Clinical Chemistry 46:12
1894–1902 (2000)

Review

Should We Use Carbohydrate-deficient Transferrin instead of γ-Glutamyltransferase for Detecting Problem Drinkers? A Systematic Review and Metaanalysis

Kay Scouller,[1,3] Katherine M. Conigrave,[1,3,4*] Petra Macaskill,[3] Les Irwig,[3] and John B. Whitfield[2]

Background: **Carbohydrate-deficient transferrin (CDT) has been used as a test for excessive alcohol consumption in research, clinical, and medico-legal settings, but there remain conflicting data on its accuracy, with sensitivities ranging from <20% to 100%. We examined evidence of its benefit over a conventional and less expensive test, γ-glutamyltransferase (GGT), and compared the accuracy of different CDT assay methods.**

Methods: **We performed a systematic review using summary ROC analysis of 110 studies prior to June 1998 on the use of CDT in the detection of alcohol dependence or hazardous/harmful alcohol use.**

Results: **We identified several potential sources of bias in studies. In studies examining CDT and GGT in the same subjects, subject characteristics were less likely to influence the comparison. In such paired studies, the original Pharmacia CDT assay was significantly more accurate than GGT, but the modified CDTect assay did not perform as well as the original and was not significantly better than GGT. The accuracy of the AXIS %CDT assay was statistically indistinguishable from modified CDTect. Several CDT assay methods appeared promising, in particular, liquid chromatography (chromatofocusing, HPLC, fast protein liquid chromatography) and isoelectric focusing, but there were insufficient paired studies from which to draw firm conclusions.**

Conclusions: **In studies published before June 1998, the results obtained with commercially available CDT assays were not significantly better than GGT as markers of excessive alcohol use in paired studies. Further high-quality studies comparing CDTect (modified) and other CDT assays with GGT in the same subjects are needed.**

Since serum carbohydrate-deficient transferrin (CDT)[5] was first described as a test for excessive alcohol consumption, there have been well over 100 reports describing its clinical use. CDT is now used routinely in many clinical centers around the world, commercial methods for CDT measurement are readily available, and CDT has been considered for use in medico-legal decision making, such as the assessment of drunk drivers *(1, 2)*.

Nonetheless, there remain conflicting data on the accuracy of CDT. Reported sensitivities range from <20% to 100%, with typical specificities of 75–100% *(3–6)*. It is difficult to compare results from different studies because of the variety of assay techniques used to measure CDT, the different cutoff points used to define an abnormal result and to define excessive alcohol consumption, and the differences in populations studied (ranging from hospitalized alcoholics with liver disease to healthy volunteers). It is also difficult to summarize the benefit CDT offers over the conventional and less expensive measurement of γ-glutamyltransferase (GGT). Furthermore, it is

[1] Drug and Alcohol and [2] Biochemistry Departments, Royal Prince Alfred Hospital, Missenden Rd., Camperdown, NSW 2050, Australia.

Department of [3] Public Health and Community Medicine and [4] Departments of Medicine and Psychological Medicine, University of Sydney, Sydney, NSW 2006, Australia.

*Address correspondence to this author at: Drug and Alcohol Department, Royal Prince Alfred Hospital, Missenden Rd., Camperdown, NSW 2050, Australia. Fax 61-2-9515-8970; e-mail katec@med.usyd.edu.au.

Received October 26, 1999; accepted September 19, 2000.

[5] Nonstandard abbreviations: CDT, carbohydrate-deficient transferrin; GGT, γ-glutamyltransferase; SROC, summary receiver-operating characteristic; FPLC, fast protein liquid chromatography; and IEF, isoelectric focusing.

not readily apparent which technique for the measurement of CDT is the most accurate.

Comparing assay methods when they have been assessed in different studies is potentially misleading because apparent differences may be attributable to the study population, study design, and biases rather than to true differences between assays. Therefore, the best evidence for comparing assays comes from "paired" studies that evaluate two or more assays using the same sample and the same reference standard.

We report the results of a systematic review of studies on the diagnostic value of CDT as a marker of excessive alcohol consumption published before June 1998. Our aim is to assist clinicians and researchers to better evaluate the available data on factors associated with the accuracy of CDT, to evaluate the relative accuracy of classes of CDT assay, and to compare CDT with GGT as a single marker of excessive alcohol consumption. Particular attention is paid to studies where two or more tests were compared in the same group.

Materials and Methods

LITERATURE RETRIEVAL

A search was made, using MEDLINE and CINAHL, for all English-language publications in refereed journals, books, and monographs published before June 1998 that described the use of CDT in humans in relation to alcohol consumption. Reports presenting original data were considered for inclusion if they provided CDT values from which sensitivity and specificity of CDT could be calculated and provided data on drinking behavior (alcohol consumption, diagnosis of abuse or dependence, or other relevant criteria for assessing drinking). Reports were excluded if they dealt primarily with fetal CDT concentrations, genetic variants of transferrin or the carbohydrate-deficient glycoprotein syndrome, or transferrin variations attributable to carcinoma.

In addition, published abstracts from several alcohol-related conferences up to May 1998 were perused, including all meetings of the International Society for Biomedical Research on Alcoholism and relevant proceedings from the European and Japanese Societies for Biomedical Research on Alcoholism, the Research Society on Alcoholism, and the North Carolina Alcoholism Research Authority.

DATA EXTRACTION

Reports were evaluated by one of the authors (K.S.), using a structured form developed by the authors with reference to published guidelines for the evaluation of diagnostic tests *(7–13)*. Information was collected on population characteristics, study design, and assay methods, as shown in Tables 1–3. Where there was uncertainty over interpretation or coding of published data, the report was referred to one (K.M.C.) or more authors for resolution. Eight studies (7.5%) were independently reviewed by a second author (K.M.C.), and there was >95% agreement on all items coded, with 100% agreement on data relating directly to test accuracy.

STATISTICAL ANALYSIS

A descriptive analysis of the 110 eligible studies (derived from 108 reports) was conducted to determine the scope and nature of the available data. Where possible, the sensitivity and specificity were computed for each assay in each study (105 of 144 sets of test results).

For any diagnostic test, there is a trade-off between sensitivity and specificity, so that if a lower threshold is used to define an abnormal test, there will be more true cases detected (higher sensitivity) but more false positives (lower specificity). The reverse is true if a higher threshold is used. Metaanalytic techniques for diagnostic tests have been developed that take into account this trade-off *(14)*. These techniques are an extension of the conventional ROC analyses. A summary ROC (SROC) curve is estimated using the log odds ratio as a global measure of test performance for each study. The log odds ratio incorporates both sensitivity and specificity and takes into account the trade-off between them.

For each assay in each study the log odds ratio (D) is computed as:

$$D = \text{logit}(sens) - \text{logit}(1 - spec)$$

$$= \log_e[sens/(1 - sens)] - \log_e[(1 - spec)/spec]$$

where *sens* is test sensitivity and *spec* is specificity (both with values in the range 0–1). An odds ratio is then obtained by exponentiating the log odds ratio. Higher odds ratios reflect higher test accuracy.

To form a SROC curve for an assay, the method developed by Moses et al. *(14)* was used to combine the log odds ratios (D) for that assay across the different studies. These values were combined by fitting the linear regression model: $D = \alpha + \beta S$. In this formula, S is a proxy for test threshold and is computed for each study as $S = \text{logit}(sens) + \text{logit}(1 - spec)$. The parameter β is used to assess whether overall test accuracy depends on test threshold (i.e., the cutoff chosen to define an abnormal test). None of our analyses supported the inclusion of S in the model; hence, overall test performance was assumed to be constant across a range of test thresholds (a symmetric SROC) and was defined by a constant log odds ratio (α). The SROC curve can be displayed graphically by plotting the predicted sensitivity across a range of values of 1 − specificity *(14)*.

Covariates, such as study design and gender, were also considered for inclusion in the regression model. Backward elimination was used to assess whether these variables were associated with test accuracy. Indicator variables were used to describe categories of study design (case control vs prospective) and gender (all male vs all female vs mixed or unknown). Models were unweighted to provide estimates consistent with a random-effects model *(14, 15)*.

After we derived summary data on test accuracy for CDT and GGT from all of the studies, our subsequent analyses focused on paired studies, i.e., studies that compared different assays or different confounding factors, such as gender and liver disease, in the same subject group. For studies comparing two tests in the same subjects, we computed indicators of test accuracy (*D*) for each of the pair of assays in each study. We used a paired *t*-test to determine whether there was a significant difference in test accuracy between the two assays performed in the same subject group. The difference in test accuracy between the two assays was expressed as the difference in the mean log odds ratios for these assays, which can be interpreted as a ratio of odds ratios. This is analogous to the manner in which polytomous logistic regression uses the ratio of odds ratios to compare estimates *(16)*.

We used the same approach to investigate whether the performance of an assay was associated with the cutoff chosen for the reference standard. For this analysis, we used studies that provided test results for three or more alcohol use groups (e.g., a study that assessed the ability of CDT to differentiate those drinking 60–79 g/day from light drinkers, and those drinking ≥80 g/day from light drinkers). Paired *t*-tests were also used to examine the effect of gender and liver disease on test performance by assessing within-study differences where results were stratified for these factors.

Results

LITERATURE RETRIEVAL

A total of 158 reports were identified, of which 112 fulfilled the inclusion criteria. One report could not be retrieved *(17)*. Where a single report combined two different studies *(18, 19)*, the data were analyzed as two separate studies. Where a report described the same study population as another report with common authors, the fuller set of results was retained and its pair was excluded *(20–24)*. This reduced the eligible reports to 108 with a total of 110 study samples.

Forty-six relevant published abstracts from 17 conferences up to 1998 were examined, and of these 31 (67%) had subsequently been published as journal articles. Because insufficient data were presented, no abstracts were included in the metaanalysis.

SYSTEMATIC REVIEW OF JOURNAL ARTICLES

Table 1 provides an overview of the information provided in the 110 eligible studies. The median sample size was 101 subjects. The mean subject age in >50% of studies was in the 40s (interquartile range, 41–48 years). When we averaged the unweighted proportions for each study, the mean percentage of male subjects was 70% and of females was 30%. The majority of studies were based in Western Europe, Scandinavia, and/or the US and Canada (37%, 35%, and 16%, respectively) with smaller numbers of

Table 1. Population tested.

		No. of studies (n = 110)	% of 110 studies
Median sample size (interquartile range)	101 (59–174)		
Mean age of samples (interquartile range), years	43.4 (40.9–48.1)		
Gender (mean %)			
M	70		
F	30		
Country in which study was performed			
US/Canada/Australia/New Zealand		24	22
Scandinavia		38	35
Western Europe		41	37
Japan		5	5
Other[a]		2	2
Spectrum of alcohol intake			
Full range		31	28
Light or non-problem drinkers vs hazardous/abusive or dependent		52	47
Dependent drinkers only		12	11
Abstaining vs drinking alcoholics		8	7
Other		7	6
Recruitment site[b]			
Clinical setting			
Patients with known alcohol problems		90	82
Patients without known alcohol problems		44	40
Community/staff population		69	63
Other population		1	1
Subjects with physical complications attributable to alcohol consumption		40	36

[a] "Other" includes a South Asian population living in Britain and a combined study performed in Sweden and US.
[b] The categories are not mutually exclusive; therefore, percentages add up to more than 100%.

studies involving Australia and New Zealand, Japan, and/or other countries (6%, 5%, and 2%, respectively).

A variety of recruitment sites were used, with 82% of studies including patients with known alcohol problems and 63% involving community or staff members. Of the 88 studies in which a single blood sample from inpatients was analyzed, 50 (57%) reported that all or most venipunctures were performed within 72 h of admission. In five studies (6%), all or most blood samples were taken later than this, whereas 32 (36%) did not report on timing of venipuncture. In the additional nine studies (9%) in which blood was taken serially, the results closest to the time of admission were used for the metaanalysis. Time since last drink was given in 43 of the 96 (45%) studies involving inpatients, and in only 2 (5%) of these were samples taken more than 1 week after the last drink.

STUDY QUALITY

The quality of reports was variable, with the method of subject selection documented in only 35% of studies, ≤50% of studies documenting gender or age comparability between cases and controls, <50% documenting criteria for diagnosing alcohol abuse or dependence, and <10% providing a statement about blinding in relation to test and diagnosis (Table 2). Stratification of results by age was provided in only 11% of studies.

Subjects were most often recruited because they were known a priori to fulfill certain consumption or diagnostic criteria, but in 32% of studies subjects were first recruited, and then a consumption threshold was subsequently used to define consumption categories. Selection of consecutive patients was done in only 22% of studies, whereas 56% used a case-control design (i.e., tests were evaluated in a diseased population and a separate control group), which has been shown to be the most important source of bias in studies of diagnostic tests *(13)*. However, despite this concern, the current data revealed no significant influence of study type (case-control vs population sampling) on the odds ratio estimate.

Studies generally used as their reference standards

Table 2. Analysis of study quality in systematic review of 110 studies of CDT as a test for excessive alcohol use.

	No. of studies (n = 110)	%[a]
Study design and subject selection		
Case-control design avoided	49	45
Recruitment of consecutive patients used	24	22
Subject selection method recorded	38	35
Evidence that verification bias[b] was avoided	25	23
Gender comparability stated between those positive and negative on reference standard	47	50[c]
Age comparability stated between those positive and negative on reference standard	36	38[c]
Spectrum of race stated	16	15
Information provided on the establishment of a reference standard		
Alcohol abuse, dependence, or hazardous/harmful drinking used as a diagnostic criterion	84	76
Criteria for diagnosing alcohol abuse/dependence stated	35	42[d]
Method for measuring alcohol consumption stated	72	66
Verification of alcohol intake self-report obtained	16	15
Alcohol consumption reference interval constructed by applying a threshold to an underlying continuum of drinkers	35	32
Information provided on the independence of the reference standard and CDT test		
Independent categorization of test and reference standard		
Test performed blind to reference standard (explicitly stated)	7	6
Reference standard established blind to test result (explicitly stated)	8	7
Compared tests assessed independently of each other (explicitly stated)	4	5[e]
Reporting of results		
Sensitivities and specificities reported or calculable	130	90[f]
Results presented on all subjects except where explanation of exclusion provided	84	75
Stratification of results		
By gender	66	60
By age	12	11
By race	5	5

[a] Percentages shown are of the total 110 studies unless otherwise indicated.

[b] Verification bias occurs when the reference standard is not assessed independently of the test results, e.g., when only persons with positive test results have their alcohol consumption recorded.

[c] Percentage of 95 studies: 16 of the 110 studies had only reference-positive or only reference-negative subjects.

[d] Percentage of the 84 studies that used hazardous/harmful drinking, or alcohol abuse or dependence as diagnostic criteria.

[e] Percentage of the 84 studies that compared two or more tests in the same subjects.

[f] Percentage of the 144 instances where CDT tests were performed (some studies assayed CDT by more than one method).

either the presence of usual self-reported alcohol consumption above a threshold value (e.g., mean alcohol consumption >60 g/day in the recent month or week) or the diagnosis of an alcohol use disorder. Alcohol dependence was more often used as a diagnostic category than was hazardous/harmful drinking or abuse (66% of studies compared with 26%). Criteria for the diagnosis were stated in 42% of such studies, whereas the method used to ascertain whether subjects fulfilled the criteria (e.g., structured interview or medical records review) was stated in only 27%.

In 56 studies (51%), both CDT results and GGT results were presented for the same subjects. However, data enabling full cross-classification of results were presented in only two of these studies, so that it was not generally possible to determine whether subjects with increased GGT were more or less likely to have increased CDT.

comparison between different alcohol consumption thresholds

When alcohol consumption was used as a reference standard, subjects were generally divided into two or three categories, usually (*a*) abstainers and light or moderate drinkers, whose upper limit of consumption was up to 60.0 g of ethanol/day (mean, 25 g/day); (*b*) hazardous or abusive drinkers, whose upper limit of consumption ranged from 40 to 143 g/day (mean, 75 g/day); and (*c*) heavy, harmful, or dependent drinkers, whose consumption was higher than that of subjects in other categories.

To determine whether use of different consumption thresholds influenced test accuracy, the nine reports that used three consumption categories were examined. These studies, which provided test results for both a higher and lower threshold, were used to calculate SROC curves for both thresholds. Because the two thresholds were compatible with the same curve ($t_8 = 0.73$; $P = 0.48$), the cutoff at or closest to 40 g of alcohol/day was chosen for further analyses.

influence of gender, age, and liver disease on cdt concentrations

On unpaired analysis, gender was not found to have a significant effect on the performance of most CDT assays or of GGT. It should be noted that studies generally used different test thresholds for men and for women. The performance of only one CDT assay [HPLC or fast protein liquid chromatography (FPLC)] was associated with gender ($P = 0.05$).

Sixteen studies with a median sample size of 85 (interquartile range, 71–147) presented stratified CDTect (modified)[6] results for males and females. A paired *t*-test revealed no statistically significant difference in test accuracy between the genders ($t_{15} = -0.65$; $P = 0.53$; Fig. 1A).

Because there were only 12 studies that provided stratified results by age and these studies used different assay techniques, there were not sufficient data to perform paired comparisons for any one CDT assay by age category.

Five studies with a median sample size of 108 (interquartile range, 82–112) presented stratified CDTect (mod-

[6] Pharmacia CDTect assay, modified method, subsequently supplied by AXIS.

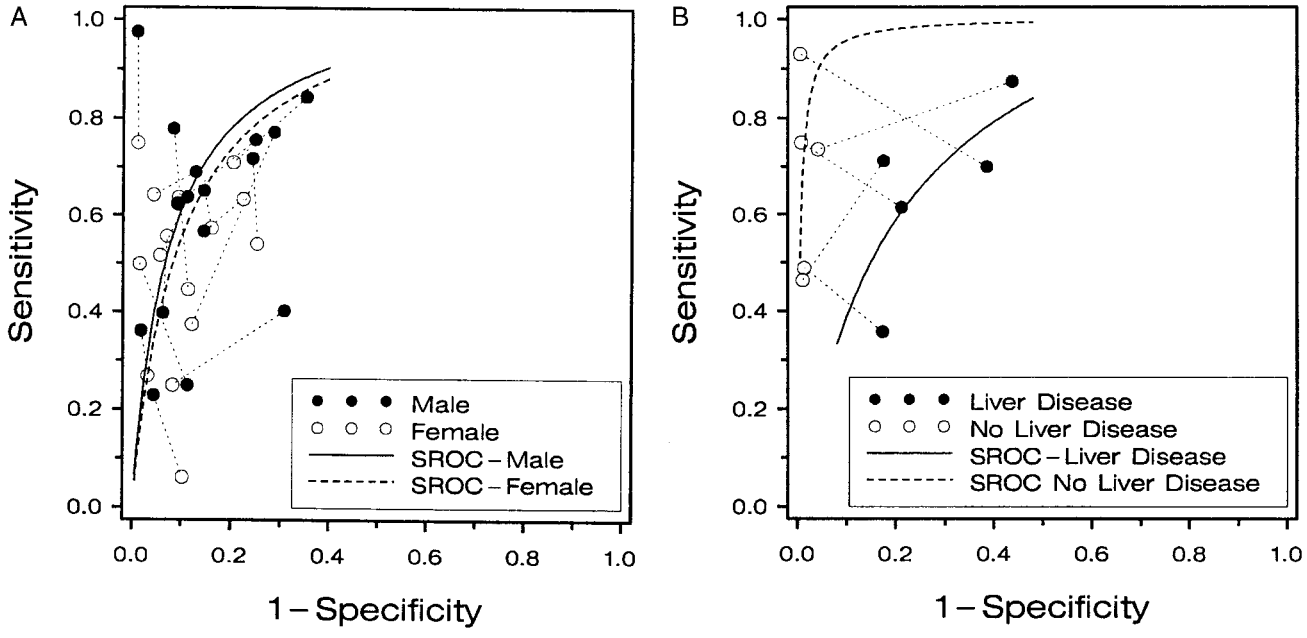

Fig. 1. SROC curves showing the effect of gender (*A*) and liver disease (*B*) on the CDTect (modified) assay results in those studies where stratified results were presented.

Dotted lines indicate results from the same study.

ified) results for subjects with and without liver disease. CDTect performed significantly better in patients without liver disease than in those patients with liver disease (t_4 = 4.25; P = 0.01; Fig. 1B).

PERFORMANCE OF DIFFERENT ASSAYS AND COMPARISONS BETWEEN ASSAYS

Categorization of assay methods. CDT assay methods were divided into seven categories (Table 3). Where an assay did not fit the major categories, the assay type was recorded as "other" and the assay category that it most closely resembled was noted. The accuracy of the less common assay method was then compared with the major assay category using SROC analysis. In all such cases there was no statistically significant difference in performance, so the less common assay was grouped with that major category for further analyses.

When we derived SROC curves for each CDT assay method and for GGT, none of the slopes of D on S were significantly different from zero, indicating that the SROC curves were defined by a constant odds ratio that could be used as an overall measure of test accuracy.

Unpaired comparisons. The odds ratio from SROC analysis for each CDT assay method and for GGT are presented in Table 3. All of the CDT methods yielded higher summary odds ratios than GGT, although for the AXIS[7] method and CDTect (modified), the confidence intervals overlapped and touched, respectively, the confidence interval for GGT. The odds ratios suggest that the FPLC/HPLC methods, chromatofocusing (the latter was used in only two studies), CDTect (original),[8] and isoelectric focusing (IEF) methods have higher odds ratios than GGT.

Paired comparisons. Where there were five or more studies providing stratified results, paired comparisons between assay types were performed:

(*a*) CDT compared with GGT. In the six studies where CDTect (original) was compared with GGT in the same subjects, CDTect (original) was found to perform significantly better (Table 4 and Fig. 2A). In contrast, in the 19 studies where CDTect (modified) and GGT were assayed in the same subjects, the performance of CDTect (modified) did not differ significantly from that of GGT (Table 4 and Fig. 2B). IEF methods appeared to perform somewhat better than GGT (paired analysis with seven studies), but this difference was not statistically significant (Table 4 and Fig. 2C). Data were not available to perform paired comparisons between HPLC/FPLC methods and GGT.

Table 3. Summary of accuracy of individual assays in unpaired SROC analyses of 105 studies.[a]

	No. of studies	Odds ratio	95% CI[b]
GGT	36	5.7	4.0–8.0
CDT assays			
CDTect (original version)	12	79.3	33.8–186.2
CDTect (modified version)	43	12.0	8.0–17.9
AXIS method (% CDT)	15	10.6	4.6–24.1
MAEC with turbidimetry	2	33.5	—[c]
HPLC or FPLC	8	117.8	27.4–506.2
Chromatofocusing	2	430.5	—[c]
IEF methods[d]	23	30.0	16.6–54.4
IEF with protein staining	4	10.4	3.0–36.4
IEF with immunofixation	14	35.5	15.9–78.9
IEF with immunoblotting	5	44.0	15.3–126.2

[a] Only tests where both sensitivity and specificity were provided for calculable are included in SROC analysis. Hence, of 144 cases where CDT testing was performed, 105 are included.

[b] CI, confidence interval; MAEC, modified anion-exchange chromatography.

[c] Only two observations were available; therefore, no confidence intervals were calculated.

[d] The three IEF methods are considered together.

(*b*) CDTect (modified) compared with AXIS. Six studies reported CDT results as assayed by the two commercially available methods, CDTect (modified) and AXIS, on the same subjects. There was no significant difference in accuracy between these methods (Table 4 and Fig. 2D).

(*c*) IEF compared with CDTect (modified). The IEF methods yielded significantly higher accuracy than CDTect (modified) in paired comparisons in six studies (Table 4).

Discussion

The technique of SROC metaanalysis provides a valuable tool for pooling and comparing results from studies on the diagnostic accuracy of tests, while taking into account differences in test reference interval, study type, and subject characteristics. A variety of factors may disturb comparisons of test performance in unpaired studies. For example, studies with a mean subject age of <30 years may produce lower accuracy estimates than studies of middle-aged patients *(3, 26)*. If one study uses a case group of hospitalized alcoholics with a control group of abstainers, it is likely to produce higher accuracy estimates than a study comparing consecutively recruited family-medicine patients who are then divided into high- or low-level drinkers using a predetermined consumption threshold *(13)*. Despite these concerns, in the current data we could not demonstrate a significant difference in results between studies that recruited subjects because they were known to fulfill a priori categories and studies that recruited a continuum of patients. Nonetheless, examination of paired test results (i.e., where results of two different tests are available for each subject) provides the best way of minimizing the influence of extraneous fac-

[7] CDT isoforms as a proportion of total transferrin (%CDT), supplied by AXIS Biochemicals AFA, Oslo, Norway.

[8] Pharmacia CDT assay (Pharmacia and Upjohn Diagnostics AB, Uppsala, Sweden), based on original method by Stibler et al. *(25)*.

tors and minimizing biases attributable to poor quality of study design and execution.

In studies on the accuracy of biological markers of alcohol use, the reference standard generally involves self-report of alcohol consumption, symptoms of dependence, or related problems. Studies vary in the criteria used to determine problem drinking. Some misclassification of subjects is likely to occur, which is likely to cause an underestimation of test accuracy *(27)*. However, identification of whether one test is better than another is not likely to be biased when the tests are compared within the same patient.

In studies where CDT and GGT were measured in the same subjects, the original CDTect method was shown to be significantly more accurate than GGT. In contrast, CDTect (modified), which is one of the commercially available methods for measuring CDT, offered no significant benefit over GGT. The accuracy of the other commercially available method, AXIS %CDT, did not differ significantly from that of CDTect (modified).

Other assay types showed considerable promise on unpaired comparison, in particular chromatofocusing and HPLC/FPLC methods. The odds ratio for HPLC/FPLC methods was 20-fold higher than that for GGT in unpaired comparisons, which translates into a sensitivity of 93% for HPLC/FPLC compared with 39% for GGT at test cutoffs that give a specificity of 90% for both. However, insufficient reports were available to compare chromatofocusing or HPLC/FPLC with GGT in the same subjects, so it remains possible that the difference in test accuracy has been exaggerated by differences in subject characteristics or study methods. IEF methods were significantly more accurate than CDTect (modified), but they were not demonstrated to be significantly more accurate than GGT on paired comparison.

Differences in the accuracy of CDT assay methods may result from differences in the ways assays treat the isotransferrins, particularly trisialotransferrin. This is the major difference between the original and modified Pharmacia methods *(28)* and illustrates the impact that variation in an assay method and hence in the analyte can have on test performance. In a similar way, immunoassays may differ in their clinical performance through measurement of different epitopes, of C- and N-terminal fragments, or of different subunits of a multimeric protein.

effect of gender and liver disease

In studies that reported separate results for CDTect (modified) in males and females, gender did not influence test accuracy. These studies generally used different test reference intervals for men and for women. There were insufficient data to perform a paired analysis for gender in other CDT assays, but it appears likely that the establishment of different test thresholds adequately compensates for differences in baseline CDT concentrations.

A major limitation of GGT is that it is affected by liver disease of any cause and by several medications. However, where CDTect (modified) results were stratified by liver disease, test performance was significantly reduced in the presence of liver disease. There were insufficient paired data to establish whether CDT is more accurate than GGT in this setting.

practical implications

Measurement of CDT by most methods (with the possible exception of the AXIS method) costs at least twice as much as measurement of GGT, and from the evidence of this study, routine use of commercial CDT testing as a single test for alcohol consumption is not indicated. CDTect (original) was significantly more accurate than CDTect (modified) or GGT; however, practical constraints may limit its clinical use. CDTect (original) was modified to form the new commercially available test, CDTect (modified), because of buffer instability. It remains a challenge to maintain the test performance of the original assay in a form that is readily available and easily used. Noncommercial CDT assay methods may be more accu-

Table 4. Comparisons between assays using paired SROC analysis.

Tests compared	No. of studies	Median sample size (interquartile range)	Estimate of OR[a]		Ratio of OR[b] (95% CI)	P[c]
CDTect (original) vs GGT	6	86 (63–233)	CDTect (original)	58.1	27.1 (3.8–193)	0.008
			GGT	2.1		
CDTect (modified) vs GGT	19	179 (88–346)	CDTect (modified)	9.3	1.3 (0.6–2.7)	0.51
			GGT	7.5		
IEF methods vs GGT	7	86 (64–167)	IEF	19.1	3.4 (0.6–19.7)	0.13
			GGT	5.5		
AXIS vs CDTect (modified)	6	173 (107–212)	AXIS	8.3	1.2 (0.4–3.1)	0.68
			CDTect (modified)	7.0		
IEF method vs CDTect (modified)	6	87 (64–107)	IEF	28.2	3.4 (1.3–9.0)	0.03
			CDTect (modified)	8.4		

[a] OR, odds ratio; CI, confidence interval.
[b] Ratio of the odds ratios is presented for the test with the higher odds ratio over the test with the lower odds ratio.
[c] *P* based on paired *t*-tests with n − 1 degrees of freedom.

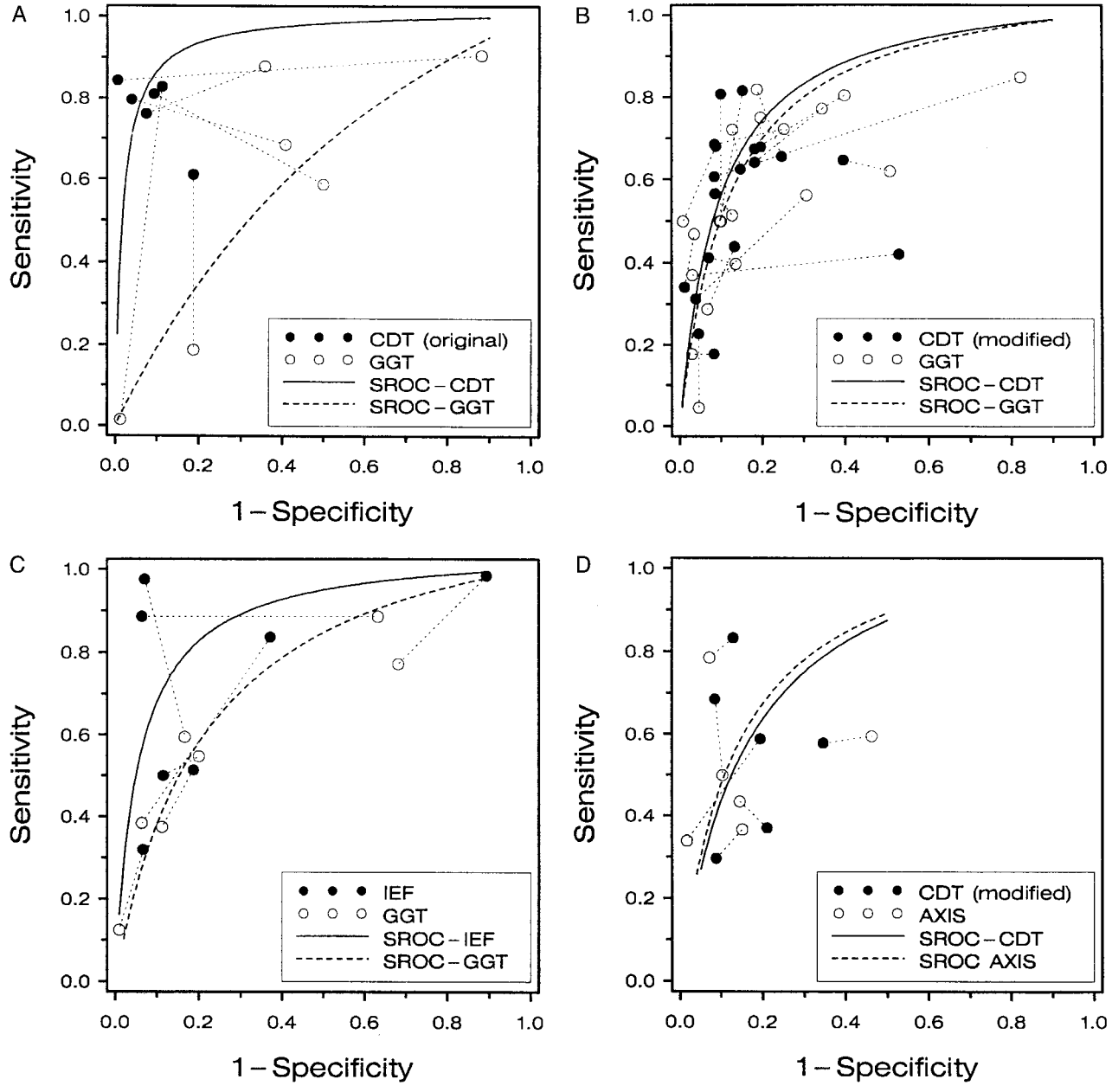

Fig. 2. SROC curves for assays compared in the same subject group.

Dotted lines indicate results from the same study. (*A*), CDTect (original) and GGT (n = 6 studies). (*B*), CDTect (modified) and GGT (n = 19 studies). (*C*), IEF and GGT (n = 7 studies). (*D*), CDTect (modified) and AXIS methods (n = 6 studies).

rate than the commercial methods, but their use may stretch the resources of many clinical laboratory facilities.

There may be situations where CDT testing by the commercially available assays is likely to be more accurate than GGT assay, e.g., in subjects using medications such as antiepileptics, where GGT would be more likely to produce false-positive results. Furthermore, there have been promising reports on the use of CDT in combination with GGT, where the combined tests have increased sensitivity without significant compromise in specificity *(3, 18, 29, 30)*. CDT also shows promise in prospective monitoring of an individual's drinking *(31–35)*, but these aspects of CDT use were outside the scope of the current report.

In conclusion, there is a tendency to enthusiastically adopt new tests that show promising results. The findings of the current study emphasize the need for ongoing and careful assessment of the evidence and for high-quality research. The original Pharmacia CDT assay, which is not now commonly used, was the only CDT assay method shown to be significantly better than GGT on paired testing. CDT as measured by commercially available assay was not shown to be a significantly better test than GGT when

performed on the same subjects. Although the use of noncommercial assay methods such as chromatofocusing and HPLC/FPLC may be more accurate than commercial methods, these may be feasible only in laboratories with a high volume of testing. Further high-quality studies using comparisons of tests in the same subjects are indicated.

This work was funded by the Motor Accidents Authority of New South Wales and the New South Wales Health Department. The reports included in the metaanalysis are listed as an Appendix available at *Clinical Chemistry Online,* and can be accessed through the December Table of Contents (http://www.clinchem.org/content/vol46/issue12).

References

1. Morgan MY, Major MG. The use of serum carbohydrate-deficient transferrin in the assessment of 'high risk offenders' in Great Britain. Alcohol Alcohol 1996;31:625–8.
2. Gjerde H, Mørland J. Concentrations of carbohydrate-deficient transferrin in dialysed plasma from drunken drivers. Alcohol Alcohol 1987;22:271–6.
3. Nystrom M, Peräsalo J, Salaspuro M. Carbohydrate-deficient transferrin (CDT) in serum as a possible indicator of heavy drinking in young university students. Alcohol Clin Exp Res 1992;16:93–7.
4. Lof K, Seppa K, Itala L, Koivula T, Turpeinen U, Sillanaukee P. Carbohydrate deficient transferrin (CDT) as an alcohol marker among female heavy drinkers: a population based study. Alcohol Clin Exp Res 1994;18:889–94.
5. Xin Y, Rosman AS, Lasker JM, Lieber CS. Measurement of carbohydrate-deficient transferrin by isoelectric focusing/Western blotting and by micro anion-exchange chromatography/radioimmunoassay: comparison of diagnostic accuracy. Alcohol Alcohol 1992;27:425–33.
6. Lof K, Koivula T, Seppa K, Fukunaga T, Sillanaukee P. Semiautomatic method for determination of different isoforms of carbohydrate-deficient transferrin. Clin Chim Acta 1993;217:175–86.
7. Bruns DE, Huth EJ, Magid E, Young DS. Toward a checklist for reporting of studies of diagnostic accuracy of medical tests. Clin Chem 1997;43:2211.
8. Irwig L, Tosteson A, Gatsonis C, Lau J, Colditz G, Chalmers T, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994;120:667–76.
9. Jaeschke R, Guyatt G, Sackett D. Users' guide to the medical literature. III. How to use an article about a diagnostic test. Part B. What are the results and will they help me in caring for my patients? JAMA 1994;271:703–7.
10. Mulrow C, Linn W, Gaul M, Pugh JA. Assessing quality of a diagnostic test evaluation. J Gen Intern Med 1989;4:288–95.
11. Oxman A, Cook D, Guyatt G. Users' guide to the medical literature. VI. How to use an overview. JAMA 1994;272:1367–71.
12. Puech M. Evaluations in Australia of the accuracy of diagnostic tests: a critical appraisal [MPH Treatise]. Sydney, Australia: University of Sydney, Department of Health, 1995:70pp.
13. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061–6.
14. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993;12:1293–316.
15. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol 1995;48:119–30.
16. Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley, 1989:307pp.
17. Prasad R, Stout RL, Coffin D, Smith J. Analysis of carbohydrate deficient transferrin by capillary zone electrophoresis. Electrophoresis 1997;18:1814–8.
18. Helander A, Carlsson AV, Borg S. Longitudinal comparison of carbohydrate-deficient transferrin and gamma-glutamyl transferase: complementary markers of excessive alcohol consumption. Alcohol Alcohol 1996;31:101–7.
19. Stowell LI, Fawcett JP, Brooke M, Robinson GM, Stanton WR. Comparison of two commercial test kits for quantification of serum carbohydrate-deficient transferrin. Alcohol Alcohol 1997; 32:507–16.
20. Fagerberg B, Agewall S, Urbanavicius V, Attvall S, Lundberg PA, Lindstedt G. Carbohydrate-deficient transferrin is associated with insulin sensitivity in hypertensive men. J Clin Endocrinol Metab 1994;79:712–5.
21. Bell H, Tallaksen CME, Try K, Haug E. Carbohydrate-deficient transferrin and other markers of high alcohol consumption: a study of 502 patients admitted consecutively to a medical department. Alcohol Clin Exp Res 1994;18:1103–8.
22. Borg S, Beck O, Helander A, Voltaire A, Stibler H. Carbohydrate-deficient transferrin and 5-hydroxytryptophol. In: Litten R, Allen J, eds. Measuring alcohol consumption. Totowa, NJ: Humana Press, 1992:149–59.
23. Fletcher LM, Kwoh-Gain I, Powell EE, Powell LW, Halliday JW. Markers of chronic alcohol ingestion in patients with nonalcoholic steatohepatitis: an aid to diagnosis. Hepatology 1991;13:455–9.
24. Spies CD, Neuner B, Neumann T, Blum S, Muller C, Rommelspacher H, et al. Intercurrent complications in chronic alcoholic men admitted to the intensive care unit following trauma. Intensive Care Med 1996;22:286–93.
25. Stibler H, Borg S, Joustra M. Micro anion exchange chromatography of carbohydrate-deficient transferrin in serum in relation to alcohol consumption (Swedish patent 8400587-5). Alcohol Clin Exp Res 1986;10:535–44.
26. Agelink M, Dirkes-Kersting A, Zeit T, Bertling R, Malessa R, Kleiser E. Sensitivity of carbohydrate-deficient transferrin (CDT) in relation to age and duration of abstinence. Alcohol Alcohol 1998;33:164–7.
27. Valenstein PN. Evaluating diagnostic tests with imperfect standards. Am J Clin Pathol 1990;93:252–8.
28. Arndt T, Hackler R, Kleine TO, Gressner AM. Validation by isoelectric focusing of the anion-exchange isotransferrin fractionation step involved in determination of carbohydrate-deficient transferrin by the CDTect assay. Clin Chem 1998;44:27–34.
29. Anton RF, Moak DH. Carbohydrate-deficient transferrin and γ-glutamyltransferase as markers of heavy alcohol consumption: gender differences. Alcohol Clin Exp Res 1994;18:747–54.
30. Behrens UJ, Worner TM, Lieber CS. Changes in carbohydrate-deficient transferrin levels after alcohol withdrawal. Alcohol Clin Exp Res 1988;12:539–44.
31. Borg S, Helander A, Voltaire Carlsson A, Hogstrom Brandt A-M. Detection of relapses in alcohol-dependent patients using carbohydrate-deficient transferrin: improvement with individualized reference levels during long-term monitoring. Alcohol Clin Exp Res 1995;19:961–3.
32. Borg S. Treatment of alcohol dependence: experiences of using biological markers in monitoring and prevention of relapse. Alcohol Alcohol 1996;31:621–4.
33. Anton RF, Moak DH, Latham P. Carbohydrate-deficient transferrin as an indicator of drinking status during a treatment outcome study. Alcohol Clin Exp Res 1996;20:841–6.
34. Huseby N, Bjordal E, Nilssen O, Barth T. Utility of biological markers during outpatient treatment of alcohol-dependent subjects: carbohydrate-deficient transferrin responds to moderate changes in alcohol consumption. Alcohol Clin Exp Res 1997;21: 1343–6.
35. Mitchell C, Simpson D, Chick J. Carbohydrate deficient transferrin in detecting relapse in alcohol dependence. Drug Alcohol Depend 1997;48:97–103.