JOURNAL OF CLINICAL MICROBIOLOGY, Apr. 1977, p. 433-438
Vol. 5, No. 4

Printed in U.S.A.

Results of a Nationwide Proficiency Test for Carcinoembryonic Antigen

ROGER N. TAYLOR,* KAREN M. FULFORD, AND ALAN Y. HUONG

Center for Disease Control, Atlanta, Georgia 30333

Received for publication 7 December 1976

A proficiency testing survey for carcinoembryonic antigen (CEA) was conducted by the Center for Disease Control. The results from 125 laboratories throughout the United States indicate that many laboratories perform satisfactorily, but some laboratories need substantial improvement. Failure to follow manufacturer's recommendations and failure to heed the indications of "out of control" control results were two of the reasons for poor performance. Results reported for samples with CEA levels of ≥20 ng/ml showed that the direct method produced significantly higher values than the indirect method on either whole or diluted plasma. Almost one-fourth of the results reported in this survey were placed in the wrong nominal group. It was determined that the results were log normally distributed and, consequently, that statistical methods that are appropriate for this distribution should be used for the analysis of CEA results. Most of the variation observed was the result of poor comparability between laboratories rather than lack of precision within the laboratory. This indicates that better performance could be achieved by better standardization and closer adherence to established procedures.

Carcinoembryonic antigen (CEA) is a glycoprotein which is associated with embryonic entodermal epithelium and which can also be detected in extracts of carcinoma cells. Gold and Freedman (2, 3; P. Gold and S. O. Freedman, J. Clin. Invest., 44:1051-1052), by using heterologous antisera, first detected the antigen in colonic adenocarcinoma tissue. Clinical studies have indicated that detection and quantitation of CEA in plasma or serum may be of use in the diagnosis and management of some types of cancer (7).

The most commonly used procedure for CEA determinations is CEA-Roche, a commercially available method (6, 7). This method specifies perchloric acid (PCA) extraction of the antigen from plasma, reaction of the antigen with specific antibody, addition of labeled CEA (^{125}I-CEA), and precipitation of the CEA-antibody complexes by zirconyl phosphate (Z gel). Other methods either do not require PCA extraction or they prescribe other means of separating the free ^{125}I-CEA from the ^{125}I-CEA-antibody complexes.

Because CEA determinations are being performed in an increasing number of laboratories, the Proficiency Testing Branch, Center for Disease Control, conducted a special proficiency testing survey for CEA determinations in June 1976. Participation in the survey was voluntary, and performance was not graded. The purpose of the survey was to assess within- and between-laboratory performance of CEA assays.

MATERIALS AND METHODS

A pool of plasma containing high levels of CEA was obtained through the courtesy of Robert H. Engel, Leary Laboratories, Boston. Dilutions of this pool were prepared in pooled plasma obtained from healthy donors. A portion of the normal pool was used as sample XS6-004. Sample XS6-002 was a 1:10 dilution of the positive plasma, XS6-001 was a 1:150 dilution, and XS6-003 and XS6-005 were 1:300 dilutions.

The samples were sterilized by membrane filtration, and 2.5-ml portions were dispensed into 6-ml bottles. Samples were stored at −20°C until they were packed in Styrofoam boxes and shipped unrefrigerated.

The samples were shipped to 140 participating laboratories in which CEA testing was being done. These laboratories were enrolled in one or more proficiency testing programs offered by CDC. Results of the tests were due 3 weeks after the date of shipment, because most of the laboratories schedule CEA testing at least biweekly. The results returned were not graded, but participants were given the median value and the range of values reported for each sample and the differences between results reported for samples XS6-003 and XS6-005 and for samples XS6-001 and XS6-004.

The distributions of results were analyzed by the Kolmogorov-Smirnov (K-S) test of goodness of fit to determine whether the results were normal or log

normal (5, 8). The K-S test was chosen over the Chi-square test because the former is a more powerful test. Analysis of variance was used to estimate the components of within- and between-laboratory variance.

RESULTS

The results reported by the participant laboratories are summarized in Tables 1 and 2. The distribution did not appear to be normal (Gaussian), and, consequently, it would have been inappropriate to use Gaussian parameters (arithmetic mean and standard deviation) for analyzing the results. In the interest of time, results were reported to participants in terms of the median and range. Subsequent analysis revealed that the distribution was log normal, as is shown in Fig. 1 and Table 3, and the geometric statistical methods and notation previously described (9) were used in this analysis.

The calculated K-S values, the critical values at different levels of significance, and the probability of obtaining that particular K-S value, assuming that the observed curve fits the theoretical curve, are given in Table 3. The means, standard deviations, and 2-standard deviation limits are also given. The observed results fit the log normal distribution as well as or better than the normal distribution. In Fig. 1 are plotted the theoretical cumulative distribution, the cumulative distribution for the standardized observations, and the cumulative distribution for the standardized observations after log transformation.

The following observations were made for the 16 laboratories reporting values above 2.5 ng/ml on sample XS6-004. Two of the laboratories did not use the Roche procedure and did not indicate to us what they considered to be a "normal" value; three of the laboratories used only a single control, and values for two of these controls were above 5 ng/ml; seven of the laboratories used only two controls, and three of the low-level controls were above their expected limits; the remaining four laboratories used four controls, and the control results were elevated in two of these laboratories. Thus, it appears that at least half of the elevated levels reported on this sample should not have been reported, because improper controls were used or the control results indicated that a problem existed in the test.

Approximately 20% of the duplicate counts reported were not within 5% of the mean count. These results did not conform to the limit suggested by the manufacturer, and therefore the test should have been repeated (6). Of the laboratories reporting greater than 5% difference on duplicate counts, very few reported this much difference on more than one specimen. Variation in counts for the same sample was smaller than variation in counts between the duplicate samples. This could indicate that some of the variation within samples (duplicate counts on the same sample) was greater than the 5% limit and was eliminated by screening and repeat testing before the results were reported.

Table 2 shows the differences between results on duplicate and paired samples within laboratories. These data were included to indicate the degree of precision that was obtained with these tests. The results on duplicate samples

TABLE 2. *Within-laboratory differences between results on duplicate and paired samples*

Differences	Laboratories reporting (%)
Duplicate samples (XS6-003 and XS6-005)	
0–0.2	36.4
0–0.4	54.5
0–0.6	68.6
0–1.0	79.3
0–2.0	90.9
0–9.5	100.0
Paired samples (XS6-001 and XS6-004)	
4.2± 0.5	38.6
4.2 ± 1.0	62.3
4.2 ± 2.0	87.7
<2.2	8.8
>6.2	3.5

TABLE 1. *Summary of results reported on CEA proficiency testing samples*[a]

Specimen no.	Median	Range	$\bar{x}_G$[b]	SD_G[c]	2 SD_G range
XS6-001	5.5	1.5–10.1	5.5	1.35	3.0–10.0
XS6-002	105.0	12.4–270.0	106.8	1.48	48.8–233.9
XS6-003	3.3	0.8–13.0	3.1*	1.54*	1.3–7.3*
XS6-004	1.1	0.1–10.3	1.1	2.03	0.3–5.1
XS6-005	3.2	0.8–12.0	3.1*	1.54*	1.3–7.3*

[a] All results are expressed in nanograms per milliliter of plasma; outlier results are not included.
[b] Geometric mean.
[c] Geometric standard deviation; antilogarithm of standard deviation of logarithms of observations.
[d] Asterisk indicates results on duplicates were combined for statistical calculations.

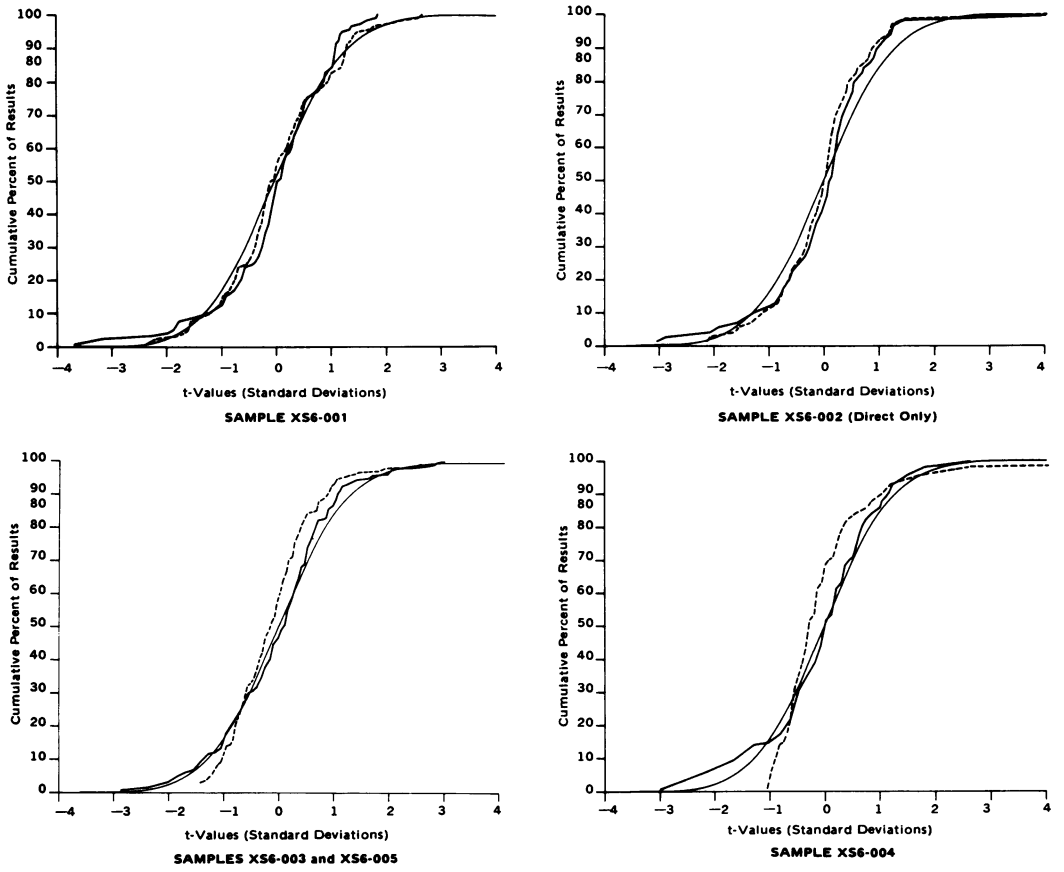

FIG. 1. *The theoretical cumulative distribution, the cumulative distribution for the standardized observations, and the cumulative distribution for the standardized observations after log transformation. Symbols: ——, Theoretical cumulative distribution function; - - -, observed cumulative distribution function; ——, observed cumulative distribution function after log transformation of results.*

(XS6-003 and XS6-005) reveal that more than one-third of the laboratories obtained results that were within 0.2 ng/ml and that more than two-thirds of the laboratories obtained results which were within 0.6 ng/ml. Over 20% of the laboratories, however, obtained results that were not within 1.0 ng/ml, and almost 10% of the laboratories obtained results that were not within 2.0 ng/ml. These figures indicate that many of the participating laboratories are capable of excellent performance with this test, but some of them need substantial improvement.

The results on paired samples (XS6-001 and XS6-004) were included as another measure of precision. The median difference between the values reported on these samples was taken as the true difference, and the variation around that difference was computed. More than one-third of the laboratories reported differences that were within 0.5 ng/ml of the median difference, and almost two-thirds of the laboratories reported differences which were within 1.0 ng/ml. More than 12% of the differences were more than 2.0 ng/ml from the median difference. The range of differences for the paired samples was larger than that for the duplicate samples, but more variation would be expected when samples of different levels are being assayed. For example, errors in the standard curve, differences in precision at different levels, or dilution errors would be more likely to be detected with paired samples than with duplicate samples.

Arithmetic and geometric statistical treatments of within-laboratory variation are presented for comparison in Table 4. The approximate limits suggested for CEA-Roche are also given (6). As expected, the geometric mean is lower than the arithmetic mean. The two standard deviation limits from the geometric treatment and the specified parameters from CEA-Roche indicate good comparability in spite of the differing assumptions concerning the distri-

TABLE 3. *Distribution analysis of CEA results*

Specimen no.	K-S	Critical values for significance levels of		Probability	$\bar{x} \pm SD^a$ ($\bar{x}_G \times/\div SD_G$)[b]	2 SD Limits (2 SD_G limits)
		0.05	0.01			
XS6-001						
Normal	0.07	0.15	0.13	<0.01	5.57 ± 1.71	2.15–8.99
Log normal	0.08			<0.01	(5.35 ×/÷1.41)	(2.69–10.64)
XS6-002 (direct)						
Normal	0.14	0.19	0.16	<0.01	116.6 ± 27.4	61.8–171.4
Log normal	0.10			<0.01	(113.9 ×/÷1.25)	(72.9–178.0)
XS6-003 and XS6-005						
Normal	0.16	0.11	0.09	>0.05	3.51 ± 1.85	−0.19–7.21[c]
Log normal	0.06			<0.01	(3.13 ×/÷ 1.61)	(1.21–8.11)
XS6-004						
Normal	0.18	0.15	0.13	>0.05	1.60 ± 1.43	−1.26–4.46[c]
Log normal	0.05			<0.01	(1.17 ×/÷ 2.28)	(0.23–6.08)

[a] Arithmetic mean plus or minus 1 standard deviation.
[b] Geometric mean multiplied or divided by 1 geometric standard deviation. Geometric standard deviation-antilog of standard deviation of logarithms of observations.
[c] Calculations from normal distribution indicates negative lower limits of 2 SD range. Negative values are impossible to obtain in reality.

TABLE 4. *Arithmetic and geometric treatment of within-laboratory variation of CEA results*

Mean	$\bar{x} \pm SD^a$ or $\bar{x}_G \times/\div SD_G$[b]	2 SD limits	Difference between 2 SD limits
Arithmetic	3.51 ± 1.12	1.27–5.75	4.48
Geometric	3.13 ×/÷ 1.25	2.01–4.87	2.86
CEA-Roche[c] (arithmetic)	3.51 ± 0.5	2.51–4.51	2.00

[a] Arithmetic mean plus or minus 1 standard deviation.
[b] Geometric mean multiplied or divided by 1 geometric standard deviation. Geometric standard deviation is the antilog of standard deviation of logarithms of observations.
[c] Mean and standard deviation suggested by CEA-Roche (8).

TABLE 5. *Summary of results on samples with elevated CEA levels by method*

Method	No. of laboratories	$\bar{x}_G$[a]	SD_G[b]	Range
CEA-Roche				
Direct	66	113.0	1.25	58–270
Indirect[c]	21	28.6	1.51	12–66
Diluted, indirect	5	55.8	1.51	33–86
Double Antibody	4	79.7	1.22	60–93

[a] Geometric mean.
[b] Geometric standard deviation, antilogarithm of standard deviation of logarithms of observations.
[c] Does not include indefinite titers such as <25, <20.

bution of results. The two standard deviation limits for CEA-Roche are smaller than the geometric limits, but it seems reasonable to assume that a laboratory could get the kind of precision indicated by Roche.

Analysis of variance revealed that there was significantly greater between-laboratory variation than within-laboratory variation in the results for samples XS6-003 and XS6-005. In other words, most of the variation was the result of poor comparability between laboratories rather than lack of precision within the laboratory. These findings indicate that better standardization and closer adherence to established procedures are needed.

Table 5 shows the geometric mean, standard deviation, and range of results on samples with elevated CEA levels by method. The geometric mean of the results obtained with the indirect CEA-Roche method performed on undiluted plasma or on diluted plasma was significantly lower ($P < 0.001$) than that obtained with the direct CEA-Roche method. The geometric mean of the results by the double antibody technique was also lower than that for results by the direct method, but the difference was not statistically significant by Student's t test.

CEA levels have been grouped according to their clinical significance. For CEA-Roche, values of 2.5 ng or less per ml are considered normal, whereas values between 2.6 and 5.0 ng/ml are intermediate and can usually be considered elevated, but should be interpreted cautiously. Values above 5.0 ng/ml are elevated

and those above 20 ng/ml are suggestive of metastatic disease. Normal levels may be different with other tests.

The results reported on the normal plasma (sample XS6-004, $\bar{x}_G$ = 1.1 ng of CEA/ml) fell into the elevated and the intermediate group as well as the normal group (Table 6). Of the reported results, 2.5% indicated that the CEA level of this specimen was elevated, and 12.5% indicated that CEA in the specimen was at the intermediate CEA level. For the samples containing slightly elevated (intermediate) levels of CEA (XS6-003 and XS6-005, $\bar{x}_G$ = 3.1 ng of CEA/ml), 21.6% of the results indicated that they contained normal levels of CEA, and 13.1% of the results indicated that they were definitely elevated. The sample that contained elevated levels of CEA (XS6-001, $\bar{x}_G$ = 5.5 ng of CEA/ml) was reported as normal by 3.4% of the laboratories and as intermediate by 30.5% of the laboratories. The sample with CEA levels considered indicative of metastatic disease (XS6-002, $\bar{x}_G$ = 106.8 ng/ml) was classified as elevated by only 3.1% of the laboratories. Almost one-fourth of all the results (24.3%) fell into the wrong group.

Most laboratories that participated in this survey used the CEA-Roche procedure. Four laboratories used a method in which a second antibody (anti-human globulin) is used to precipitate the anti-CEA and bound CEA. One laboratory used a procedure for CEA-S, a homogeneous species of CEA that is reported to be more specific for adenocarcinoma than other heterogeneic preparations of CEA (1). All results except those for CEA-S were included in the tabulations.

The CEA-Roche results and the results from laboratories that use methods other than CEA-Roche were not distributed differently. All results from the individual laboratories tended to be consistent: consistently high, consistently low, or consistently close to the means of the results reported for the CEA-Roche procedure. (See Table 7.)

We would suggest that quality control results and similar analyses of CEA determinations be calculated and plotted by the methods that we have previously described (9).

DISCUSSION

The CEA assay cannot be used as a screening test for carcinoma because of the number of carcinoma patients with normal CEA levels, as well as the number of healthy individuals with levels greater than 2.5 ng/ml. The assay has been advocated as an adjunct in diagnosis of individual patients suspected of having carcinoma or in following the progress of patients after surgery or during therapy. In both situations, serial tests are preferred to single determinations. A knowledge of the extent of interlaboratory and intralaboratory variation in test results is important in such cases so that the

TABLE 6. *Distribution of CEA results by nominal group*

Specimen no.	$\bar{x}_G$ (ng/ml)	Normal (<2.5 ng/ml)	Intermediate (2.5< x ≤5.0 ng/ml)	Elevated (5.0< x <20.0 ng/ml)	Indicative of metastasis (≥20.0 ng/ml)
XS6-001 (elevated)	5.5	3.4[a] (4/118)[b]	30.5 (36/118)	66.1 (78/118)	0
XS6-002 (metastatic disease)	106.8	0	0	3.1 (4/127)	96.9 (123/127)
XS6-003 (intermediate)	3.1	29.1 (34/117)	59.8 (70/117)	11.1 (13/117)	0
XS6-004 (normal)	1.1	85.0 (102/120)	12.5 (15/120)	2.5 (3/120)	0
XS6-005 (intermediate)	3.1	14.3 (17/19)	70.6 (84/119)	15.1 (18/119)	0

[a] Percentage of results in this nominal group.
[b] Number of results in this group/total number of results for this specimen.

TABLE 7. *CEA results obtained with methods other than CEA-Roche*

Laboratory	Method	Units	Sample No.				
			XS6-001	XS6-002	XS6-003	XS6-004	XS6-005
A	Double antibody	ng/ml	10	90	5.0	4.0	5.0
B	Double antibody	ng/ml	<1	60.4	<1	<1	<1
C	Double antibody	ng/ml	7.8	93.1	4.4	1.6	3.9
D	Double antibody	ng/ml	5.5	80	3.5	1.5	3.5
E	CEA-S	U/ml	6.47	33.8	5.77	6.50	8.00
Participant $\bar{x}_G$		ng/ml	5.5	106.8	3.1	1.1	3.1

physician can determine whether changes in CEA levels are significant. Reporting incorrect results, either consistently biased or sporadically high or low, could cause problems in the diagnosis and treatment of patients.

Nearly one-fourth of the results reported for the samples in this survey were placed in the wrong nominal group (normal, intermediate, elevated, or indicative of metastasis) as compared with the means of all the results. Twenty laboratories (16%) failed to detect differences in positivity of samples XS6-001 and XS6-004. Sixteen of these laboratories reported results which were higher than normal (>2.5 ng/ml) for XS6-004 and which therefore were similar to the results for sample XS6-001. The others reported normal levels (<2.5 ng/ml) for sample XS6-001, and their results were likewise similar for the two samples.

Many of the aberrant results could have been eliminated by strict adherence to the protocol of the test and to quality control procedures. All laboratories reported using four or more standards (including zero concentration), but 57 (46%) used three or fewer control samples. A number of the results reported for the controls were not in agreement with the expected values or ranges given.

When serial samples from an individual are being tested, results obtained with the direct (no extraction) method and the indirect (PCA extraction) method may be different. This difference is particularly important when the CEA level is around 20.0 ng/ml (4). Because of the shape of the inhibition curve for CEA, results around 20.0 ng/ml may be imprecise, since small changes in counts per minute suggest large changes in concentration. Moreover, in this range the antibody becomes saturated, and the ^{125}I-CEA is unable to compete effectively, so the resultant counts no longer reflect the amount of CEA present in the plasma. Therefore, results higher than 20.0 ng/ml obtained by the indirect assay are not valid, but 21 laboratories (17%) reported such values on sample XS6-002 of this survey.

PCA extraction is used to remove the CEA from cross-reacting material in the plasma because CEA is soluble in PCA because of its high carbohydrate content. When CEA values are greater than 20.0 ng/ml, the cross-reacting material apparently does not contribute appreciably to the results. However, PCA extraction gives lower CEA values than those obtained by direct methods, and the differences in results obtained are not consistent with all sera (11; A. C. Madsen, D. L. Brown, Z. Y. Madsen, J. T. Wu, P. F. Bray, Clin. Res. 23:115A). In this survey, the results reported for plasma that was diluted and then assayed by the indirect method were significantly lower than the results for plasmas tested by the direct method.

Several species of CEA have been reported, and assay results may vary with the standards or antiserum used (10, 11). The samples used in this survey were prepared from a single positive pool and a normal pool, so the proportions of the variants in each sample should have been relatively constant. Variation in results due to different species would not have been detected in this survey.

ACKNOWLEDGMENTS

We wish to thank Ross J. Wood, Patricia L. Rossing, and Phyllis L. Strobel of the Center for Disease Control, and J. LaMar Robbins of the LDS Hospital in Salt Lake City, Utah, for reviewing the manuscript.

LITERATURE CITED

1. **Edgington, T. S., R. W. Astarita, and E. F. Plow.** 1975. Association of an isomeric species of carcinoembryonic antigen with neoplasia of the gastrointestinal tract. N. Engl. J. Med. 293:103-107.
2. **Gold, P., and S. O. Freedman.** 1965. Demonstration of tumor-specific antigens in human colonic carcinomata by immunological tolerance and absorption techniques. J. Exp. Med. 121:439-462.
3. **Gold, P., and S. O. Freedman.** 1965. Specific carcinoembryonic antigens of the human digestive system. J. Exp. Med. 122:467-481.
4. **Lowenstein, M. S., H. Z. Kupchik, and N. Zamcheck.** 1976. Disparity between CEA-Roche "indirect" and "direct" carcinoembryonic antigen values: clinical relevance. N. Engl. J. Med. 294:1123.
5. **Ostle, B.** 1963. Statistics in research, p. 471-472. Iowa State University Press, Ames, Iowa.
6. **Roche Diagnostics.** 1974. Procedure manual for CEA-ROCHE carcinoembryonic antigen assay. Hoffman-LaRoche, Inc., Nutley, N.J.
7. **Roche Diagnostics.** 1974. CEA-ROCHE carcinoembryonic antigen assay. A clinical monograph. Hoffman-LaRoche, Inc., Nutley, N.J.
8. **Siegel, S.** 1956. Nonparametric statistics for the behavioral sciences, p. 47-52. McGraw-Hill Book Co., New York.
9. **Taylor, R. N., A. Y. Huong, and K. M. Fulford.** 1975. Measurement of variation in serologic tests. Center for Disease Control, Atlanta, Ga.
10. **Vrba, R., E. Alpert, and K. J. Isselbacher.** 1975. Carcinoembryonic antigen: evidence for multiple antigenic determinants and isoantigens. Proc. Natl. Acad. Sci. U.S.A. 72:4602-4606.
11. **Zamcheck, N., and H. Z. Kupchick.** 1976. Summary of clinical use and limitations of the carcinoembryonic antigen assay and some methodological considerations, p. 753-764. *In* N. R. Rose and H. Friedman (ed.), Manual of clinical immunology. American Society for Microbiology, Washington, D.C.