Printed December 1982 DOT HS-806 193
Interim Report

U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

Statistical Evaluation of Blood Alcohol Measurements

James A. Lechner

National Bureau of Standards
U.S. Department of Commerce
Washington, D.C. 20234

Contract No. DOT-HS-020-2-2901A

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT-HS-806 193		

4. Title and Subtitle	5. Report Date
Statistical Evaluation of Blood Alcohol Measurements	October 1981
	6. Performing Organization Code
	8. Performing Organization Report No.
7. Author(s) James A. Lechner	
9. Performing Organization Name and Address National Bureau of Standards U.S. Department of Commerce Washington, DC 20234	10. Work Unit No. (TRAIS)
	11. Contract or Grant No. DOT-HS-020-2-290IA
	13. Type of Report and Period Covered Interim 1977 - October 1980
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration U.S. Department of Transportation Washington, DC 20590	14. Sponsoring Agency Code

15. Supplementary Notes This report was prepared by the Law Enforcement Standards Laboratory of the National Bureau of Standards' National Engineering Laboratory under the direction of Lawrence K. Eliason, Chief of LESL, and Paul H. Krupenie.

16. Abstract

The U.S. Department of Transportation, National Highway Traffic Safety Administration (NHTSA) has instituted a voluntary program to evaluate the proficiency of laboratories measuring the amount of alcohol in blood. In this report, data from that program are examined, and the variability of those measurements assessed. Differences between labs, between dates for the same lab, and between samples on the same date are quantified. Differences in overall bias for the six different periods (covering two years) are noted. A few of the roughly 120 labs participating were observed to perform considerably less well than the others.

17. Key Words	18. Distribution Statement
accuracy; blood alcohol; clinical laboratory; reliability; statistical analysis; validity	Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	28	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

METRIC CONVERSION FACTORS

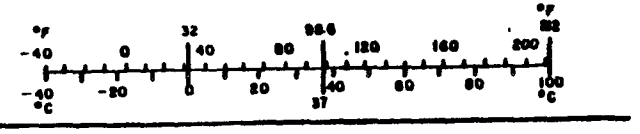
Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.

Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



ACKNOWLEDGMENTS

This report was prepared by the Law Enforcement Standards Laboratory of the National Bureau of Standards' National Engineering Laboratory under the direction of Mr. Paul H. Krupenie and Lawrence K. Eliason, Chief of LESL: Dr. James F. Frank, National Highway Traffic Safety Administration, Office of Driver and Pedestrian Research, Contract Technical Manager. The Data that are analyzed in this report were provided by Dr. Arthur L. Flores, Transportation Systems Center, Cambridge, MA.

EXECUTIVE SUMMARY

The effort reported herein was undertaken at the request of the National Highway Traffic Safety Administration (NHTSA), as part of an effort to evaluate how accurately U.S. clinical labs can determine the alcohol level in whole blood.

Data used in this effort came from the Voluntary Blood-Alcohol Proficiency Test Program conducted by the Transportation System Center (TSC), Department of Transportation, for the NHTSA. Each test involves the analysis of four unknowns. Summaries of the results, and some detailed data, were made available to the National Bureau of Standards. From these data, systematic errors and variability were assessed. The results of that assessment, together with observations on possible areas for improvement, are presented.

The questions asked of these data and the answers were:

(1) Is there an overall bias between lab results and target values? (Yes) How large is it? (Down to -1% in March 1977.) (Note that we do not have true values--only target values that are thought to be close to the true values.)

(2) Do the average results for different dates show a pattern? (Yes; negative bias, decreasing in magnitude.)

(3) Do analytical techniques differ significantly from each other? (No.)

(4) Do labs differ, beyond the difference (if any) due to analytical techniques? (Yes, in offset and variability; offsets have a standard deviation of 3.8%.)

(5) How big is the variability due to measurement error (generally called repeatability)? How does it depend on concentration? (Four percent for a single determination, but not less than 0.0033 g/100 mL.)

(6) Does the act of sample preparation, or do the differences between "identical" samples sent to different labs, seem to contribute variability? How much? (These two sources cannot be differentiated in this study, since presumably there is only one act of "sample preparation" for each sample.) (Yes; 4.9%, standard deviation.)

(7) Are there any other obvious patterns or sources of variation? For example, are the errors on the four samples measured by one lab on one date correlated? (This could be due to the presence of a "system setup" error, differing from date to date, but affecting all measurements made by the lab on that date.) (Yes; "system setup" seems to introduce a random offset, with standard deviation of 5.1%.)

This last component is the largest one listed. As explained in section 4, the data studied provide no way to tell how much this effect differs from day to day or week to week, because the data are obtained once every four months. It could be a long-term drift.

The setup, sample preparation, and between-lab variances, combined by root-mean-square, would produce an overall standard deviation of 8 percent (as compared with 8.2 percent standard deviation observed). Thus there is little point in trying to improve repeatability, or in taking more

determinations, or worrying about date or technique differences, until these three sources are reduced considerably.

Miscellaneous other considerations:

(1) Measurements ought to be taken to three decimal places; fewer places increases the effects of roundoff.

(2) The overall bias is negative, but decreasing in magnitude -- it was -1 percent at the last (March 1977) date. The implications for the two kinds of error--reading a sample too high, or reading it too low--might well be considered in the light of the use of blood alcohol measurements in legal proceedings.

(3) A few labs did significantly poorer than the general run of labs, having (jointly) more than one-third of their measurements off by more than 20 percent. A useful role might be envisioned for an "expert" lab to provide individual help to labs that are in need of upgrading.

(4) Measurements deviating from target by more than 25 percent were adjusted to $\pm$25 percent for these analyses, in order to obtain more stable estimates of the different sources of variation. In other words, there are occasional wild observations in addition to the variability already described. These concentrate in--but are not limited to--a small percentage of labs.

CONTENTS

STATISTICAL EVALUATION OF BLOOD ALCOHOL MEASUREMENTS

James A. Lechner*
National Bureau of Standards
Washington, DC 20234

The U.S. Department of Transportation, National Highway Traffic Safety Administration (NHTSA) has instituted a voluntary program to evaluate the proficiency of laboratories measuring the amount of alcohol in blood. In this report, data from that program are examined, and the variability of those measurements assessed. Differences between labs, between dates for the same lab, and between samples on the same date are quantified. Differences in overall bias for the six different periods (covering two years) are noted. A few of the roughly 120 labs participating were observed to perform considerably less well than the others.

Key words: Accuracy; blood alcohol; clinical laboratory; reliability; statistical analysis; validity.

1. INTRODUCTION

The effort reported herein was undertaken at the request of the National Highway Traffic Safety Administration (NHTSA), as part of an effort to evaluate how accurately U.S. clinical labs can determine the alcohol level in whole blood. This is an important question, since overindulgence in alcohol contributes greatly to highway accidents, and determination of intoxication is frequently based on measured blood alcohol concentration.

Data used in this effort came from the Voluntary Blood-Alcohol Proficiency Test Program conducted by the Transportation Systems Center (TSC), Department of Transportation, for the NHTSA. Originally intended as a short-term survey of the status of proficiency in blood alcohol analysis, the program is now conducted as a continuing service to approximately 160 participating laboratories. During the early years of the program tests were performed about every four months. Currently, tests are performed twice a year, available funding permitting. Each test involves the analysis of four unknowns. Summaries of the results, and some detailed data, were made available to the National Bureau of Standards. From these data, systematic errors and variability were assessed. The results of that assessment, together with observations on possible areas for improvement, are presented.

The questions asked of these data were:

(1) Is there an overall bias between lab results and target values? How large is it? (Note that we do not have true values--only target values that are thought to be close to the true values.)

(2) Do the average results for different dates show a pattern?

(3) Do analytical techniques differ significantly from each other?

(4) Do labs differ, beyond the difference (if any) due to analytical techniques?

---

*Statistical Engineering Division, Center for Applied Mathematics, National Engineering Laboratory.

(5) How big is the variability due to measurement error (generally called repeatability)? How does it depend on concentration?

(6) Does the act of sample preparation, or do the differences between "identical" samples sent to different labs, seem to contribute variability? How much? (These two sources cannot be differentiated in this study, since presumably there is only one act of "sample preparation" for each sample.)

(7) Are there any other obvious patterns or sources of variation? For example, are the errors on the four samples measured by one lab on one date correlated? (This could be due to the presence of a "system setup" error, differing from date to date, but affecting all measurements made by the lab on that date.)

The next section describes the data used for this study. Section 3 presents the study of variability within laboratories, based primarily on detailed analyses of approximately 780 measurements from 77 labs for one set of four samples (May 1976); section 4 describes the between-labs study of summary data for six sets of four samples, from May 1975 through March 1977. Finally, section 5 contains the summary and conclusions, including possible directions for improvement.

2. DATA USED IN THIS STUDY

Two kinds of data were used for this study: detailed data for May 1976, and summary data for six dates, May 1975 through March 1977. These data are described below in the course of describing the procedure used by the TSC to conduct the exercises.

Roughly every four months, TSC sends out to each partipating laboratory four samples of blood, spiked with known amounts of alcohol, and with various preservatives added. Each participating laboratory is instructed to analyze the samples as they would any ordinary sample, and send the results (including repeat measurements if any) to TSC. A summary[1] of the results for that date, identifying different labs by number only, is then distributed to each lab by TSC. Each lab can then compare its results with the target values and the results of other labs, but without being able to identify the different labs by name. All concentrations, both target and measured, are stated in grams of alcohol per 100 mL of blood.

Summary reports were made available to NBS for 11 dates. The number of "labs" reporting was 52 for the first of these, rose to a high of 103, and leveled off at 101. The last six of these summaries were analyzed, it being felt that a series of six dates would be sufficient to assess the current state of the art. The number of "labs" for these six dates was 79, 92, 101, 103, 101, and 101. (The word "labs" is in quotations because, in those few cases where one lab analyzed the samples by more than one technique, that lab appears in the summary separately for each technique used. Thus the actual number of different labs participating is somewhat less, ranging from 74 to 101 for the six periods of interest.)

In the process of cleaning up the data, three of the 2308 values were estimated by setting them equal to the mean value obtained for that sample by the other labs which used that technique. The 21 techniques used were grouped into five general techniques, following the TSC: gas chromatography by headspace, by whole blood injection, or by other injection; dichromate oxidation; and enzymatic oxidation. These are labeled in this report as techniques 1 through 5, respectively.

[1] Illustrative portions of one such summary are presented in appendix A.

Since there is no direct evidence upon which to estimate the variability of repeat measurements (sometimes called "repeatability") from the summary reports, all the useful data for one date, May 1976, were obtained. Labs submitting only one measurement for a given sample were ignored for that sample; depending upon the sample, this left 75 to 77 labs, with 194 to 199 measurements per sample. The next section describes the analyses performed on these detailed data, and the following section describes the analyses done on the summary data for six different dates.

3. WITHIN-LABS ANALYSIS

The first thrust of this investigation was addressed to the internal variability of measurements; how variable are repeated measurements from the same laboratory? Does this variability depend on the sample concentration? (Question 5 of sec. 1.)

3.1 Within-Lab Variability

The detailed data from May 1976 were subjected to an analysis of variance, for each sample separately. The results are shown in table 1, with standard deviation estimates given both in absolute units and in percent of the target value (i.e., relative s.d.).

Table 1. Repeated-measurements results: May 1976 data.

Statistic	Sample Number 1	2	3	4	Combined
Number of labs	77	75	76	76	
Number of measurements	199	194	199	197	
Target value, g/100 mL	.115	.380	.093	.060	
Sample average, g/100 mL	.1126	.3806	.0921	.0596	
Estimates of standard deviations:					
Within-lab[a], g/100 mL	.0034	.0078	.0037	.0033	
Between-labs[b], g/100 mL	.0079	.0284	.0071	.0044	
Relative standard deviation estimates (in percent):					
Within-lab[a]	3.0	2.1	4.0	5.5	3.9
Between-labs[b]	6.9	7.5	7.6	7.3	7.3

[a]"Within-lab" refers to the variability of repeated measurements, performed at the same time, on one sample. It is often termed "repeatability."

[b]This variation has three sources: technique differences, overall differences between labs, and (the largest) "setup" variations from one lab or date to another. (This latter source actually may include other causes, but the term "setup" is chosen because that seems likely to be the predominant cause.) See section 4.

The standard deviations (s.d.'s) seem consistent with the frequently-occurring situation wherein the standard deviation of a measurement is a given percentage of the measurement value, but with a "floor" that keeps the s.d. above a certain minimum value regardless of how small the concentration is. Thus, in the four samples analyzed here, with concentrations of about .06, .09, .12, and .38 g/100 mL, the s.d.'s are essentially identical (about .0035) for the three smaller concentrations, and about twice as large for the largest concentration. Thus the relative s.d.'s steadily decrease from 5.5 percent to 2.1 percent as the concentration increases. Since only four different concentrations are represented here, this cannot be considered a definitive characterization of the repeatability. For the concentration range 0.09 to 0.38, the relative s.d. seen in repeated measurement ranges from 4 percent to 2 percent. Since laboratories generally do repeat their measurements, the contribution of these measurement errors to the total error is decreased by the averaging process. The total error is discussed in section 5, using the combined rsd (3.9 percent), which is the root-mean-square (RMS) average of the four values for this date.

The analysis of variance also produces information about the between-labs variability. These four samples exhibited between-lab relative s.d.'s from 6.9 to 7.6 percent. these numbers being unrelated to the magnitude of the concentration being estimated. These figures estimate the variability between labs on any given sample (which may or may not be independent from sample to sample). This between-lab variability is sufficiently larger than the within-lab variability of 2-4 percent given above, so that the existence of between-lab differences is verified at the .0005 probability level. Note that the existence of differences has been demonstrated at this high level of significance, because there is a large amount of data. The estimated magnitude of the between-labs component of variability is, of course, less than 7 percent; it will be shown later that there are several contributing sources of this variation. The principal contributor to these differences will be called the "setup effect," since (as it turns out) measurements made by one lab on different dates vary much like measurements made by different labs on the same (or different) dates.

Note that this source of error cannot be reduced by repeating measurements at one lab at one time.

3.2 Round-Off Error in Reported Values

The question of round-off error needs to be addressed: for two of the labs, the repeat measurements showed no variation at all, presumably because these labs reported one less significant digit than the others. (Three other labs reported one less digit and still did not produce identical measurements.) There is always some effect from the finite representation of a number, but when several observations come out identical, the effect becomes obvious--even if it should be unimportant. How significant is this? Eisenhart, in Eisenhart et al. [1][2], has studied this question. His results showed that when the standard deviation of the measurements is at least three times the reporting interval, the estimate of the variance is likely to be high by at most 1 percent--so that when estimating a standard deviation of .003, the estimate will (on the average) be .00302 at most, an inconsequential difference. Our standard deviations are all above .003, so that reporting to .001 is no problem.

---

[2]Numbers in brackets refer to references in section 6.

There are nearly 500 degrees of freedom for estimating the within-lab variances. Only 32 of these come from measurements made to 0.01. Their effect is therefore negligible. In fact, the average within-lab s.d. for these few labs was even (slightly) higher than the average for all labs. It is important, however, to report the measurements to three decimals, because the measurement s.d. is considerably less than .01. The saving feature in this study is that most labs did indeed report to three decimals.

3.3 Correlated Errors for Different Samples

One might suspect that a lab which gets too high a value for one sample would tend to get high values for the others too. (Question 7 in sec. 1.) A qualitative indication of this condition is seen from the listing in table 2. This list was obtained by considering each sample separately, and listing the lab numbers in the order of their average determination for that sample--i.e., the number of the lab obtaining the highest average value for sample 1 is listed first in the column headed sample 1, the lab obtaining the next-highest value is listed next, etc. Now if there were no tendency for a lab to err in the same direction on all samples, the ordering of the four columns would be independent. As it turns out, the lab numbers near the top of one column tend to be near the top on all columns, and similarly for the bottom.

A more quantitative measure of the presence of this effect was obtained by counting how many labs measured all four samples higher than the median, or all four samples lower than the median. This was done for each date. This is a crude measure, but the effect is so pronounced that it is good enough. The last date (March 1977) is typical: 53 of the 101 labs had all four values on the same side of the median. Since the probability that any one value will be above the median is 1/2, the probability that all four values will be above the median is 1/16, if they behave independently. Similarly, the probability that all four values will be below the median is also 1/16, so the probability that all four values will deviate in the same direction is 1/8. Thus if the four sample errors are independent, the number of labs having all four deviations in the same direction is a binomial random variable, with mean equal to (101)(1/8) or 12.6, and standard deviation equal to

$$\sqrt{(101)(1/8)(7/8)} = 3.32.$$

Thus the actual number of "one-sided" labs, 51.5 (allowing for ties), differs from the expected number, 12.6, by about 12 standard deviations! Thus we have convincing proof of a persistence of lab bias across samples, or in different terms, evidence of correlations among the results for four samples all measured on the same date by one lab.

One might ask whether there is a long-term tendency for a given lab to read high (or low). This requires looking at results over several time periods, which will be done in the next section; it turns out that there are indeed significant differences among lab means, in addition to an overall tendency to read low.

4. STUDY OF SUMMARY DATA FOR SIX TIME PERIODS

The summary data for each laboratory consisted of average values for each of four samples, for six dates, with the average number of "labs" being 96. The target values and the observed between-lab relative s.d.'s (in percent) are shown in table 3, and the relative s.d's, hereafter denoted rsd's, are plotted against target value in figure 1. Note that these are the observed rsd's, and therefore include a contribution from the internal (within-lab) variability as well.

Table 2. Lab numbers listed in order according to the (average) value obtained on each sample. May 1976 data; labs submitting only one value per sample are not included.

	Ordered lab values[a] (scaled)				Corresponding lab identification numbers			
Sample	1	2	3	4	1	2	3	4
Rank								
1	8.36	11.35	7.43	6.00	104	7	104	97
2	5.29	8.28	4.56	3.63	30	98	97	104
3	4.32	7.76	4.20	2.46	97	52	7	7
4	3.83	7.27	3.61	2.16	52	2	30	17
5	3.68	6.84	3.02	2.11	11	100	2	30
6	3.44	6.58	2.79	1.81	7	86	101	101
7	3.10	6.47	2.79	1.76	101	101	98	9
8	2.76	3.70	2.65	1.65	9	47	100	100
9	2.51	3.15	2.24	1.50	115	50	12	68
10	2.17	3.00	2.11	1.35	86	15	117	2
11	2.07	2.97	1.74	1.25	55	62	86	106
12	2.07	2.59	1.56	1.20	17	109	68	12
13	2.07	2.42	1.56	1.20	15	84	11	11
14	2.07	2.29	1.47	.95	12	99	109	55
15	2.07	2.29	1.47	.95	2	11	1	49
16	1.78	2.29	1.29	.90	39	1	51	108
17	1.68	2.20	1.20	.90	100	106	17	15
18	1.59	1.84	1.11	.85	49	87	39	34
19	1.49	1.77	1.02	.85	98	39	50	1
20	1.49	1.77	.93	.74	91	35	103	29
21	1.49	1.64	.88	.74	1	103	15	14
22	1.30	1.51	.74	.64	106	48	106	39
23	1.20	1.39	.74	.64	84	80	44	4
24	1.20	1.39	.74	.59	44	3	9	98
25	1.05	1.04	.74	.59	85	44	4	95
26	1.05	1.00	.65	.44	35	9	14	118
27	.91	.96	.47	.44	103	94	91	109
28	.91	.83	.47	.44	99	97	55	103
29	.71	.78	.47	.44	50	54	29	94
30	.61	.78	.38	.34	109	40	49	20
31	.52	.68	.33	.29	88	23	87	91
32	.32	.61	.33	.29	94	108	41	51
33	.32	.53	.29	.24	87	4	110	89
34	.32	.14	.29	.14	68	20	96	117
35	.32	-.12	.29	.14	29	19	89	84
36	.22	-.16	.20	.14	110	12	20	50
37	-.26	-.29	.06	.04	108	29	81	114
38	-.26	-.35	.02	.04	89	85	60	80
39	-.26	-.38	.01	-.01	60	14	34	61
40	-.26	-.46	-.08	-.07	54	88	58	88
41	-.26	-.61	-.08	-.07	14	95	3	83
42	-.26	-.72	-.38	-.07	3	93	99	44
43	-.36	-.89	-.35	-.17	40	17	95	93
44	-.36	-.93	-.35	-.17	20	81	85	87
45	-.41	-.93	-.35	-.17	111	42	52	60
46	-.70	-.93	-.35	-.17	51	6	6	52
47	-.70	-1.00	-.44	-.27	23	30	54	110
48	-.85	-1.06	-.62	-.27	117	49	35	58
49	-.85	-1.19	-.71	-.32	62	91	88	3
50	-.85	-1.19	-.71	-.37	61	38	80	86
51	-.85	-1.32	-.80	-.57	47	117	93	6
52	-.85	-1.32	-.90	-.62	46	112	48	48
53	-.85	-1.36	-.90	-.67	42	60	40	45
54	-1.14	-1.36	-.99	-.77	48	58	72	99
55	-1.43	-1.45	-1.08	-.77	93	118	77	85
56	-1.43	-1.45	-1.17	-.77	19	111	118	38
57	-1.53	-1.62	-1.35	-.87	83	77	47	62
58	-1.53	-1.88	-1.35	-.92	58	110	19	42
59	-1.58	-2.01	-1.44	-.97	95	83	113	81
60	-1.63	-2.26	-1.44	-.97	6	114	23	77
61	-1.73	-2.41	-1.53	-1.18	77	115	62	54
62	-1.73	-2.41	-1.58	-1.23	72	61	111	113
63	-1.73	-2.74	-1.58	-1.28	38	89	42	33
64	-2.02	-2.80	-1.62	-1.53	33	113	81	35
65	-2.17	-3.32	-1.71	-1.58	27	21	38	19
66	-2.21	-3.70	-1.99	-1.68	81	72	112	47
67	-2.31	-3.90	-1.99	-1.68	118	33	21	40
68	-2.31	-4.54	-2.17	-1.78	4	5	83	72
69	-2.46	-4.63	-2.17	-1.83	113	63	33	112
70	-2.51	-4.63	-2.44	-1.83	80	34	114	23
71	-2.60	-4.67	-2.53	-1.88	34	55	27	63
72	-2.76	-6.60	-2.62	-1.99	21	27	63	111
73	-3.04	-7.31	-2.81	-2.19	112	68	108	5
74	-4.36	-7.82	-3.35	-2.74	114	104	5	27
75	-4.55	-11.37	-4.90	-2.90	5	45	45	46
76	-4.94		-6.08	-3.50	63		46	21
77	-7.38				45			

[a]For each sample (column) in turn, each lab value was scaled by subtracting the average of all the lab values and dividing by the estimated within-lab standard deviation. The ordering of the numbers is unaffected by this scaling.

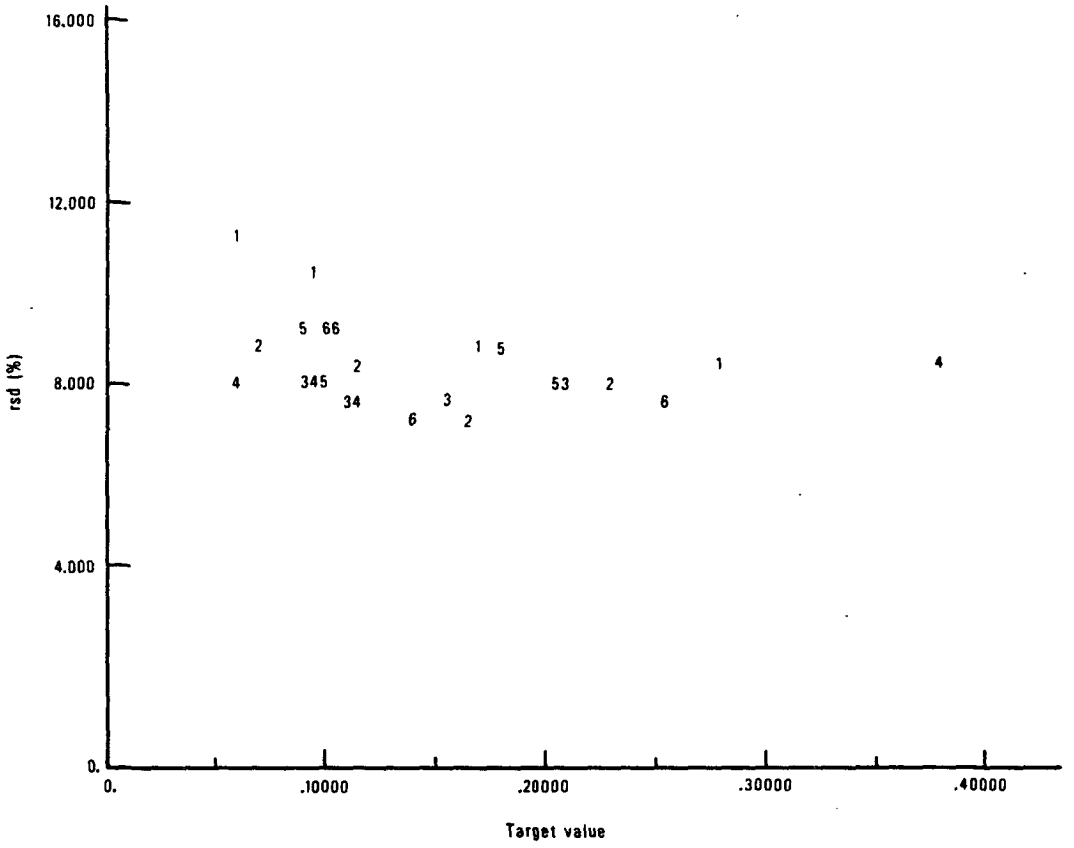


Figure 1. Relative standard deviation for each sample, (plotted against target value (plotting symbol is date).

4.1 Data Expressed as Relative Differences

From figure 1, it is obvious that the target value has essentially no effect on the rsd, except for the low concentrations for the earliest date. (The set of all 24 rsd's is nonhomogeneous (at the 0.3 percent significance level, by the Bartlett-Box test), but with these two values deleted, the remaining 22 values show no evidence of nonhomogeneity.) For this reason, it was decided to convert each observation to a relative difference from the target value, by subtracting and dividing by the target value. That is, if the target value is .20, and one lab's determination is .230 (15 percent above the target value), this lab's value for this sample becomes (.23 - .20)/.20 or .15. All further analyses in this report are in terms of these relative differences. It would have been possible to do a weighted analysis, but the effects of such weighting were judged to be negligible since only 2 of 24 sets would be downweighted, and not very heavily at that.

In order to assess differences between techniques, between dates, and between labs, it helps to minimize the effects of random error and sample preparation variability. Thus, since the relative differences just described have equal variability for any concentration being measured, it is helpful to

Table 3. Target value and average measured value, with between-laboratory relative standard deviation (in parentheses, in percent), by date and sample.[a]

Date	Sample number 1	2	3	4
1st	.281,.274 (8.31)	.060,.0556 (11.20)	.171,.164 (8.63)	.093,.0910 (10.31)
2nd	.232,.229 (7.82)	.113,.114 (8.23)	.069,.0716 (8.73)	.163,.161 (7.39)
3rd	.093,.0901 (8.17)	.153,.147 (7.54)	.110,.101 (7.70)	.208,.204 (7.83)
4th	.115,.112 (7.63)	.380,.376 (8.30)	.093,.0922 (8.16)	.060,.0595 (8.05)
5th	.207,.204 (8.12)	.092,.0900 (9.04)	.181,.175 (8.69)	.093,.0921 (8.16)
6th	.254,.251 (7.71)	.099,.0986 (9.39)	.141,.138 (7.04)	.106,.105 (9.36)

[a]The upper pair of numbers are the target value and the average of all measured values. The numbers is parentheses are the between-laboratory relative standard deviations.

average the relative differences for the four samples on any one date, for each lab. In order to make clear which numbers are being analyzed, we adopt the following conventions: the term "lab value" will henceforth mean the relative difference obtained by one lab on one sample on a given date; the term "lab average" will mean the average of the four relative differences obtained by one lab on one date.

4.2 Treatment of "Outlying" Values

One more important detail: before analysis and before calculation of table 3, the lab results were adjusted to decrease the effect of gross errors [2]. (For tables 1 and 2, which are based on May 1976 data, the adjustment makes essentially no difference.) All measured values differing by more than +25 percent from the target values were adjusted to differ by ±25 percent. There were 56 such outliers out of the total of 2308 measurements (2.4 percent of the measurements), ranging from -100 percent to +82 percent of target values; the number of such measurements on any one date range from four in May 1976 to 15 (including both the extremes quoted above) in March 1977. The lab values derived from these discrepant measurements had (among themselves) a 60 percent standard deviation; their elimination from the total sample reduces the average standard deviation among lab values for a given sample from about 12 percent to 8 1/2 percent, so their contribution is not negligible. Interestingly, 3 of the 122 labs contributed 18 of the 56 outlying measurements, and 9 of these labs' remaining 46 measurements were more than 20 percent away from the target value. Another 2 labs had 10 of their total of 36 measurements more than 20 percent off target. More than one-third of the measurements reported by these five labs were in error by more than 20 percent. Perhaps if such labs would carefully scrutinize their procedures, considerable improvement could be realized.

One might question the value 25 percent used for this adjustment of outliers. Surely some adjustment is wise, and 25 percent corresponds to about 3 sigmas. A further check was made by redoing the analyses after adjusting the data to ±15 percent. No surprises appeared, and most of the standard deviation estimates were only a little smaller.

These "outliers" were adjusted in order to obtain better estimates of the other effects and their relative importance. However, caution is in order when interpreting variance estimates based on the adjusted data, which underestimate the variability to be found in real life. A few labs seem to have much larger errors routinely than the adjusted data would indicate, and even among the other labs, one would expect an occasional very large error; but adjusting eliminates these large errors.

4.3 A Model for the Measurement Process Under Consideration

Each measurement reported in this program can be considered to be a sum of "effects" due to assignable causes, and an "error." In terms of the lab values expressed as relative differences from the target value, we can write the following:

$$X_{jk\ell} = \mu + D_k + L_\ell + S_{k\ell} + P_{jk\ell} + e_{jk\ell} ,$$

where:

$X_{jk\ell}$ is the lab value for the j-th sample on the k-th date by the ℓ-th lab;

μ is the grand mean;

D_k is the effect of the k-th Date [which includes the average (over four samples) of any bias in the "target" values for that date];

L_ℓ is the effect of the ℓ-th Lab, relative to the mean effect of all labs;

$S_{k\ell}$ is the system Setup effect for the ℓ-th lab on the k-th date, and is common to all measurements made on that date by that lab;

$P_{jk\ell}$ is the sample Preparation effect for the j-th sample on the k-th date at the ℓ-th lab;

$e_{jk\ell}$ is the random Error for the j-th sample, k-th date, ℓ-th lab.

The remainder of this subsection is devoted to an explanation of this model, one term at a time.

Grand Mean: If there were no overall bias in the measurements, then there would be no need for the constant μ. It might turn out that the estimated value of this constant is so close to zero as to be insignificant. However, to allow for the possibility of such an overall bias, the constant is included.

Date Effect: It is conceivable that the bias (if there is one) varies from date to date. Perhaps the labs are slowly (on the average) removing their bias. It is also possible that the target values are in error, thus contributing a fixed error to every deviation for that date. This effect could (at least in part) be considered a random effect--i.e., it could be effectively drawn from a random distribution, independently for each date--but one would also be interested in other possibilities, as for example a time trend in the effect (as would be expected if the labs were indeed improving their performance). Thus this has been treated as a fixed effect: there are six dates, each with its value of the "date effect," and we

investigate whether these six values are different, and whether there is a pattern to the numbers.

Lab and Technique Effects: Each participating laboratory may have a characteristic long-term bias. This effect is allowed for by the L_ℓ terms.

The variation exhibited by the L_ℓ terms could be further subdivided, into one part due to technique differences and a second part due to lab differences within a given technique. In this case, one could define a technique effect, say T_t, representing the average effect of all labs using that technique, and then define the lab effect relative to this average. There are five general classes of techniques used by the participating laboratories. However, this is not a designed experiment. Any one lab used only one technique on any one date (with a few exceptions); some labs changed techniques during the experimental period, but not on any systematic basis; and presumably there are many other labs using each technique which did not participate in the study. Thus it is not easy to quantify any differences due to technique. One thing that can be examined, however, is whether the results obtained by the particular set of labs using a given technique differ from the results for those labs using a different technique. If such differences are observed, and are significant in relation to the differences between labs using the same technique, then it can be considered established that there are differences between techniques. On the other hand, if no such significant differences are observed, it appears that any technique differences are small relative to differences between labs. In this case, the complication of subdividing the lab variation can be avoided.

Since the remaining effects are "nested," they will be explained in the order of increasing aggregation.

Random Error: The random error, $e_{jk\ell}$, refers to the contribution from nonassignable causes--i.e., the differences between repeat determinations at the same time on the same sample by the same lab. Of course, this cannot be evaluated without repeat determinations, so that it cannot be evaluated from just the lab averages or even from the lab values (unless it can be assumed that different samples behave alike, which is an unwarranted assumption). But we do have detailed data for one date, consisting of repeat measurements on individual samples. From this, the relative standard deviation (rsd) of repeats on individual samples can be estimated; then, knowing how many repeats were done on individual samples, we can estimate the rsd to be attached to the error in a lab value; and finally, knowing that four lab values go into any lab average, we can estimate the rsd due to the random error in a lab average.

Sample Preparation: Now suppose one were to compare lab values within a lab and date. If they vary more than the calculations of the previous paragraph would say they ought to, to what can we attribute the extra variation? In this study, this variation has been modeled as a random effect, independently drawn for each sample, lab, and date combination, and called the Sample Preparation Effect. The rationale is that it is an effect that persists over all measurements on that sample, but nowhere else, as if it came about from the act of sample preparation.

System Setup: One could compare lab averages for a given lab, for different dates. These might vary more than the considerations of the previous two paragraphs and the overall date effect would indicate. What kind of an effect would this be, varying from date to date within a lab, but affecting all measurements made that date? In this study, it is called the System Setup Effect, and is modeled as a random effect. The rationale is that it affects all measurements made that date at that lab--i.e., all measurements made with the system as set up for that date. Since these "dates" are four months apart, we may be seeing not a day-to-day variation but instead a long-term gradual drift. In order to determine how much of the

variation is actually a long-term drift and how much a day-to-day random variation, one would need measurements made on successive days (or perhaps weeks). Such measurements may have been made by some labs. If not, they could be made. The establishment of a Measurement Assurance Program [3] would involve this kind of evaluation, of course. Note that if the overall date effect is large, it will be difficult to measure this effect. One way out, if necessary, is to adjust the data for the date effect before evaluating this effect.

4.4 Analyses of Summary Data

Several analyses of variance were performed on the summary data expressed as relative differences from the target values. These are summarized below.

A. Two-way analysis of variance by technique and date.

This analysis of the lab averages, for which the lab identification was ignored, resulted in the following analysis of variance (ANOVA) table [4]:

Source	Degrees of freedom	Sum of squares	Mean squares	F ratio	(Significance)
Mean	1	.2308	.2308	47.1	(<.001)
Date	5	.06315	.01263	2.56	(.025)
Technique	4	.02488	.00622	1.26	ns
D x T	20	.09745	.00487	1.0	ns
Error	547	2.69621	.00493		

There is a statistically significant date effect, but no evidence of technique effect or of interaction between date and technique. An estimate of the rsd for the date effect is 0.84 percent (i.e., if the date effect were randomly, independently drawn from a normal distribution, the rsd of that distribution is estimated at 0.84 percent).

The above table contains a line for the mean. This line is often omitted from ANOVA tables, since generally there is no doubt that the mean is nonzero. However, in the present case, the data being analyzed are (relative) differences from the target values; thus there is reason to hope that the values cluster around zero, corresponding to a lack of bias. Therefore the line for the mean was included. The high level of significance found indicates that there is indeed an overall bias in the data (question 1 of sec. 1). Another way of looking at this question is to notice that the number of "labs" obtaining an overall average less than 0 is 129 (out of 177). In the absence of an overall bias, this number ought to be binomially distributed with mean equal to 177/2 and s.d. equal to $(177/4)^{1/2}$. Actually, 129-(177/2)=40.5, over six times the s.d. of 6.65. The average relative deviation from the target value, for each date and technique, is shown in table 4 and figure 2.

Table 4. Average relative deviation from target, by date and technique.[a]

Date	Technique 1	2	3	4	5
1st	-.0036	.0137	-.0273	.0021	-.0523
2nd	.0100	-.0165	.0023	.0022	-.0302
3rd	-.0112	-.0359	-.0283	-.0175	-.0252
4th	.0205	-.0000	.0546	-.0225	.0158
5th	-.0038	.0113	.0208	-.0019	.0080
6th	.0163	.0099	.0434	-.0301	.0291

[a]The tabulated values are obtained by subtracting the grand mean, -0.0217, from each value. Note that the average number of labs using each of the techniques was 44, 21, 5, 16, and 11 for techniques 1 through 5, respectively. Thus the behavior of technique 3 is relatively less well determined than the others. (Techniques are identified in sec. 2.)

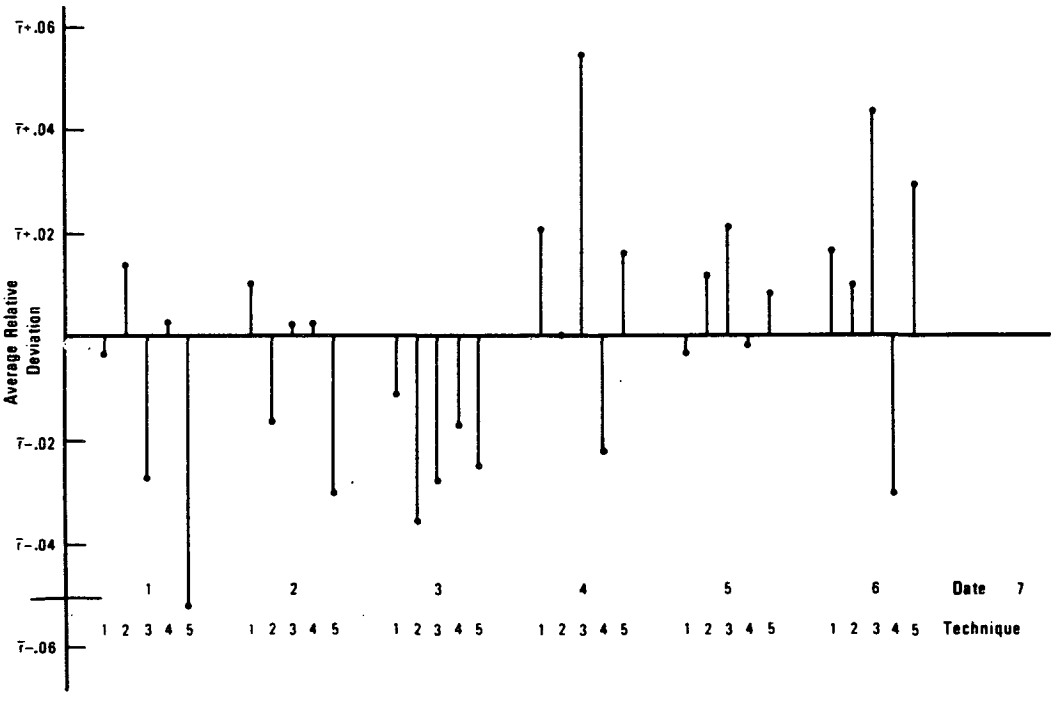


Figure 2. Graphical presentation of the data in table 4, giving the average relative deviation from target, by date and technique; $\bar{r}$(-.0217) is the overall average relative deviation from target of the entire set of data.

B. One-way by date.

Addressing question 2 of section 1, a one-way analysis of variance of lab averages by date was performed, ignoring all other variables. The variability with date, as estimated by this analysis, agrees with the value obtained in paragraph A. The overall average biases for the six dates are shown in figure 3, plotted against date. An upward trend is suggested; however, the fitted slope coefficient, based on only these six points, is not significantly different from zero. Perhaps the labs are learning to remove their built-in negative bias.

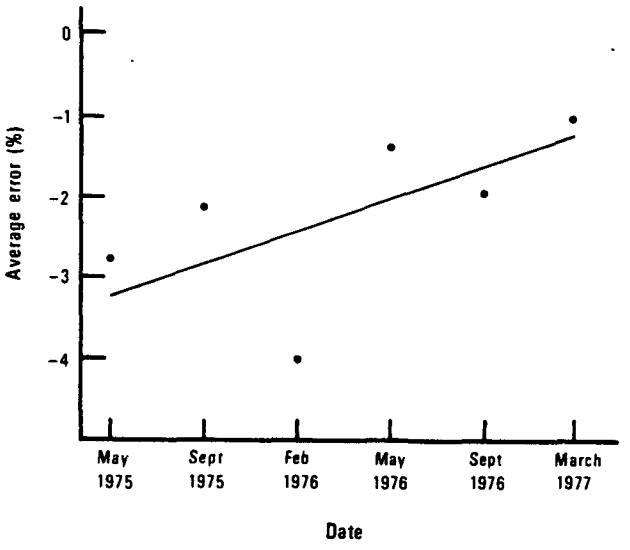


Figure 3. Average of all results (expressed as percent deviation from the target value) for a given date, plotted against date. Line fitted by least squares.

What have we shown about the "date effect"? Assume that the errors in the lab averages for a given date are drawn at random from a population whose mean can be termed the "date effect" for that date. We have examined the differences within dates (to estimate how variable those populations are); compared these with the differences between dates; and concluded that the differences between dates are too large to be consistent with the hypothesis that there is no date effect. (The significance level is correct only if the populations are normally distributed, with common variance, and the sample independently drawn. The last condition is violated if individual labs have biases which persist over time. In the presence of such persistent biases, the "within" sum of squares is too big, and the test for a date effect is consequently conservative--i.e., the effect is even more significant statistically than shown above.)

C. Analysis of variability by technique (question 3 in sec. 1)

A one-way analysis of lab averages by technique was performed, both with and without adjustment of the data for the date effect. The effect of technique is insignificant both statistically and practically. Since this finding agrees with the finding of paragraph A above, the technique effect was assumed nonexistent for the remaining analyses. This simplifies the laboratory analysis, since with the technique effects assumed nonexistent, L_ℓ is simply the effect of lab ℓ relative to the overall bias (after adjusting for date).

D. Analysis of within-lab and between-lab variability (question 4 in sec. 1)

A one-way analysis of the 577 lab averages by "lab" was performed, after adjusting for the date effect. The "lab" means (i.e., averages over all the dates for which that "lab" participated) range from -24.2 percent to +22.9 percent, significantly different at a probability level less than 0.005. This is a conservative test: the between-labs differences are evaluated relative to the between-date variation within labs, which may include time trends or other time effects within labs beyond the overall date effect. Thus the true significance level may be even smaller. If the laboratories were randomly drawn from a large population of laboratories, an estimate of the s.d. of the corresponding population of lab effects is 3.8 percent. As it is, this figure is simply a measure of the variation among the lab effects for these particular labs, since there may well be important differences between the labs which chose to participate in this analysis program and those which chose not to participate.

The within-labs rsd's -- i.e., the variability across dates for those labs submitting data on more than one date -- range from 0.1 percent to over 20 percent. This variation is significant (at the 0.05 percent level) -- i.e., there is essentially no doubt that the rsd does vary from one lab to another. This conclusion holds true also when the data are first adjusted for date. A histogram of the sample rsd's is presented in figure 4.

Finally, the analysis was done for true lab number (i.e., combining all results obtained by a given lab, even if several techniques were used). The same conclusions apply, except that the largest rsd was 17 percent.

The true rsd's probably do not span this range of values. These estimates are based on sets of two to six measurements, so the individual rsd estimates have uncertainties (rsd's) of 29 percent or more. (In fact, all seven rsd's less than 1.1 percent, and all four rsd's greater than 11.3 percent, are based on at most three measurements, so the uncertainties of these 11 estimates are 40 percent or more.)

Were we to assume a common within-lab rsd value, its estimate would be 5.9 percent. Since this estimate is based on 400 degrees of freedom, its s.d. would be about 0.2 percent, under the (unrealistic) assumption of a common rsd value. Note that this is an estimate of the variability from one date to another (sufficiently far away), of the average of the percent deviations of four samples after correction for the overall date effect; the variability for a single sample would be somewhat higher, although nowhere near twice as high.

This "within-labs" variability is not to be confused with the "repeatability" discussed in section 3. The variability under consideration here has three parts: the setup variability from time to time within one lab; the sample preparation variability; and an appropriate fraction of the repeatability. Thus the within-lab variance, $(5.9)^2$, is an estimate of the

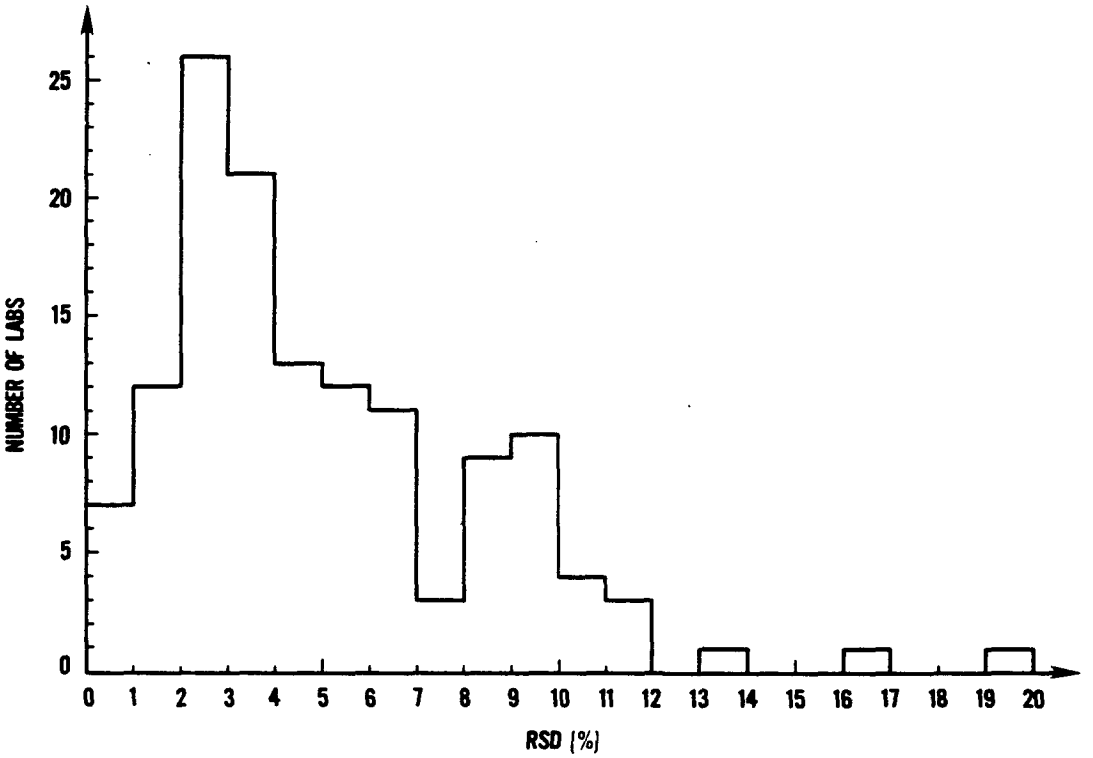


Figure 4. Histogram of percentage rsd's (over dates) of lab averages, for the 134 "labs" reporting data for more than one date.

setup variance plus one-fourth the sample preparation variance plus 0.137[3] times the repeatability variance, $(3.9)^2$, and one obtains an estimate of (setup variance plus one-fourth the sample preparation variance)$^{1/2}$ of 5.7 percent (rsd). Thus it appears that the overall, or average, or long-term differences between labs are not as large as the setup/sample preparation variability, but both are significant in a practical sense, being at least as large as the "repeatability" error for a single measurement within a lab (at least for concentrations of 0.09 or higher).

E. Analysis of lab values (not averages) in sets of four.

Up to this point, we have been discussing lab averages--each such number being the average of four lab values, one for each of the four samples distributed together on a given day. (Most of these lab values are averages, since most labs do more than one determination.) The four lab values

[3]The reciprocal of the properly-weighted average number of measurements entering into one analyzed lab average.

obtained by one lab at one time by one technique will not be identical, of course. However, the variation among these four values will not be due to variation between labs, dates, or technique, and (by definition) will not be affected by "system setup." Thus the variance "within" these sets of four is simply the sample preparation variance (σ_p^2) plus the proper fraction of the repeatability variance (.54 σ_{rpt}^2). As in paragraph D, we can subtract the latter term, to obtain an estimate of σ_p^2.

The "within" mean square for this analysis was 0.00322, and its expected value is σ_p^2 + .54 σ_{rpt}^2. Using the value $(.039)^2$ = .00152 for σ_{rpt}^2 from table 1, we obtain .0024 for our estimate of σ_p^2; i.e., the relative standard deviation of the sample preparation error is about 4.9 percent (question 6 of sec. 1).

Now consider the "between" mean square, from which system setup (question 7) can be addressed. Out of roughly 166,000 comparisons between two of the 577 lab averages, only about 330 are between averages from the same lab (on different days). Therefore, while the system setup effect is present in every comparison, the lab effect is absent in a fraction (330/166,000) of them. Thus the expected value of the "between" component is σ_S^2 + .998 σ_L^2. (The data were adjusted for the date effect before this analysis, and the technique effect is negligible.) The observed value was .004092. Subtracting the estimate of lab variance, .00144, from paragraph D, we obtain .00265 for the component due to system setup--i.e., the estimated relative standard deviation of the system setup effect is 5.1 percent. Incidentally, we obtained a value 5.7 percent for $(\sigma_S^2 + 1/4\ \sigma_p^2)^{1/2}$, in paragraph D. We obtain the same figure by combining the two estimates just obtained.

F. Analysis of variability of within-lab rsd's by number of dates

This analysis was performed, to see if more- or earlier-participating labs performed better or worse than less- or later-participating labs. The within-lab variances, i.e., the apparent variability of a lab over dates, were analyzed according to the number of dates for which the lab participated. No trend was observed, and no great difference between different groups. Apparently there is no significant difference between the labs participating in most of the tests and those participating in few.

5. SUMMARY AND CONCLUSIONS

In this report, several sources of variation of blood alcohol measurements were evaluated. Some should be considered as "fixed" effects, while others are best considered to be "random" effects. The lab effect is classed with the fixed effects, since it does not seem reasonable to assume that labs not participating in the study would be like those which did participate--i.e., the labs represented cannot justifiably be considered to be a random sample from any identifiable larger class of labs. (One could, of course, conceptualize a population of "similar" labs; then inferences to such a population would be legitimate.) A tabulation of these sources of variability, with estimated values, is presented as table 5.

Repeatability refers to variability between repeated measurements on one sample. Sample preparation refers to additional variability between samples on one date at one lab, over and above that which would be expected from the repeatability. System setup refers to still more variability between different dates and/or labs, beyond what would be expected from within-lab and date variability. This last component, the largest one listed, may come as a surprise. It is as if there is a component of error introduced by the act of setting up the system, randomly drawn each time the system is set up. As explained in section 4, the data studied provide no way to tell how much this effect differs from day to day or week to week, because the data are obtained only once every four months. It could be a long-term drift.

Table 5. Summary of sources of variability of blood alcohol measurements.

Source	Symbol	Relative Standard Deviation (percent)
FIXED EFFECTS:		
Between dates (the particular six dates analyzed, over 2 y), sec. 4.4(B)	σ_D	0.9[a] (the data were adjusted to account for this effect, in determining the values below)
Between techniques (five classes of technique), sec. 4.4(C)	σ_T	Insignificant
Between labs (i.e., between long-range averages for different labs), sec. 4.4(D)	σ_L	3.8[a]
RANDOM EFFECTS:		
System setup variation (between different "setups," whether in different labs or on different dates or both), sec. 4.4(E)	σ_S	5.1
Sample preparation (common to all measurements on a given sample by a given lab, but varying across samples for a given date and lab), sec. 4.4(E)	σ_P	4.9
Repeatability (the variability observed in repeated measurements of the same quantity), table 1	σ_{rpt}	3.9 (for a single determination) (σ_e is approximately $(.54)^{1/2}\sigma_{rpt}$ for these data)

[a]Even though these effects are considered fixed, not random, the calculated relative standard deviation still serves as a reasonable measure of the variation among the different values for the effect.

The setup, sample preparation, and between-lab variances, combined by root-mean-square, would produce an overall standard deviation of 8 percent (as compared with 8.2 percent standard deviation observed). Thus there is little point in trying to improve repeatability, or in taking more determinations, or worrying about date or technique differences, until these three sources are reduced considerably.

Miscellaneous other considerations:

(1) Measurements ought to be taken to three decimal places; fewer places increases the variability due to roundoff.

(2) The overall bias is negative, but decreasing in magnitude--it was -1 percent at the last (March 1977) date. The implications for the two kinds of error--reading a sample too high, or reading it too low--might well be considered in the light of the use of blood alcohol measurements in legal proceedings.

(3) A few labs did significantly poorer than the general run of labs, having (jointly) more than one-third of their measurements off by more than 20 percent. A useful role might be envisioned for an "expert" lab to provide individual help to labs that are in need of upgrading.

6. REFERENCES

[1] Eisenhart, Churchill; Hastay, Millard W. and Wallis, W. Allen, eds. Selected techniques of statistical analysis, New York: McGraw-Hill; 1949. p. 212.

[2] Huber, Peter J. Robust statistics. New York: Wiley; 1981. p. 18.

[3] Cameron, J. M. Measurement assurance, National Bureau of Standards (U.S.) NBSIR 77-1240; 1977 April.

[4] Brownlee, K. A. Statistical theory and methodology in science and engineering. 2d ed. New York: Wiley; 1965.

Appendix A--A Sample of the Data

The data summary for one of the dates is presented here, together with the full list of techniques. Shown in the body of the summary are lab number, date, average value for each of four samples, and technique used. Underneath these are the four values obtained by TSC and the technique used; the four target values; and the mean and standard deviation for each sample, calculated without regard to technique and then within technique groups.

Summary of Test Results - May 1976

Lab	Date	A	B	C	D	Tech
1	05/14/76	.118	.398	.098	.062	1
2	05/20/76	.120	.437	.103	.064	1
3	05/17/76	.112	.391	.092	.059	4
4	05/26/76	.105	.384	.095	.062	13
5	05/13/76	.097	.345	.080	.052	13
6	05/14/76	.107	.373	.091	.058	18
7	05/24/76	.125	.468	.108	.068	4
9	05/14/76	.122	.388	.095	.065	40
11	05/12/76	.126	.398	.098	.064	13
12	05/13/76	.117	.379	.101	.064	40
13	05/19/76	.111	.374	.092	.052	10
14	05/26/76	.112	.377	.095	.062	40
15	05/14/76	.120	.404	.096	.063	19
16	05/28/76	.107	.376	.093	.063	1
17	05/12/76	.120	.373	.097	.067	50
18	05/18/76	.108	.377	.088	.055	1
19	05/21/76	.108	.379	.087	.054	18
20	05/18/76	.112	.381	.093	.061	15
21	05/28/76	.104	.354	.085	.048	1
23	05/13/76	.111	.386	.087	.053	4
25	05/18/76	.106	.342	.085	.058	40
26	05/13/76	.113	.321	.102	.066	50
27	05/21/76	.106	.329	.083	.051	40
28	05/13/76	.104	.352	.077	.063	40
29	05/17/76	.114	.378	.094	.062	10
30	05/15/76	.131	.373	.106	.067	1
31	05/21/76	.110	.410	.090	.050	50
32	05/26/76	.118	.371	.099	.063	13
33	05/13/76	.106	.350	.084	.055	4
34	05/12/76	.104	.344	.092	.062	50
35	05/14/76	.117	.394	.089	.055	4
36	05/16/76	.075	.378	.090	.059	50
37	05/20/76	.100	.290	.080	.050	40
38	05/21/76	.107	.371	.086	.057	1
39	05/14/76	.119	.394	.096	.062	4
40	05/14/76	.112	.386	.089	.054	40
42	05/12/76	.110	.373	.087	.057	13
43	05/17/76	.119	.387	.096	.061	1
44	05/13/76	.117	.388	.095	.059	1
45	05/20/76	.088	.292	.074	.057	40
46	05/27/76	.110	.300	.070	.050	2
47	05/14/76	.110	.409	.087	.054	1
48	05/14/76	.109	.392	.089	.058	1
49	05/13/76	.118	.372	.094	.063	50
50	05/16/76	.116	.407	.096	.060	6
52	05/15/76	.126	.440	.091	.059	6
54	05/12/76	.112	.386	.091	.056	14
55	05/13/76	.120	.344	.094	.063	2
56	05/22/76	.133	.343	.100	.069	50
57	05/18/76	.121	.398	.099	.063	14
58	05/13/76	.108	.370	.092	.059	4
59	05/18/76	.115	.395	.095	.061	50
60	05/17/76	.112	.370	.092	.059	1
61	05/19/76	.110	.362	.094	.059	40
62	05/13/76	.110	.403	.087	.057	4

Summary of Test Results - May 1976 (Continued)

Lab	Date	A	B	C	D	Tech
63	05/14/76	.096	.344	.083	.053	13
64	05/19/76	.120	.463	.102	.060	13
65	05/13/76	.098	.325	.089	.062	13
66	05/14/76	.104	.342	.083	.057	13
67	05/15/76	.123	.400	.118	.065	6
68	05/17/76	.114	.323	.098	.065	30
70	05/17/76	.100	.347	.085	.056	1
73	05/17/76	.104	.364	.087	.054	4
74	05/17/76	.114	.392	.092	.054	40
75	05/14/76	.105	.359	.088	.061	6
77	05/18/76	.107	.368	.088	.056	13
79	05/17/76	.106	.361	.092	.061	4
80	05/17/76	.104	.391	.090	.060	32
81	05/19/76	.105	.373	.086	.056	40
83	05/18/76	.108	.365	.084	.059	50
84	05/13/76	.117	.399	.093	.060	1
85	05/18/76	.117	.378	.091	.057	4
86	05/12/76	.120	.431	.099	.058	4
87	05/13/76	.112	.370	.095	.061	40
87	05/13/76	.114	.395	.094	.059	13
87	05/13/76	.113	.389	.092	.060	4
88	05/19/76	.115	.377	.090	.059	13
89	05/14/76	.112	.359	.093	.060	1
91	05/20/76	.118	.371	.094	.061	1
92	05/14/76	.107	.390	.095	.060	40
93	05/19/76	.108	.375	.089	.059	14
95	05/14/76	.107	.376	.091	.062	13
97	05/12/76	.128	.387	.109	.079	4
98	05/19/76	.118	.444	.103	.062	1
99	05/20/76	.116	.398	.091	.057	1
100	05/19/76	.119	.433	.102	.065	1
101	05/17/76	.123	.431	.103	.066	1
102	05/14/76	.108	.367	.090	.060	40
103	05/13/76	.116	.393	.096	.061	13
104	05/20/76	.141	.319	.120	.072	31
105	05/24/76	.090	.320	.080	.050	1
106	05/14/76	.117	.397	.095	.064	10
107	05/21/76	.118	.337	.087	.052	1
108	05/21/76	.112	.385	.082	.063	40
109	05/17/76	.115	.400	.098	.061	1
110	05/20/76	.114	.366	.093	.059	4
111	05/14/76	.112	.369	.087	.053	40
112	05/18/76	.103	.370	.085	.054	13
113	05/17/76	.105	.359	.087	.056	40
114	05/20/76	.098	.363	.083	.060	17
115	05/13/76	.122	.361	.095	.064	40
117	05/12/76	.110	.370	.100	.060	1
118	05/14/76	.105	.369	.088	.061	17
TSC		.118	.409	.097	.062	2
Target		.115	.380	.093	.060	

All samples beef blood containing potassium oxalate 4 mg/mL, sodium fluoride 5 mg/mL.

Summary of Test Results - May 1976 (Continued)

AMEAN	STDA	BMEAN	STDB	CMEAN	STDC	DMEAN	STDD
.112	.0093	.376	.0317	.092	.0077	.059	.0049

TECH	AMEAN	ASTD	BMEAN	BSTD	CMEAN	CSTD	DMEAN	DSTD
1-9	.114	.0078	.385	.0333	.093	.0083	.059	.0053
10-29	.110	.0078	.378	.0253	.091	.0055	.059	.0036
30-39	.120	.0193	.345	.0402	.102	.0154	.065	.0059
40-49	.109	.0078	.363	.0310	.088	.0070	.058	.0046
50	.111	.0158	.367	.0275	.094	.0054	.062	.0056

TECH:

1 GC headspace
2 " " acetonitrile int. stand.
3 " " methanol " "
4 " " n-propanol " "
5 " " s-butanol " "
6 " " t-butanol " "
7 " " dioxane internal int. stand.

10 whole blood injection
11 " " acetonirile int. stand.
12 " " methanol " "
13 " " n-propanol " "
14 " " s-butanol " "
15 " " t-butanol " "
16 " " n-butanol " "
17 " " methyl ethyle ketone int. stand.
18 " " acetone

30 extract injection
31 supernatant injection, i-propanol int. stand.
32 distilate injection

40 dichromate oxidation

50 enzymatic oxidation