibrations. The components interfered with one another substantially even though there were only 3 of them. The situation is exacerbated in the near-infrared, where the signals consist largely of overtones and combinations of fundamental infrared vibrations.¹⁶

The Fitness-Fair blood-sample pool contained normal samples, icteric samples, hyperlipoproteinemic samples, samples from both sexes, samples from fasting and nonfasting individuals, samples containing drugs, samples that were refrigerated prior to analysis and samples that were not, samples with hemolysis and turbidity, *etc.* When all the blood samples in the Fitness Fair pool (a total of 162) were used to develop a near-infrared calibration equation for cholesterol, the standard error of estimate for the calibration rose to 36.5 mg/dl. Dividing the samples into two groups and calibrating with 81 training samples gave an SEE of 33.5 mg/dl. On the surface it appeared that the pool contained a number of sample subgroups, each of which might best be

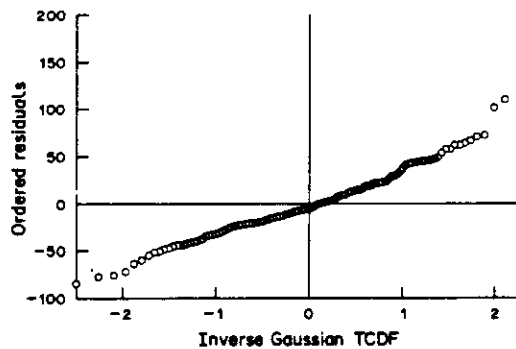


Fig. 1. An empirical QQ plot of the residuals from the ordinary least-squares line through the values for all the samples in the pool. Fitting was performed between the principal-component spectra and the reference cholesterol values. The ordered residuals from the fitting process are plotted on the ordinate, and the values of the inverse Gaussian theoretical cumulative distribution function (Φ^{-1}) are plotted on the abscissa. With the exception of the three outlying points at the top of the graph, all the points form a single line, suggesting a single group.

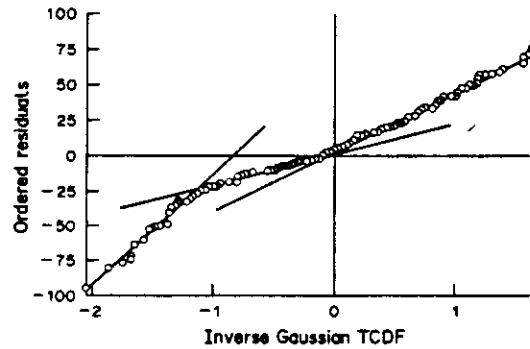


Fig. 2. An empirical QQ plot of the ordered residuals from a robust-fit line through the results for all the samples in the pool. Fitting was performed between the principal-component spectra and the reference cholesterol values. The outlying points at the top of the graph (see Fig. 1) are omitted by scaling to expand the range around the 3 lines that suggest 3 groups.

considered separately in attempts to develop a near-infrared calibration for cholesterol.

Leverage effects in least-squares regression force calibration lines to the best simultaneous fit through all of the groups present in a sample pool.¹⁷ This forced fit increases the error in application of the calibration to new samples, and may also obscure the fact that subgroups are present. Determining subgroup membership from sample spectra is analogous to detection of samples that have been tampered with, and the extent to which false samples can be detected largely determines the suitability of pattern-recognition methods for blood analysis.

Robust methods of regression,¹⁷ combined with quantile-quantile (QQ) plots of residuals from the fitting process, are useful in identifying the presence of subgroups in spectral data. Distinct patterns emerge in these plots, that are indicative of underlying structures in the data.¹⁹ In general, every straight line or line segment in the plot corresponds to a group of residuals. When the ordered residuals from a fitting process are plotted (on the ordinate) *vs.* the inverse Gaussian cumulative distribution function (on the abscissa) the slope of each segment in the plot is equal to the standard deviation of the corresponding group of residuals (see Fig. 1). Furthermore, the intercept of each line in the plot is equal to the mean of the corresponding group of residuals. Finally, curvature in lines indicates skew in the residuals.

The ARCANE multivariate robust regression,²⁰ based on a repeated-medians procedure²¹ and weighted multiple least-squares, was used to generate a cholesterol calibration line by use of all the test specimens obtained at the Fitness Fair. A QQ plot of the residuals from the ARCANE fit appears in Fig. 2. Some scatter appears in the points because they represent actual experimental data, hence the best-fit straight line to each cluster is also shown in order to delineate the groups better. We will call the three groups apparent in Fig. 2 groups 1, 2, and 3.

Table 1. Probability* of two cholesterol groups sharing the same mean level of a background constituent

	Albumin		Sodium	
	Group 2	Group 3	Group 2	Group 3
Group 1	88	23	11	2
Group 2	—	27	—	1
	Protein		Triglycerides	
	Group 2	Group 3	Group 2	Group 3
Group 1	82	13	1	24
Group 2	—	23	—	45
	Glucose			
	Group 1	Group 3		
Group 1	5	63		
Group 2	—	19		

*Probability given as % by a 2-tailed *t*-test.

If the sample matrix were the same for each of the cholesterol groups, the same cholesterol calibration would work for every group. Table 1 gives the percentage overlap ($P[\text{data} | H_0: \mu_1 = \mu_2]$, 2-tailed *t*-test) between two cholesterol groups with different albumin, sodium, protein, triglyceride, or glucose values. The sodium concentration appears to differentiate the cholesterol groups most effectively; for instance, the probability that groups 2 and 3 have the same Na^+ concentration is only 1%. Other differences can be found between the groups as well, e.g., although the overall pool membership was nearly 75% female, the membership of group 3 was 55% male. Interestingly, group 3 also had the lowest average cholesterol concentration of the three groups (177.5 mg/dl). These compositional differences do not lead to distinct spectroscopic groups when the *n*-wavelength spectra are projected as points into an *n*-dimensional hyperspace, however. For example, when the BEAST algorithm,¹⁸ a nonparametric method of measuring distances in hyperspace, is used, the mean distance of the group-2 spectral points from the center of the group-1 training set is 5.4 standard deviations (SDs). (Complete separation between two groups of similar size is ordinarily defined as 6 SDs.) The mean distance from the center of the group-1 points to the group-3 points is just 4.3 SDs. The combination of the spectral overlap of the cholesterol groups with the prediction error of the sodium calibration equation is more than enough to prevent assignment of a single sample to a single cholesterol group on the basis of a sodium concentration determined by near-infrared spectroscopy. Of course, sets of samples can often be assigned to a certain group even when the groups overlap.²² Unfortunately, while this set-assignment technique is useful for process-control applications and exploratory analyses of multidimensional distributions, its utility in serum analysis is limited because sets of samples from a given individual are seldom available.

If additional information vectors could be found that resulted in the formation of a space in which the groups were all separated by more than 6 SDs, fairly

accurate cholesterol determinations could still be performed by a pattern-recognition technique such as principal-component regression. Additional information vectors could be anything from use of new near-infrared wavelengths to results from completely separate analytical techniques such as chromatography. If a single serum sample could be unambiguously assigned to a specific cholesterol group, the prediction errors for the three groups would be 12.5 mg/dl (for group 1), 14.7 mg/dl (for group 2), and 10.7 mg/dl (for group 3). These prediction errors were obtained by dividing each of the three cholesterol groups in half, generating a near-infrared calibration equation with one half of each group, and using this to predict the cholesterol concentrations of the samples in the other half. While these are acceptable errors by themselves, they do not reflect the error introduced by improper assignment of a specific sample to a certain group. Combining the assignment error with the prediction error could be expected to raise the prediction error to the level obtained for calibration with the entire 162-sample pool.

Grouping effects similar to those observed in the cholesterol system have also been noted in triglyceride calibrations (see Fig. 3). Perhaps the worst problem encountered in performing cholesterol and triglyceride determinations with an 18-filter instrument, however, is that even after principal-axis transformation, the largest signal for both cholesterol and triglycerides is found on the same axis (the 8th). Cholesterol and triglyceride levels are not particularly correlated (see Fig. 4) and thus give substantial mutual interference at the analytical wavelengths used in our instrument (see Table 2). (It should be noted that the filters in the InfraAlyzer were not chosen for serum analysis in particular, but instead for their suitability for other analyses.) The standard error of the triglyceride calibration based on all 162 samples in the pool was 52 mg/dl.

Other blood constituents can also be determined with the near-infrared technique; for example, albumin and total protein determinations on the entire sample pool produced RSDs of 6 and 4%, respect-

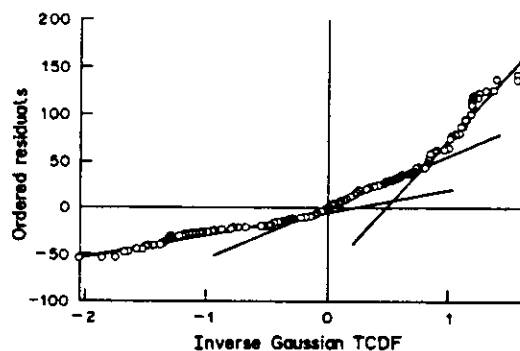


Fig. 3. An empirical QQ plot of the ordered residuals from a robust-fit line for all the samples in the pool. In this case, fitting was done between the principal-component spectra and the reference triglyceride values.

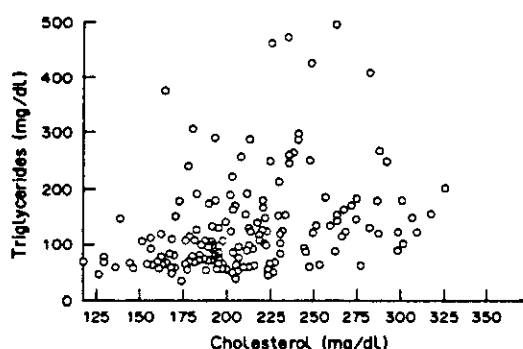


Fig. 4. A scatter-plot showing the relationship between the reference cholesterol and triglyceride values of all the samples in the pool ($r^2 = 0.14$).

ively. These constituents are present at fairly high concentrations, however. Unfortunately, there is a large number of rather similar, partially correlated serum constituents present at or near the levels of the lipoproteins, greatly complicating the near-infrared spectral analysis. In a practical sense, an instrument that produces a useful cholesterol result for only half the samples seems unlikely to gain wide clinical acceptance, even if the instrument is able to tell the operator that it cannot predict the cholesterol concentration in the remainder of the samples.

CONCLUSIONS

This study provides an indication of what still needs to be done in order to create a completely optical, reliable near-infrared method of serum analysis. Obviously, the subclustering observed inside training sets can be the result of a number of phenomena, including (1) a lack of sufficient training samples that are well-distributed across the range of sample variability, (2) a shortage of independent wavelengths available for describing the variability of the training set and its natural groups, and (3) noise factors, including not only random instrumental noise but also operator effects (such as work by different analysts on different days), temperature variations, lack of sample reproducibility (noted in other cholesterol determinations described earlier), and others. Whenever a system has a large number of independent variables (*e.g.*, chemical constituents), and monitoring is accomplished by using a small number of independent information vectors (*e.g.*, wavelengths), variations in the system that are

Table 2. Peak transmission wavelengths (nm) of the filters used in the InfraAnalyzer 400 (10-nm FWHM bandpass)

1654	1794	2116
1659	1841	2131
1678	1902	2158
1703	1934	2164
1713	2021	2255
1726	2087	2305

theoretically meaningful will appear to be random noise, complicating the problem of analysis.

Attempts to correlate group membership with known parameters of the sample pool indicated that the groups are different in a number of ways. To a certain extent these differences, too, could be a function of the limited number of samples available. Determining the source or sources of serum spectral subclustering will inevitably require the gathering of a larger sample pool.

The development of a broadly applicable near-infrared method of analyzing blood serum, using relatively small training sets (30 samples) and a few wavelengths (less than 20) in a standard filter instrument, appears to be confounded by the false-sample problem (the problem that arises in pattern-recognition when something unexpected appears in a sample). Nevertheless, considerable evidence indicates that accurate near-infrared determination of a number of blood constituents is possible. The next step is the acquisition of a high S/N training set with a thousand or more samples, the full near-infrared spectra of which will be analyzed at a few hundred wavelengths.

Acknowledgements—This work was supported in part by the National Science Foundation through Grant CHE 87-22639, by the Office of Naval Research, and by Miles Laboratories, Inc.

REFERENCES

1. *J. Am. Med. Assoc.*, 1985, **253**, 2080.
2. D. H. Blankenhorn, S. A. Nessim, R. L. Johnson, M. E. Sanmarco, S. P. Azen and L. Cashin-Hemphill, *ibid.*, 1987, **257**, 3233.
3. D. Edwards, *Science News*, 1987, **132**, 342.
4. J. M. Hoeg, R. E. Gregg and B. Brewer, *J. Am. Med. Assoc.*, 1986, **255**, 512.
5. *Science News*, 1987, 25 July, 63.
6. P. D. Hoagland and P. E. Pfeffer, *J. Agric. Food Chem.*, 1987, **35**, 316.
7. D. Edwards, *Science News*, 1987, **132**, 348.
8. L. Roberts, *Science*, 1987, **237**, 28.
9. D. Edwards, *Science News*, 1987, **132**, 376.
10. D. M. Hegsted and R. J. Nicolosi, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 6259.
11. M. Waldholz, *Wall Street Journal*, 1987, **68**, No. 249, 23.
12. L. Roberts, *Science*, 1987, **238**, 482.
13. H. J. Kisner, C. W. Brown and G. J. Kavarnos, *Anal. Chem.*, 1983, **55**, 1703.
14. E. Peuchant, C. Salles and R. Jensen, *ibid.*, 1987, **59**, 1816.
15. R. A. Lodder and G. M. Hieftje, *Appl. Spectrosc.*, 1988, **42**, 518.
16. O. H. Wheeler, *J. Chem. Educ.*, 1960, **37**, 234.
17. P. J. Rousseeuw, *J. Am. Stat. Assoc.*, 1984, **79**, 871.
18. R. A. Lodder, M. Selby and G. M. Hieftje, *Anal. Chem.*, 1987, **59**, 1921.
19. M. B. Wilk and R. Gnanadesikan, *Biometrika*, 1968, **55**, 1.
20. R. A. Lodder and G. M. Hieftje, *25th Colloquium Spectroscopicum Internationale*, Toronto, Canada, June 1987; paper H3.5.
21. A. F. Siegel, *Biometrika*, 1982, **69**, 242.
22. R. A. Lodder and G. M. Hieftje, *Appl. Spectrosc.*, 1988, in the press.