CLIN. CHEM. 23/9, 1744–1752 (1977)

Analytical Performance and Comparability of the Determination of Cholesterol by 12 Lipid-Research Clinics

Kenneth Lippel,[1] Susan Ahmed,[2] John J. Albers,[3] Paul Bachorik,[4] Gerald Cooper,[5] Ronald Helms,[2] and Jackie Williams[5]

Twelve Lipid-Research Clinic laboratories performed automated cholesterol analyses on four control-serum pools of known cholesterol concentration, using the Liebermann–Burchard reaction. The analyses were done during a two-year period, with the same standards, methodology, and quality-control procedures. Estimates of analytical bias, variability, and short- and long-term trends for each instrument and for the entire group of LRC instruments are presented. High accuracy, precision, and interlaboratory comparability were achieved through the rigorous standardization and control of the entire analytical procedure. The significance of these results for long-term collaborative studies is discussed. Individual laboratory biases averaged from 0.5 to 2.0% below Abell–Kendall reference values. Between-run variability was about equal to within-run variability and inter-laboratory variation was substantially less than intra-laboratory variation. The total standard deviation for all instruments was about 0.04 g/liter. Only 8–15% of this variation was due to differences between instruments. The between-instrument standard deviation ranged from 0.011 to 0.015 g/liter; the between-run, within-instrument standard deviation ranged from 0.023 to 0.030 g/liter; and within-run standard deviation ranged from 0.023 to 0.028 g/liter. The significance of the achieved results for long-term collaborative studies is discussed.

The variability in automated cholesterol determinations among laboratories due to differences in methodology, standards, and instruments is well recognized (*1*). In an effort to minimize interlaboratory and intralaboratory variability, the 12 Lipid-Research Clinic (LRC) laboratories used identical methods, standards, and internal control procedures, and were also monitored by the same blind external-surveillance process. The analytical methods and the internal-control procedures have been described in detail in Volume I of the LRC Program "Manual of Laboratory Operations" (*2*).

Rigorous adherence to a common protocol was required, to ensure that data gathered during the studies in the various LRC's met prescribed accuracy and precision limits and allowed assessment of interlaboratory and intralaboratory comparability over a long period of time. So that we could assess the accuracy, precision, and comparability of the laboratories, a three-part quality-control program was developed, consisting of an external surveillance program, an internal quality control program, and a split-sample quality-control program.

The results of analyses for cholesterol performed with AutoAnalyzer AAII (Technicon Corp., Tarrytown, N.Y. 10591) instruments by the 12 LRC laboratories for the internal quality-control pools (pools of known cholesterol concentration) during the first two years of operation are presented in this communication.

Materials and Methods

Primary Standards

The purity of commercial preparations of cholesterol was tested by the Lipid Standardization Laboratory of

[1] National Institutes of Health, National Heart, Lung, and Blood Institute, Lipid Metabolism Branch, Federal Building, Rm. 308, Bethesda, Md. 20014.

[2] University of North Carolina, Chapel Hill, N. C. 27514.

[3] University of Washington, Seattle, Wash. 98104.

[4] The Johns Hopkins University, Baltimore, Md. 21205.

[5] Center for Disease Control (CDC), Atlanta, Ga. 30333.

Members of the Laboratory Methods Committee who participated in this experiment included: Herman Arnett, Ph.D., Washington University, St. Louis, Mo.; H. Nordean Baker, Ph.D., Baylor College of Medicine, Houston, Tex.; Carl Breckenridge, Ph.D., University of Toronto, Toronto, Canada; Mary E. Dempsey, Ph.D., University of Minnesota, Minneapolis, Minn.; Gerald Hillerman, B.S., MT(ASCP), Oklahoma Medical Research Foundation, Oklahoma City, Okla.; Lillian Karlsson, Stanford University, Stanford, Calif.; Gerald Kessler, Ph.D., Washington University, St. Louis, Mo.; Kanta Kuba, M.S., University of Minnesota, Minneapolis, Minn.; Myron Kuchmak, Ph.D., Center for Disease Control, Atlanta, Ga.; Richard Muesing, Ph.D., George Washington University, Washington, D.C.; Maura Nava, Baylor College of Medicine, Houston, Tex.; Donald Puppione, Ph.D., University of California, San Diego, Calif.; Louis Smith, Ph.D., Baylor College of Medicine, Houston, Tex.; Paula Steiner, B.A., University of Cincinnati, Cincinnati, Ohio; Charles Stewart, Ph.D., Center for Disease Control, Atlanta, Ga.; John Turner, M.D., University of California, San Diego, Calif.; Russell Warnick, M.S., University of Washington, Seattle, Wash.; Stuart Weidman, Ph.D., Washington University, St. Louis, Mo.; Donald Wiebe, Ph.D., University of Iowa, Iowa City, Iowa; Carol Winn, Center for Disease Control, Atlanta, Ga.; and Peter Wood, Ph.D., D.Sc., Stanford University, Stanford, Calif.

Received April 23, 1977; accepted May 20, 1977.

CDC by thin-layer and gas–liquid chromatography and spectrophotometry (*3*). The cholesterol was used to prepare sets of solutions of different concentrations in isopropanol (0.50, 1.00, 2.00, 3.00, and 4.00 g/liter) for distribution to LRC laboratories if the purity was equal to the purity of National Bureau of Standards reference cholesterol (99.9%).

Cholesterol Analysis with the AAII

The procedures used by the LRC laboratories for cholesterol analysis have been described in detail (*2*). All analyses were performed on 20-fold diluted (as compared to the sample volume) isopropanol extracts of the samples. The extracts were treated with a zeolite mixture to remove interfering substances and then analyzed with the AAII. Standards prepared in isopropanol/water (95/5 by vol) were similarly treated.

For the cholesterol analyses on the AAII, the Liebermann–Burchard reaction was used. Samples were analyzed at a rate of 50 samples/h with an 8/1 sample/wash ratio. The linearity of the instrumental response to standards was evaluated by running an ascending series of standards (1.00–4.00 g/liter) at the beginning of each day's run. Recorder response was adjusted to 60% of full-scale deflection with a 3.00 g/liter standard and readjusted after every sample tray (40 samples) to compensate for baseline drift and scale expansion. Results were printed on a digital printer tape. Duplicate extracts of a serum calibrator were always included in the first tray of samples in each day's run.

Calibration of the AAII

On a molar basis, cholesterol esters develop more color with the Liebermann–Burchard reaction than does unesterified cholesterol (*4*). Since 60–70% of the cholesterol in human plasma is esterified, and the analytical procedure used does not include hydrolysis of the cholesterol esters, the apparent cholesterol concentrations are falsely high when referred to unesterified cholesterol standards. It was therefore necessary to use a calibration procedure to correct AAII-assayed cholesterol values so that they corresponded to Abell–Kendall cholesterol values (*5*).

The calibration procedure used daily in the LRC laboratories has been described in detail (*2*). The rationale for and validation of the serum calibration technique will be described separately (manuscript in preparation). Briefly, the serum calibration pool was analyzed in each day's run vs. unesterified cholesterol standards. Control limits for the serum calibration pool were established after analyzing two extracts of the pool on each of 20 consecutive runs. The control limits were recalculated after 50 consecutive runs. The pool was subsequently used at the beginning of each run to calibrate the Liebermann–Burchard cholesterol values. Corrected cholesterol values for individual samples were calculated by multiplying the apparent AAII cholesterol value by a correction factor. The correction factor was based on the relationship between the apparent cholesterol value and the reference cholesterol value of the serum calibration pool as follows:

(*a*) Correction factor =

$$\frac{\text{(reference cholesterol value of serum calibration pool)}}{\text{(apparent AAII cholesterol value of serum calibration pool)}}$$

(*b*) Corrected cholesterol value (for individual sample) = (correction factor) · (apparent AAII cholesterol value of the individual sample)

Before the run was continued, the following two criteria had to be met for that run: (*a*) the instrument response to standards had to be linear, and (*b*) the response of the serum calibrator relative to the standards had to be within control limits.

Analysis of Control Pools and Serum Calibrator Pools

A total of four serum pools of known cholesterol content was analyzed by all LRC laboratories during 24 months; two pools were analyzed daily in any one period. These pools were used for bench control of cholesterol measurements. The pools were prepared by the Lipid Standardization Laboratory at the CDC, from sterile human sera. Cholesterol-rich lipoprotein fractions, made from a part of the collected human sera, were added to other aliquots of the pooled human sera to make serum pools of different cholesterol concentrations (*6*). The pools were then dispensed into glass Wheaton vials. The vials were sealed and stored frozen in an upright position at −20 °C, and tested for homogeneity by automated cholesterol analysis of samples randomly selected from the entire pool. Homogeneous pools were subsequently labeled with a reference value (target value) by replicate analyses performed by the method of Abell et al. (*5*). The precision of the determinations was such that the width of a 95% confidence interval was less than 1% of the target value. Frozen aliquots were mailed to the LRC laboratories and stored at −15 °C until analyzed. Additional serum pools (SERCAL) were used as AAII calibrators. Calibrator pools were prepared from pooled human sera and labeled by the method of Abell et al. (*5*). The precision of the calibrator determinations was such that the width of a 95% confidence interval was less than 0.5% of the target value. The control pools were monitored quarterly for stability during the period of their use, and the calibrator pools were monitored weekly by analyses with the reference method of Abell et al. (*5*).

Standardization of Lipid Determinations

Each of the LRC laboratories met the requirements of a lipid standardization program before becoming operational. Phase I of the program, the evaluation phase, consisted of the analysis of three pools of unknown and two pools of known cholesterol concentration. Six analyses were performed on samples from each pool in each of four runs during two to four weeks. A single analysis was performed on each sample. Any

Table 1. Criteria for Accuracy and Precision for Various Cholesterol Concentrations

Cholesterol concentration	Deviation from target value (T.V.) T.V. ± 5% of T.V.	Overall standard deviation[a]
	mg/100 ml	
100–199	5.00–9.95	<7.00
200–299	10.00–14.95	<8.00
300–399	15.00–19.95	<9.00
400–499	20.00–24.95	<10.00

[a] Overall standard deviation is the total of the within day and day-to-day components of variance derived by the analysis of variance technique. Nine pools at various concentrations were used in the surveillance program. Thirty-two samples from each pool were run each year. The limits in Table 1 were used annually to evaluate the results from each pool.

problems related to the analysis were identified and resolved before proceeding with Phase II.

Phase II, the standardization phase, was a more intensive evaluation of the laboratory's performance. Its purpose was to determine precision and accuracy more thoroughly and to document that analyses remained stable over a longer period of time. Three sets of frozen serum samples (a total of 320 samples with unknown cholesterol concentrations) were analyzed on 32 different days, over a period of 8 to 16 weeks. During this time, results had to be within acceptable limits of accuracy and precision. Upon satisfactory completion of Phase II, the laboratory was designated as "standardized" and permitted to start analyses on LRC specimens. Table 1 summarizes the criteria used for standardization. The data reported in this communication were collected after all laboratories were standardized.

Subsequent to being standardized, 24 unknown samples were analyzed each month for the duration of the program (Phase III). Six samples were analyzed in one run each week for four weeks of the month. During this surveillance phase, laboratories meeting the criteria specified in Table 1 were redesignated annually as "standardized." All 12 LRC laboratories remained standardized during the first two years of internal surveillance reported in this communication.

Data and Results

Data Source

Four samples from each of two serum pools of known cholesterol content were analyzed in each day's run made by an LRC laboratory. Before using a pool for quality control, the laboratory established its own internal quality-control limits for the pool (2). At the end of each run, the results of the four sample determinations on a pool were examined in relation to these internal limits, to determine whether the day's run was acceptable ("in control"). Samples from an "out of control" run were reanalyzed.

The internal quality-control data (data from pools of known cholesterol content) were collected from each instrument from the initial date of operation. Early data, tabulated before standardization, primarily reflect startup difficulties. The results reported here are based upon data collected from each instrument's first run in Phase III of standardization through the end of December 1974. The first Phase III run was made on March 8, 1973. The "Red" and "Blue" pools (named after the colors of the vials' caps) were the initial pools used throughout the LRC Program; then the "Q1" and "Q2" pools were phased in, beginning in September 1973, as the samples from the "Red" and "Blue" pools were exhausted. The time period of overlap between the two sets of pools ranged from zero to three months. The results presented here are based on all quality-control data from all runs at all laboratories except for extreme outliers from runs declared "out of control" by a laboratory.

Systematic Differences from Target Value

The very accurate manual Abell–Kendall (MAK) "target values" established by CDC permit examination of the data for both systematic and random error sources.

Systematic errors were measured by a bias statistic, i.e., the difference between an average and the target value; the magnitude of random errors was measured by a variance or standard deviation. Systematic errors may arise from three sources: type of instrument, individual instruments, and serum pools.

Systematic Type of Instrument Bias

The first line of Table 2 summarizes the average biases exhibited by the AAII instruments for the four quality-control pools. Except for the Blue pool, the average AAII biases are negative, indicating lower-than-target-value determinations, on the average. The "% Deviation" column contains averages of percent deviations, defined as:

$$\% \text{ deviation} = \frac{\text{instrument value} - \text{target value}}{\text{target value}} \times 100.$$

The percent deviation column shows the AAII instruments averaged 1.3% below target values. The standard errors in Table 2 are quite small, primarily as a result of the large sample sizes, and indicate that the biases are estimated with a high degree of precision. Also, except for results for the Blue pool, the bias estimates are all significantly different from zero.

Individual Instrument Biases

Table 2 also contains bias estimates and their standard errors for each instrument and pool combination and for percent deviation averaged over pools for each instrument. These results are plotted in Figure 1.

Small individual biases, which varied from pool to pool, were observed for each instrument. For the Red pool (target value = 1.66 g/liter), the biases ranged from −26 to +19 mg/liter, a range of 45 mg/liter about the all-AAII average of −11 mg/liter. The Q1 pool (target value = 1.71 g/liter) exhibited more negative biases

Table 2. Overall Bias of Instruments and Types of Instruments

Instrument No.	Red pool (1.66[a])			Q1 pool (1.71)			Blue pool (2.63)			Q2 pool (2.51)			% deviation (all pools)		
	Bias	SE	n	Bias	SE	n	Bias	SE	n	Bias	SE	n	% bias	SE	n
	g/liter			g/liter			g/liter			g/liter			%		
All	−0.011	0.0009	1965	−0.023	0.0004	7788	0.002	0.0009	1942	−0.042	0.0005	7807	−1.3	0.01	8766
1	−0.013	0.0088	24	−0.013	0.0012	846	0.013	0.0163	23	−0.042	0.0016	842	−1.2	0.05	1735
2	0.017	0.0026	168	−0.004	0.0010	1144	0.024	0.0033	168	−0.039	0.0014	1141	−0.7	0.04	2621
3	--	--	--	−0.042	0.0016	504	--	--	--	−0.047	0.0016	503	−1.8	0.06	1007
4	−0.010	0.0029	197	−0.031	0.0009	1116	−0.002	0.0030	198	−0.048	0.0011	1114	−1.6	0.03	2625
5	0.002	0.0028	102	−0.015	0.0018	296	0.007	0.0034	94	−0.037	0.0021	304	−0.8	0.06	796
6	−0.007	0.0022	224	−0.018	0.0019	360	0.022	0.0025	216	−0.020	0.0020	368	−0.6	0.06	1168
7	−0.005	0.0026	52	−0.023	0.0020	331	0.011	0.0030	148	−0.031	0.0022	336	−0.9	0.06	967
8	−0.007	0.0048	60	−0.036	0.0016	355	−0.004	0.0046	56	−0.041	0.0019	354	−1.7	0.06	825
9	−0.014	0.0018	288	−0.046	0.0013	480	−0.017	0.0015	284	−0.070	0.0014	479	−2.0	0.04	1531
10	−0.007	0.0039	125	−0.008	0.0013	543	0.001	0.0044	136	−0.016	0.0016	547	−0.5	0.05	1351
11	+0.019	0.0017	7	−0.031	0.0017	356	−0.045	−0.0106	8	−0.052	0.0021	356	−1.9	0.07	727
12	−0.026	0.0024	320	−0.031	0.0010	835	0.007	0.0020	312	−0.047	0.0012	836	−1.5	0.04	2303
13	−0.020	0.0021	298	−0.026	0.0015	622	−0.016	0.0023	299	−0.044	0.0016	627	−1.4	0.05	1846

[a] The numbers in parentheses are target values for the pools established at CDC by the Abell-Kendall method.

ranging from −46 to −4 mg/liter, a range of 42 mg/liter about the all AAII average of −23 mg/liter. The Blue and Q2 pools have higher cholesterol concentrations (Blue target value = 2.63 g/liter, Q2 target value = 2.51 g/liter). For the Blue pool the AAII instruments' biases ranged from −17 to +24 mg/liter. The Q2 pool again exhibited more negative biases, ranging from −70 mg/liter (the largest absolute bias of any AAII) to −16 mg/liter, a range of 54 mg/liter about an all-AAII average bias of −42 mg/liter.

The average (over pools) percent deviations (Table 2) vary from −2.0 to −0.5%, a range of 1.5%. The instrument with the largest average negative bias, number 9, is 0.7% below the all-AAII average of −1.3% and the instrument with the largest average positive bias, Number 10, is 0.8% above the average.

Serum Pool Biases

The AAII results reveal strikingly clear pool-related biases; the Blue pool resulted in the most positive biases (averaging very close to zero for all AAII's combined), followed in order by the Red, Q1, and Q2 pools. There is only one exception to this ordering: the instrument 9 bias of the Red pool is 3 mg/liter more positive than the Blue pool bias. The two sets of pools (Red, Blue vs. Q1, Q2) exhibited different levels of bias.

The percents bias in the Red and Blue pools were −0.66 and +0.08%, and the percents bias in the Q1 and Q2 pools were −1.34 and −1.67%. The reference cholesterol method exhibits cycles of 1 to 2% over variable time periods that apparently depend on such factors as the analyst's technique, equipment variability, and reagent response. Examination of the quality-control charts for the reference method for the period during which quality-control target values were established for the Q1 and Q2 pools revealed that the reference method possessed approximately a 1.5% positive bias when compared with the average values over several years. The attainable coefficient of variation with the reference cholesterol method as performed in the CDC Lipid Standardization Laboratory approximates 1 to 1.5%, independent of the bias cycle. The combined effect of this variation and the positive bias in cycle could partly explain the difference from target values shown by the two different sets of quality-control pools. Also, very small differences in instrument response (resulting from variable cholesterol/cholesterol ester ratios in the different control pools) may have been detected as a consequence of the large number of replicate analyses used to estimate the bias of each pool.

At the same time that the Q1 and Q2 pool target values were set, the target value of the "SER 299" serum calibrator pool (a pool used to calibrate the AAII ma-

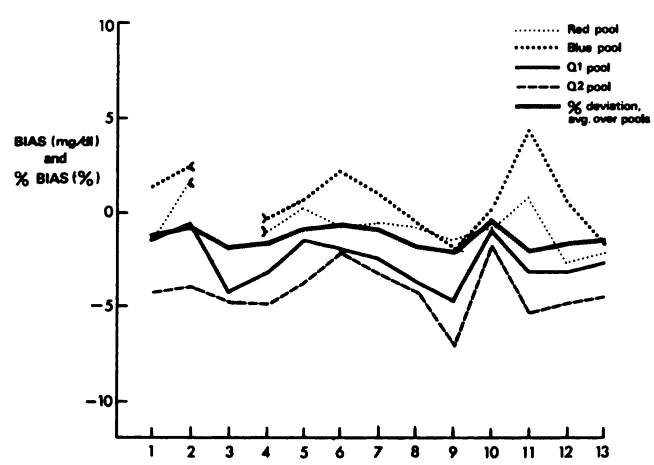

Fig. 1. Biases of instruments and quality-control pools

Table 3. Variance Components Analysis for Each Instrument and Quality Control Pool

Instr. No.	Source of Variation	Red pool (1.66[a])			Q1 pool (1.71)			Blue pool (2.63)			Q2 pool (2.51)			% deviation (all pools)		
		df	SD g/liter	Component %	df	SD g/liter	Component %	df	SD g/liter	Component %	df	SD g/liter	Component %	df	SD %	Component %
1	Total	23	0.043	100.0	845	0.034	100.0	22	0.078	100.0	841	0.046	100.0	1734	2.00	100.0
	Between runs	5	0.024	30.0	210	0.024	49.9	5	0.057	53.7	210	0.035	59.4	433	1.50	56.2
	Within run	18	0.036	70.0	635	0.024	50.1	17	0.053	46.3	631	0.029	40.6	1301	1.32	43.8
2	Total	167	0.047	100.0	1143	0.035	100.0	167	0.042	100.0	1140	0.047	100.0	2620	2.19	100.0
	Between runs	41	0.034	52.0	285	0.025	53.7	41	0.026	38.4	285	0.037	59.3	655	1.72	62.2
	Within run	126	0.033	48.0	858	0.024	46.3	126	0.033	61.6	855	0.030	40.1	1965	1.35	37.8
3	Total	---	-----	-----	503	0.036	100.0	---	-----	-----	502	0.037	100.0	1006	1.82	100.0
	Between runs	---	-----	-----	125	0.029	67.2	---	-----	-----	125	0.028	58.8	251	1.47	65.2
	Within run	---	-----	-----	378	0.021	32.8	---	-----	-----	377	0.024	41.2	755	1.08	34.8
4	Total	196	0.041	100.0	1115	0.030	100.0	197	0.043	100.0	1113	0.035	100.0	2624	1.76	100.0
	Between runs	48	0.028	48.5	276	0.019	37.5	48	0.028	43.1	276	0.022	39.3	651	1.20	45.9
	Within run	148	0.029	51.5	839	0.024	62.5	149	0.032	56.9	837	0.027	60.7	1973	1.30	54.1
5	Total	101	0.029	100.0	295	0.031	100.0	93	0.033	100.0	303	0.037	100.0	795	1.73	100.0
	Between runs	24	0.013	20.9	74	0.017	30.9	23	0.006	3.4	76	0.025	45.7	200	1.11	41.2
	Within run	77	0.025	79.1	221	0.026	69.1	70	0.033	96.6	227	0.027	54.3	595	1.33	58.8
6	Total	223	0.032	100.0	359	0.037	100.0	215	0.036	100.0	367	0.039	100.0	1167	1.93	100.0
	Between runs	53	0.022	44.7	89	0.029	60.4	53	0.030	70.2	91	0.032	68.9	289	1.55	64.6
	Within run	170	0.024	55.3	270	0.023	39.6	162	0.020	29.8	276	0.022	31.1	878	1.15	35.4
7	Total	151	0.032	100.0	330	0.037	100.0	147	0.037	100.0	335	0.040	100.0	966	1.95	100.0
	Between runs	36	0.025	63.1	82	0.031	67.6	36	0.032	76.3	83	0.037	71.8	240	1.65	72.0
	Within run	115	0.019	36.9	248	0.021	32.4	111	0.018	23.7	252	0.021	28.2	726	1.03	28.0
8	Total	59	0.037	100.0	354	0.031	100.0	55	0.034	100.0	353	0.035	100.0	824	1.74	100.0
	Between runs	14	0.031	69.7	87	0.020	43.4	13	0.023	45.9	87	0.026	55.6	204	1.30	56.2
	Within run	45	0.021	30.3	267	0.023	56.6	42	0.025	54.1	266	0.023	44.4	620	1.15	43.8
9	Total	287	0.030	100.0	479	0.028	100.0	283	0.025	100.0	478	0.030	100.0	1530	1.73	100.0
	Between runs	71	0.016	28.4	119	0.021	57.1	70	0.018	48.8	119	0.022	54.1	382	1.37	63.1
	Within run	216	0.026	71.6	360	0.018	42.9	213	0.018	51.2	359	0.020	45.9	1148	1.05	36.9
10	Total	124	0.043	100.0	542	0.031	100.0	135	0.052	100.0	546	0.037	100.0	1350	1.80	100.0
	Between runs	31	0.034	62.3	136	0.022	50.6	33	0.043	69.8	136	0.030	63.6	339	1.38	58.9
	Within run	93	0.027	37.7	406	0.022	49.4	102	0.028	30.2	410	0.022	36.4	1011	1.16	41.1
11	Total	6	0.028	100.0	355	0.033	100.0	7	0.030	100.0	355	0.039	100.0	726	1.80	100.0
	Between runs	1	0.013	22.4	88	0.025	60.6	1	0.011	12.6	88	0.032	68.4	181	1.45	65.3
	Within run	5	0.025	77.6	267	0.021	39.4	6	0.028	87.4	267	0.022	31.6	545	1.06	34.7
12	Total	319	0.042	100.0	834	0.029	100.0	311	0.036	100.0	835	0.034	100.0	2302	1.83	100.0
	Between runs	79	0.035	70.3	208	0.022	57.6	77	0.029	63.9	208	0.028	65.6	575	1.53	69.4
	Within run	240	0.023	29.7	626	0.019	42.4	234	0.022	36.1	627	0.020	34.4	1727	1.01	30.6
13	Total	297	0.037	100.0	621	0.036	100.0	298	0.040	100.0	626	0.040	100.0	1845	1.93	100.0
	Between runs	74	0.009	5.5	155	0.017	20.8	74	0.020	24.6	156	0.023	32.9	462	0.94	23.9
	Within run	223	0.036	94.5	466	0.032	79.2	224	0.035	75.4	470	0.033	67.1	1383	1.68	76.1

[a]The numbers in parentheses are target values for the pools established at CDC by the Abell-Kendall method.

chines daily) was set at 2.99 g/liter. Later, more extensive Abell–Kendall analyses established a value of 2.93 g/liter for that pool. (All serum-calibrator related results reported here are based upon the 2.93 g/liter value.) If the Q1 and Q2 pool target values were too high, and if the discrepancy was proportional to the serum calibrator difference, the appropriate target values would have been 1.676 (vs. 1.71) g/liter and 2.460 (vs. 2.51) g/liter, respectively. With these target values the all-AAII average biases would have been 11 mg/liter (±0.66%) and 8 mg/liter (±0.33%), respectively, values that are somewhat closer to the average biases realized with the Red and Blue pools.

Random Variability

Nested analysis of variance estimates (7) of between-run and within-run variance components were computed separately for each pool–instrument combination (Table 3). The same analysis was applied to the percent deviation variable, defined in the preceding section, which included results from all four pools.

The "SV" column in Table 3 indicates the source of variation (as used in ANOVA tables). Lines identified as "Between Runs" contain statistics related to run-to-run variability. Lines identified as "Within Run" contain within-run variability statistics. "Total" lines are explained below. The "df" column refers to the degrees of freedom for each source of variation, as would be found in an ANOVA table. A large "df" indicates great statistical precision of the estimates of the standard deviations.

Table 3 presents three standard deviations, the square roots of the corresponding variance components, for each instrument-pool combination. The "within-run" standard deviation estimate is the pooled (over runs) estimate of within-run variability, i.e., a measure of the variability among the four essentially identical samples in one run. The "between-runs" standard deviation estimate is a measure of run-to-run variability. The "total" standard deviation is the square root of the sum of the "between runs" and "within runs" variance components and is an estimate of the standard deviation of a single determination of a quality-control sample from the specified instrument and pool. The "Variance Component %" column indicates the relative magnitudes of the between-runs and within-runs variance components; the "Variance Component %" is computed at 100 × (VC/TVC) where VC is the between- or within-run component and TVC is the "Total" variance components.

Total Variances and Standard Deviations

Better precision over the four pools was consistently observed for AAII Instruments 5, 9, and 11 than for the other AAII's. Poor precision with the Blue pool was observed for Instrument 1. However, only six runs were made with the pool, and the large total standard deviation estimate may have been due to poor statistical precision resulting from few degrees of freedom. Very good precision was generally achieved with all AAII instruments. Most of the total standard deviations are less than 40 mg/liter.

One expects the precision (standard deviation) of an instrument to be smaller for pools with lower cholesterol concentrations. As shown in Table 3, 12 of 13 instruments had smaller total standard deviations for the Q1 pool (target value = 1.71 g/liter) than for the Q2 pool (target value = 2.51 g/liter), and 7 of 11 instruments had smaller total standard deviations for the Red (1.66 g/liter) than for the Blue (2.63 g/liter) pool. The Red and Blue pools were used simultaneously and, at later dates, the Q1 and Q2 pools were used simultaneously. Comparison of results from the Red and Q1 pools or the Blue and Q2 pools would not be appropriate, because widely different dates are involved.

Overall, of 50 CV's for instrument-pool combinations shown in Figure 2, none had a total CV greater than 3.0%, 11 had total CV's between 2.0% and 3.0%, and 39 had total CV's of 2.0% or lower. All 13 instruments had a total CV of less than 2.0% for the Q2 pool and eight of 12 had total CV's of less than 2.0% for the Q1 pool.

Within-Run Variability

The within-run standard deviation estimates are summarized in Table 3 in lines labeled "Within Run," in the "SV" column.

The within-run standard deviations tend to follow the same pattern as the total standard deviations. All instruments performed well with all pools with the possible exception of instrument 1 on the Red and Blue pools. Again, this is probably due to the small sample sizes. For the Blue and Q2 pools (target values 2.63 and 2.51 g/liter, respectively) the within-run standard deviations generally fell in the 20 to 33 mg/liter range. Similar results were observed for the Red pool (target value = 1.66 mg/liter). For the Q1 pool, the within-run

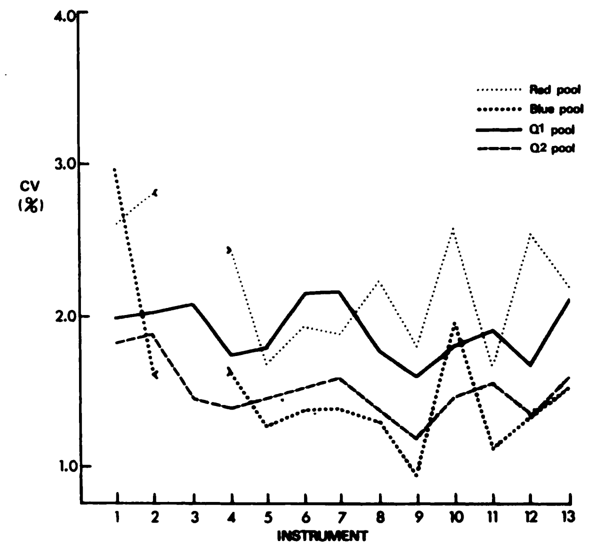

Fig. 2. Coefficient of variation for each instrument-pool combination

standard deviations were generally less than 25 mg/liter.

Between-Run Variability

The standard deviations representing run-to-run variability are shown in the "Between Runs" lines of Table 3. Between-run standard deviations follow the same pattern as the within-run and total standard deviations. Small run-to-run variability was generally observed with all four quality-control pools on all AAII instruments.

Instrument Comparability

The results of a nested variance components analysis in which data from all AAII instruments were pooled for analysis are presented in Table 4. These results confirm the conclusions drawn above. The within-run variance components of AAII's are comparable for all pools. Similar conclusions apply to the "Between Runs" variance components.

The total AAII standard deviation for all instruments is 30 to 40 mg/liter. Only 8 to 15% of this variation is due to differences between instruments. Most of the variability occurs within and between days on a single instrument. The variability was presumably due to the different batches of reagent and tubing as well as slight changes in instrument response.

Trends

Short-term trends. Short-term trending can be examined through autocorrelations. Lag 1, lag 2, lag 3, and lag 10 autocorrelations were computed from daily means for each instrument/pool combination. A "lag d" correlation is the correlation between values "d" days apart. A high "lag d" correlation indicates that if a machine is high on a given day, it also tends to be high "d" days later.

Table 4. Variance Components Analysis for Each Quality Control Pool—All Instruments Combined

Source of Variation	Red pool (1.66[a])			Q1 pool (1.71)			Blue pool (2.63)			Q2 pool (2.51)			% deviation (all pools)		
	df	SD g/liter	Component %	df	SD g/liter	Component %	df	SD g/liter	Component %	df	SD g/liter	Component %	df	SD %	Component %
Total	1964	0.039	100.0	7787	0.035	100.0	1941	0.041	100.0	7806	0.041	100.0	19,501	1.97	100.0
Among instruments	11	0.011	8.3	12	0.014	14.6	11	0.015	12.8	12	0.012	8.8	49	0.83	17.9
Between runs	477	0.025	42.5	1934	0.023	42.4	474	0.027	43.6	1940	0.030	51.3	4,825	1.28	42.0
Within run	1476	0.028	49.2	5841	0.023	43.0	1456	0.027	43.6	5854	0.026	40.0	14,627	1.25	40.1

[a]The numbers in parentheses are target values for the pools established at CDC by the Abell-Kendall method.

Table 5. Lag Correlations between Consecutive Runs, Every Second Run, Every Third Run, and Every 10th Run for All AA-II's

	Red pool				Q1 pool				Blue pool				Q2 pool			
	Lag correlations				Lag correlations				Lag correlations				Lag correlations			
	Lag 1	Lag 2	Lag 3	Lag 10	Lag 1	Lag 2	Lag 3	Lag 10	Lag 1	Lag 2	Lag 3	Lag 10	Lag 1	Lag 2	Lag 3	Lag 10
Mean	0.24	0.11	0.11	0.05	0.26	0.19	0.21	0.11	0.17	0.07	0.01	0.02	0.27	0.24	0.19	0.09
SD	0.20	0.21	0.25	0.12	0.15	0.19	0.15	0.09	0.21	0.23	0.18	0.29	0.16	0.16	0.16	0.16

Table 5 contains the average (over instruments) lag 1, lag 2, lag 3, and lag 10 autocorrelations for each pool and the corresponding standard deviations. The relatively large standard deviations indicate considerable variability among the instruments in their various autocorrelations. Most 10-day correlations were less than 0.20, indicating only a slight tendency towards trends as long as 10 days.

Longer-term trends. Longer trends may be examined by standard time-series techniques, but the following approach leads to more easily interpretable results. The mean of all available data was calculated for each instrument/pool combination. A "high series" was then defined to be that number of consecutive runs for which the daily run mean was greater than the instrument's overall mean for the pool. Similarly, a "low series" was defined as the number of consecutive runs below the instrument's overall mean for the pool. If there were no trending, the probability that a given run's mean would fall above (below) the instrument mean would be very close to ½, independent of the results of any previous run. The probability of two consecutive high (low) runs would be $(\frac{1}{2})^2 = \frac{1}{4}$, the probability of three consecutive high (low) runs would be ⅛, etc. Examining the frequencies of high and low series and comparing observed frequencies with theoretical non-trend frequencies can reveal the extent of trending. A long series, whether high or low, reveals the presence of a trend by the instrument.

Table 6 contains a summary of the distribution of lengths of series for low and high series combined. The first column of the table indicates the length of the series, in runs. Lengths greater than 10 runs have been grouped into four groups: 11–20, 21–30, 31–40, and 41–50 runs. In addition, statistics were compiled for "long series," defined here as any series more than nine runs long.

The second column of Table 6 presents the theoretical frequencies, in percentages, for *non-trend* situations. The instruments are known to experience trends; the theoretical non-trend frequencies are simply presented as a baseline for comparison.

When results for all instruments and pools are combined, ignoring whether a series is high or low, the percentage of one-run "series" is very close to 50%, the theoretical non-trend value. The frequencies of series of lengths 3–10 are also reasonably close to the theoretical non-trend values. But only 22.0% of the series were of length two runs, compared to 25% theoretical

Table 6. Distribution of Length of Series (Trends)

Series length (runs) (1)	Theoretical non-trend % (2)	Frequency (3)	Percent (4)
1	50.0	981	49.1
2	25.0	439	22.0
3	12.5	231	11.6
4	6.3	109	5.5
5	3.1	78	3.9
6	1.6	50	2.5
7	0.8	26	1.3
8	0.4	18	0.9
9	0.2	18	0.9
10	0.1	11	0.6
11–20	0.098	32	1.6
21–30	1×10^{-4}	2	0.1
31–40	1×10^{-7}	1	0.1
41–50	1×10^{-12}	1	0.1
Total long series			
10–50	0.20	36	1.8
Total series		1997	100.0

Table 7. Instruments Experiencing Long Series (Lengths Greater than Nine Runs)

Instrument number	Total no. of series	Total no. of long series	Percent of long series	Long series lengths and high(+)/low(−) indicators
10	94	8	8.5	11,15,16,17,19,19,23,45 + + + − + + − +
6	104	5	4.8	10,10,12,14,16 − − − + +
2	234	10	4.3	10,11,11,12,13,13,17,17,19,31 − + − + + − − − + +
13	188	6	3.2	10,10,11,12,14,16 − − − + + −
1	195	4	2.1	10,10,11,14 + − − +
12	245	5	2.0	10,11,12,14,22 − + − + −
7	106	2	1.9	12,12 + −
3	107	2	1.9	10,11 − −
11	73	1	1.4	14 −
4	313	4	1.3	10,10,12,12 + + − −
9	160	1	0.6	12 −

non-trend, and the frequencies of longer series are substantially higher than one would expect in a non-trend situation, i.e., the results reflect a tendency to trend by some of the instruments.

Examination of results for individual instruments, too voluminous to be shown here, verifies the interpretation given above for the most of the instrument/pool combinations. Two instruments had no long series: AAII Instrument 5 had no series longer than six runs in a series of 94 total runs, and AAII Instrument 8 had no series longer than nine runs in a total of 84 runs. (In both cases the total numbers of runs are counted separately for different pools, so the actual number of analytical runs is about half the number given.)

On the other end of the distribution a few instruments had some very long series. As shown in Table 7, AAII Instruments, 2, 6, and 10 exhibited long-term trends, more or less equally divided between high and low series.

Discussion

Previous surveys have demonstrated that methodology, standardization technique, type of standard, reagent source, and the quality of control sera affect the accuracy and precision of interlaboratory cholesterol measurements (*8, 9*) and that interlaboratory performance is significantly improved when a common standardization material is used (*9*).

The LRC laboratories attempted to minimize bias and variability of cholesterol analyses by establishing procedures whereby all participating laboratories used identical instrumentation, aliquots of common primary standard solutions, aliquots of common bench control serum pools, and, for cholesterol analyses performed with the Liebermann–Burchard reaction, common serum calibration pools. The cholesterol concentrations of the bench control pools and the serum calibrator pools were established by reference methodology. The LRC laboratories also used the same sequence of standards and common procedures for instrument start-up, instrument calibration, and analysis.

The average cholesterol bias for the 12 laboratories was −1.3% and ranged from −0.5 to −2.0%. The degree of bias varied from pool to pool, ranging from −1.67% on the Q2 pool to +0.08% on the Blue pool. The small individual biases exhibited were revealed only by the many analyses performed. The small but definite average negative bias may be due in part to inherent chemical and physical differences in the measurement of cholesterol by the reference method and automated systems, in part to the variability of the reference method itself, and in part to the small changes in physical properties that may occur in reference sera as they age.

Surprisingly, the run-to-run variance components observed for the LRC cholesterol analyses compared favorably with the within-run variability. Many factors can significantly affect an instrument between runs, but typically they can have only a minor or negligible effect within the course of a run. The converse is not true:

virtually any factor that can affect instrument precision within a run can also affect it between runs. As a consequence, run-to-run variance components are frequently substantially larger than within-run components. However, the results in Table 3 show that run-to-run variability was tightly controlled for most of the instrument/pool combinations, since the run-to-run variability was less than 50% of the total (i.e., less than the within-run variability) in many cases. This surprisingly good control of run-to-run variability was attributed to strict adherence to analytical protocol.

Even more impressive than the small run-to-run variability is the small instrument-to-instrument variability (Table 4). The AAII instrument-to-instrument variability was substantially smaller than the average within-run and among-run sample-to-sample variability for every quality-control pool. This demonstrates impressive comparability among instruments and enables cholesterol results from the various laboratories to be pooled in the collaborative study.

The small lag correlations between every tenth run (lag 10) indicate that trend problems were corrected quickly. Runs 10 days apart are only slightly correlated.

Cycles of different lengths within the allowable quality-control limits can be detected by inspection of the LRC laboratory quality-control charts. Individual instruments showed different tendencies toward trends, with great variability in the duration of a cycle and some variability in the distance from the overall mean. This tendency to cycle is also exhibited in the distribution of series of runs of various lengths (Table 5). One-run series occurred in approximately the non-trend frequency. Series of length 2 occurred less often than expected in a non-trend situation and series of length greater than 9 occurred more often than expected. possible explanation for this observed pattern is the following: when an instrument is operating very close to the overall mean, the runs occur in approximately the theoretical non-trend frequencies, but as an instrument's bias swings on either side of zero the series tend to get longer, interrupted from time to time by a single daily mean on the other side of the long-term mean. These occasional single "other side" daily means terminate the series which is on the same side as the bias and increase the frequency of one-run series. Thus, over a long period of time, the one-run series have approximately the non-trend frequency, but the two-run series occur less frequently, being replaced by longer series.

As noted earlier, trending or cycling is almost unavoidable with these instruments, because of batch-to-batch variation in tubing and reagents. One objective of our quality-control program was to maintain bias and precision of the instruments within preset limits so that the effect of cycling was minimized.

The overall results reported here indicate that a high degree of accuracy, precision, and interlaboratory comparability were achieved through the rigorous standardization and control of the entire analytical procedure in the LRC laboratories. Interlaboratory and intralaboratory variability were probably as good as can be achieved with the AAII system. In addition, the average bias was very small and probably approaches the practical limits of accuracy that can be achieved in any collaborative multilaboratory study.

More generally, the results show that collaborative studies involving laboratories across the country can be done in a such a way that resulting data can be legitimately pooled for analysis. Care must be taken in standardizing methodology, standards, control pools, calibration material, etc. Internal and external quality-control programs should also be implemented, so that one can monitor and assess the performance and comparability of the various laboratories.

The following personnel are acknowledged:

Directors, Lipid Research Clinics: William Connor, M.D., Francois Abboud, M.D., Reagan Bradford, M.D., Daniel Steinberg, M.D., W. Virgil Brown, M.D., John W. Farquhar, M.D., Ivan Frantz, M.D., Charles Glueck, M.D., Antonio Gotto, M.D., Edward Bierman, M.D., William Hazzard, M.D., Peter O. Kwiterovich, M.D., John LaRosa, M.D., Alick Little, M.D., Gustav Schonfeld, M.D., Elena H. Gerasimova and Anatoli N. Klimov (USSR), and Yechezkiel Stein (Israel).

Director, Central Patient Registry and Coordinating Center: James Grizzle, Ph.D. and O. Dale Williams, Ph.D.

Director, Central Electrocardiographic Laboratory: L. Thomas Sheffield, M.D.

Program Office (National Heart, Lung, and Blood Institute): Conrad Blum, M.D., Gary Fisher, M.D., Barry Greenberg, M.D., Richard Havlik, M.D., Basil M. Rifkind, M.D. (Project Officer), and Barbara Roberts, M.D.

References

1. Tonks, D. B., The estimation of cholesterol in serum: A classification and critical review of methods. *Clin. Biochem.* **1,** 12 (1967).

2. *Manual of Laboratory Operations, Lipid Research Clinic Program,* 1, *Lipid and Lipoprotein Analysis.* DHEW Publication No. (NIH) 75-628.

3. Williams, J. H., Kuchmak, M., and Witter, R. F., Purity of cholesterol to be used as a primary standard. *J. Lipid Res.* **6,** 461 (1965).

4. Brown, W. D., Errors in determination of serum cholesterol. *Aust. J. Exp. Biol. Med. Sci.* **39,** 209 (1961).

5. Abell, L. L., Levy, B. B., Brodie, B. B., and Kendall, F. E., Simplified method for the estimation of total cholesterol in serum and demonstration of its specificity. *J. Biol. Chem.* **195,** 357 (1952).

6. Williams, J. H., Taylor, L., Kuchmak, Mr., and Witter, R. F., Preparation of hypercholesterolemic and/or hypertriglyceridemic sera for lipid determinations. *Clin. Chim. Acta* **28,** 247 (1970).

7. Anderson, R. L., and Bancroft, T. A., *Statistical Theory in Research,* McGraw-Hill, New York, N.Y., 1952.

8. Fasce, C. F., Jr., and R. E. Vanderlinde, Factors affecting the results of serum cholesterol determinations: An interlaboratory evaluation. *Clin. Chem.* **18,** 901 (1972).

9. Laessig, R. H., Schwartz, T. H., and Paskey, T. A., Cooperative survey of "SMA" analyzers in Wisconsin—a method of improving interlaboratory performance. *Clin. Chem.* **19,** 62 (1973).