COMBINING CARBOHYDRATE-DEFICIENT TRANSFERRIN AND GAMMA-GLUTAMYLTRANSFERASE TO INCREASE DIAGNOSTIC ACCURACY FOR PROBLEM DRINKING

JACK CHEN[1]**, KATHERINE M. CONIGRAVE[2], PETRA MACASKILL[3], JOHN B. WHITFIELD[4] and LES IRWIG[3] ON BEHALF OF THE WORLD HEALTH ORGANIZATION AND THE INTERNATIONAL SOCIETY FOR BIOMEDICAL RESEARCH ON ALCOHOLISM COLLABORATIVE GROUP***

[1]School of Public Health, University of Sydney, [2]Drug Health Services, Royal Prince Alfred Hospital, Departments Psychological Medicine and Medicine and School of Public Health, University of Sydney, [3]Screening and Test Evaluation Program, School of Public Health, University of Sydney, and [4]Department of Biochemistry, Royal Prince Alfred Hospital and Department of Medicine, University of Sydney, Sydney, New South Wales, Australia

Abstract — Aim: To examine methods for combining quantitative results for serum carbohydrate deficient transferrin (CDT), gamma-glutamyltransferase (GGT) and/or aspartate aminotransferase (AST), and refining these by inclusion of patient characteristics. **Methods:** Data from 1684 subjects, recruited from the general population, abstainer groups and alcohol treatment centres (participants in the five nations WHO/ISBRA study of biological markers of alcohol use), were used to develop clinical rules for combining results of GGT, AST and CDT. The algorithm derived by Sillanaukee and Olsson was tested, and compared with new algorithms derived by logistic regression and discriminant analysis. Diagnostic accuracy was assessed by area underneath the receiver operator characteristic curve. Effects of adding gender and clinical information to the algorithm were estimated. **Results:** The predictive ability of combination rules derived from the two studies and by two different statistical techniques was remarkably consistent. For men, combining lnCDT and lnGGT provided the best accuracy for detecting daily consumption of 60 g ethanol or more in the past 30 days. For women, GGT alone provided the best accuracy for that consumption level. Clinical variables added significantly to the diagnostic accuracy of the models for both men and women, and conversely the test results modified the probability of problem drinking as assessed from clinical data alone. A graphic method was produced to help clinicians estimate probabilities for consumption of 60 g or more per day. **Conclusions:** Combining biochemical markers enhances detection of problem drinking in men but not in women. Information on clinical variables increases the ability to correctly detect problem drinking.

INTRODUCTION

The use of multiple biochemical tests supplemented by clinical, demographic or physiological information, may offer diagnostic improvements over single tests. Biological markers play an important role in the early detection of alcohol problems, monitoring of rehabilitation and evaluation of intervention programmes. The tests most often used to detect problem drinking are the liver enzymes serum gamma-glutamyltransferase (GGT), aspartate aminotransferase (AST) and alanine aminotransferase (ALT) and, more recently, the protein isoform carbohydrate deficient transferrin (CDT). It has been well documented that these liver enzymes have only modest sensitivity (Conigrave *et al.*, 1995). They also have limited specificity as they are affected by common medications and conditions such as non-alcoholic liver disease, hepatic congestion and biliary tract disease. CDT appears to be more specific than the liver enzymes, though patients with primary biliary cirrhosis, inborn errors of glycoprotein metabolism, the genetic D-variant of transferrin and other conditions can have higher CDT levels (Scouller *et al.*, 2000). The findings on sensitivity vary widely (Scouller *et al.*, 2000; Conigrave *et al.*, 2002).

Several studies have shown that use of a combination of tests such as GGT and CDT may increase diagnostic accuracy. Often the tests have been combined in a qualitative manner (i.e. whether one or more results are above the reference range). In addition several researchers have developed rules or algorithms for combining the quantitative results of multiple biochemical tests to increase diagnostic accuracy for chronic excessive alcohol consumption (Shaper *et al.*, 1985; Sillanaukee, 1992; Hartz *et al.*, 1997; Huseby *et al.*, 1997). Recently, Sillanaukee and Olsson (2001) provided an algorithm for combining CDT and GGT, and the resulting combination showed higher accuracy than either test alone. Accuracy was evaluated as the area under the receiver operating characteristic (ROC) curve. Based on discriminant analysis, the authors concluded that the combination of CDT and GGT provided a powerful tool in discriminating between problem drinkers and social drinkers.

That study raises several practical and methodological questions regarding the use of a combination of biochemical markers in the diagnosis of problem drinking:

Can the predictive function derived from the study (i.e. 0.8*lnGGT + 1.3*lnCDT) be usefully applied to other settings?

How consistent are the results obtained if logistic regression is used to derive the predictive function instead of discriminant analysis (Su *et al.*, 1993)?

Does the addition of information on non-alcohol related clinical factors such as obesity and smoking improve the accuracy of the predictive function?

To what extent does the addition of information on past history of alcohol dependence improve the accuracy of the predictive function?

*Author to whom correspondence should be addressed at: Drug Health Services, RPAH, Missenden Road, Camperdown, NSW 2050, Australia. Tel.: +61 2 9515 8650; Fax: +61 2 9515 8970; E-mail: katec@med.usyd.edu.au
**Current address: Simpson Centre for Health Services Research, University of NSW, Locked Bag 7103, Liverpool BC, NSW 1871, Australia.
***Other WHO/ISBRA study group members: B. Tabakoff (Project Coordinator), S. Borg, M. Dongier, H. Edenberg, C. J. P. Eriksson, M. L. O. S. Formigoni, B. F. Grant, P. L. Hoffman, K. Kiianmaa, T. Koyama, L. Legault, T-K. Li, M. G. Monteiro, M. Ogata, T. Saito, M. Salaspuro, S. Tufik.

Should different combination rules be used for population subgroups such as men and women?

Does a nonlinear combination of markers provide greater predictive accuracy than a linear combination?

These issues can best be addressed by a large scale study with sufficient sample size and population heterogeneity. The multicentre study of biological markers of alcohol use initiated by the World Health Organization (WHO) and the International Society for Biomedical Research on Alcoholism (ISBRA) provides such an opportunity (Helander and Tabakoff, 1997; World Health Organization, 1999; Conigrave *et al.*, 2002). This report uses the data from this study to address the above questions.

MATERIAL AND METHODS

Sample

Subjects were recruited from five countries: Australia (Sydney), Brazil (São Paolo), Canada (Montreal), Finland (Helsinki) and Japan (Sapporo) (Helander and Tabakoff, 1997). Recruitment targeted persons aged 18 years and over both from the community and from alcoholism treatment services. At each centre stratified recruitment was used, seeking equal subject numbers in three age groups (18–29, 30–44 and 45–60 years) in each of four drinking categories: non-drinkers, light/moderate drinkers (at least once per month and <210 g/week for men, <140 g/week for women), heavy drinkers (≥210 g/week for men, ≥140 g/week for women and no past treatment for an alcohol-related problem), and alcohol-dependent persons. Within each category, a fixed number of men and women were sampled, except in Australia and Finland where only men were studied. Subjects were recruited using convenience sampling rather than consecutive or randomized sampling. In most sites newspaper or other advertisements were used to recruit members of the public to non-drinker, light or heavy drinker categories. In at least one site (Australia) religious groups which commonly practice abstinence (e.g. Seventh Day Adventists) were also approached to obtain sufficient numbers of abstainers. Most alcohol-dependent patients were recruited from inpatient alcohol treatment services. In at least one centre (Canada) additional alcohol-dependent subjects were sought by newspaper advertisements. Overall, 1863 participants were recruited. For this report, analyses were restricted to the 1684 subjects (1125 men and 559 women) who did not report alcoholic or non-alcoholic liver disease.

Interview and blood sample

All patients were interviewed using the WHO/ISBRA schedule. This is closely based on the AUDADIS interview (Grant *et al.*, 1995) and takes 70–90 min to complete. The questionnaire includes demographic characteristics, amount of alcohol consumption in the past 30 days, symptoms of harmful alcohol use, abuse or dependence (ICD-10 and DSM criteria), affective changes associated with drinking, illicit drug use; medical and psychiatric history. Alcohol-dependent inpatients were interviewed within 72 h of entering the treatment facilities and where possible, were reinterviewed approximately 2 weeks later. Following the Universal Precautions Guidelines for biological samples, two 20-ml blood samples were obtained from each individual at the completion of the interview. Serum samples were then sent packed in dry ice to the Helsinki sample processing centre, from which aliquots were distributed to the relevant assay centre. A detailed description of the sampling and testing procedure has been presented elsewhere (World Health Organization, 1999). Questionnaires were dispatched to the data coordinating centre in Denver.

GGT and AST were assayed by reflectance spectrophotometry using a Vitros 250 Analyser (Ortho Clinical Diagnostics). Serum CDT determinations were carried out using the Pharmacia CDTect™ test. This test employs separation of transferrin isoforms on an anion exchange chromatography microcolumn followed by quantification by a double antibody radioimmunoassay. It provides test performance that correlates strongly with the BioRad %CDT method ($r = 0.84$) (Anton *et al.*, 2001).

Statistical methods

The distributions of the markers (CDT, GGT and AST) were examined and summary statistics computed on both the original and natural logarithmic (lnCDT, lnGGT and lnAST) scales. The logarithmic transformation was used because the marker distributions were positively skewed.

To allow direct comparison with the results of Sillanaukee and Olsson (2001), 'problem drinking' was defined as an average daily alcohol consumption of 60 g/day or more for both men and women. Lower levels of consumption were classified as 'social drinking'. Discriminant analysis was used to carry out stepwise selection of the three biochemical markers (lnCDT, lnGGT and lnAST) to determine the best algorithm for differentiating between problem drinkers and social drinkers/ abstainers. A split-sample strategy was adopted to examine the stability of the results. Half of the total sample for each recruitment site was chosen at random to form a calibration sub-sample for estimating the model parameters. This differed from the approach of Sillanaukee and Olsson (2001) who used two of their six study sites as the calibration sample, then validated the findings using data from the other four sites. The models were then applied to the whole sample. The analysis was repeated using logistic regression to determine if the results were comparable with those obtained by discriminant analysis.

Further analyses explored the contribution of clinical variables to the predictive function. For these we used logistic regression in preference to discriminant analysis. Logistic regression may be more appropriate in deriving predictive functions when the clinical variables are categorical (e.g. sex and study site) so that the multivariate normality assumptions of discriminant analysis are less likely to be met. Also, the logistic regression procedures available in many standard software packages can be used to assess the diagnostic performance of a prediction rule by estimating the ROC curve and the area under the curve (AUC). For simple models with single markers only, whether using untransformed data (CDT, GGT or AST) or logarithmic transformation, the ROC curve and AUC will be the same. However, for complex models, where markers were combined with clinical variables, logarithmic transformation produced statistically significant better discriminating power (i.e. significantly higher AUC) compared with algorithms using untransformed data. For the purposes of consistency, and to allow assessment of the incremental benefit of adding variables, log transformation is used for the markers throughout.

The relative performance of alternative prediction rules was assessed by comparing their AUC using *roccomp* and *rocgold* in Stata (Statacorp, 2001). These procedures implement a nonparametric test developed by DeLong *et al.* (1988) which takes into account the paired nature of the analysis (i.e. alternative prediction rules are compared within the same patient). Listwise deletion of missing values was used to ensure that model comparisons were based on the same set of subjects. After deletion of cases with missing values, models that included covariate information were based on 1552 subjects.

Interaction effects between gender, body mass index (BMI) and markers were examined. Because a significant interaction between gender and markers was found, separate analyses were then undertaken for men and women to derive predictive functions based on both clinical and laboratory data. Backward elimination of nonsignificant variables was used to select the 'best' predictive model. Nonlinear relationships of the markers to the 'diagnosis' of excessive consumption were also considered using cubic polynomial regression. Both statistical significance and the improvement in the AUC were used as criteria for including non-linear terms.

To assist clinical application of the results, a plot of the predicted post-test probability of problem drinking was created for the best marker or combination of markers separately for females and males. For both women and men, three post-test probability plot curves were produced: for the whole sample with only marker(s) included, and for higher risk and lower risk groups. These groups were derived from the complex models with clinical information included. As results indicated (see below) that both CDT and GGT contribute significantly to diagnostic accuracy in men, we re-estimated the complex model in men using $\ln(CDT^2 \times GGT)$ as a single independent variable. This was because the ratios of the regression coefficient of lnCDT to lnGGT were close to 2 in both Sillanaukee's and our analyses. The post-test probabilities were estimated using the formula:

$$P_{post} = \frac{1}{1 + \exp[-(const + \beta_1 x_1 + \ldots + \beta_n x_n)]}$$

where $x_i \ldots x_n$ represent the variables included in the logistic regression model and $\beta_i \ldots \beta_n$ are their estimated coefficients. All statistical analyses were carried out using Stata 7.0 and SPSS 6.1 for Windows.

RESULTS

Descriptive results

There were 1125 male and 559 female subjects, with mean ages of 37.1 (range 18–67) years and 36.7 (range 18–60) years, respectively. Of these, 343 men and 103 women reported mean daily alcohol intake ≥60 g in the previous 30 days. A further 624 men and 363 women reported either no alcohol use or consumption within commonly recommended limits (up to 30 g/day for men and up to 20 g/day for women). For those drinking within these limits, median and 95th-percentile GGT values were 27 and 83 U/l for men, 20 and 57 U/l for women; while for CDT the corresponding values were 13 and 23 U/l (men) and 17 and 30 U/l (women). The mean, median and interquartile range of the three markers are presented in Table 1 for social drinkers/abstainers and for problem drinkers as defined by Sillanaukee (2001). For each marker, both the mean values and standard deviations are higher in problem drinkers than in the group of social drinkers/abstainers.

Derivation and replication of combination rules

Based on the calibration sample, the predictive functions that included both lnCDT and lnGGT were $Y_1 = 0.90*$lnGGT + 1.68*lnCDT, and $Y_2 = 1.04*$lnGGT + 1.96*lnCDT, using discriminant analysis and logistic regression, respectively. These two functions, together with the function previously published by Sillanaukee and Olsson (2001) ($Y_3 = 0.8l*$lnGGT + 1.3*lnCDT), were then applied to the full data set to obtain the ROC curves. The resulting AUC were each 0.81 (0.811, 0.810 and 0.812 respectively). Comparisons between the ROC based on DeLong's method confirmed that the AUCs were not significantly different ($P = 0.31$ for the comparison

Table 1. The mean, standard deviation, median, inter-quartile range and range for each biochemical marker for problem drinkers and social drinkers/abstainers (control group), stratified by sex in a study on combining carbohydrate-deficient transferrin and gamma-glutamyltransferase to increase diagnostic accuracy for problem drinking

Marker	<60g ethanol/day		≥60g ethanol/day	
	Male (*n* = 748)	Female (*n* = 439)	Male (*n* = 328)	Female (*n* = 100)
CDT				
Mean (SD)	15.0 (7.5)	18.5 (6.8)	27.3 (18.4)	23.7 (10.5)
Median (IQR)	13 (11–16)	17 (14–21)	21 (15–33)	22 (17–28)
Minimum–Maximum	3–85	5–53	5–136	5–73
lnCDT: Mean (SD)	2.62 (0.4)	2.85 (0.4)	3.13 (0.6)	3.08 (0.4)
GGT				
Mean (SD)	44.8 (59.2)	30.9 (50.8)	110.7 (238.0)	86.1 (195.0)
Median (IQR)	30 (22–45)	21 (17–29)	50 (31–99)	37 (26–64)
Minimum–Maximum	7–947	7–699	13–3072	12–1730
lnGGT: Mean (SD)	3.52 (0.6)	3.14 (0.6)	4.11 (0.9)	3.83 (0.9)
AST				
Mean (SD)	33.0 (61.3)	25.7 (16.3)	51.7 (53.8)	43.7 (63.7)
Median (IQR)	27 (24–32)	23 (20–27)	33 (27–51)	25 (22–37)
Minimum–Maximum	11–1638	9–231	7–512	10–525
lnAST: Mean (SD)	3.36 (0.4)	3.18 (0.3)	3.7 (0.6)	3.44 (0.7)

n = 1615. AST, aspartate aminotrasferase; CDT, carbohydrate-deficient transferrin; GGT, γ-glutamyl transferase. SD, standard deviation; IQR, inter-quartile range.

between the two functions derived by discriminant analysis; $P = 0.07$ between the two functions derived by different methods using the current data).

AUC based on different models for the whole sample

The AUC was estimated for each individual marker giving 0.74 for CDT, 0.74 for GGT and 0.70 for AST. Table 2 shows the logistic regression results for each possible two marker combination (lnCDT + lnGGT, lnGGT + lnAST, lnCDT + lnAST) with and without adjustment for gender, age groups, body mass index (BMI), ethnicity, smoking status and past alcohol dependence.

For the whole sample, lnCDT + lnGGT was the best combination with an AUC of 0.81. The AUC for this combination was significantly larger than either of the alternative combinations (both $P < 0.001$). Including demographic and non-alcohol related clinical information in the model significantly improved the AUC to 0.86. Adding information on past history of alcohol dependence further increased the AUC to 0.90. Past history of alcohol dependence was a strong predictor of problem drinking and hence, the AUC of the three models that include past history are similar.

Does the diagnostic value of markers vary with sex or body mass index?

There were significant interaction effects between sex and markers ($P < 0.001$) but no significant interaction effects between BMI and markers. Stratifying by sex, lnCDT + lnGGT resulted in the highest AUC (0.82) for men. The AUC for lnGGT + lnAST and lnCDT + lnAST were 0.73 and 0.80, respectively. For women, the model with lnGGT as the only marker yielded an AUC of 0.78. This did not differ significantly ($P = 0.42$) from the AUC for lnCDT + lnGGT (0.80). The AUC for the other combinations, lnGGT + lnAST and lnCDT + lnAST, were 0.78 and 0.77, respectively. The above AUC are for biochemical markers alone without clinical information. Based on these results, all subsequent analyses are stratified by sex.

The contribution of clinical information to the AUC

Table 3 presents three different models for men based on lnCDT + lnGGT: (1) markers without any clinical information; (2) markers with non-alcohol clinical variables; and (3) the above plus past alcohol dependence. The site effect was not included in the model as the aim was to examine the

Table 2. The areas under the curve derived from marker combinations, with and without other clinical information (using the whole sample and based on a logistic regression model) in a study on combining carbohydrate-deficient transferrin and gamma-glutamyltransferase to increase diagnostic accuracy for problem drinking

Variable	lnCDT + lnGGT	lnGGT + lnAST	lnCDT + lnAST
Markers only	0.81	0.75	0.77
Markers plus non-alcohol clinical information[a]	0.86	0.83	0.84
Markers plus full clinical information[b]	0.90	0.89	0.89

[a]Markers with non-alcohol clinical variables (i.e. sex, age group, body mass index, ethnicity, smoking status). [b]The same plus history of alcohol dependence.

Table 3. Four logistic regression models and areas under the curve based on lnCDT + lnGGT for men ($n = 1026$) in a study on combining carbohydrate-deficient transferrin and gamma-glutamyltransferase to increase diagnostic accuracy for problem drinking

Parameter	No clinical information (model 1) Coefficient (st error)	Non-alcohol clinical variables included (model 2) Coefficient (st error)	Clinical variables including past alcohol dependence (model 3) Coefficient (st error)	Same as the model (3) but with marker combination[a] (model 4) Coefficient (st error)
Constant	−10.24 (0.70)***	−9.93 (0.75)***	−9.69 (0.79)***	−9.64 (0.79)**
lnCDT	2.21 (0.20)***	2.04 (0.22)***	1.77 (0.22)***	—
lnGGT	0.83 (0.11)***	0.82 (0.13)***	0.74 (0.13)***	—
$\ln(\text{GGT} \times \text{CDT}^2)$	—	—	—	0.83 (0.08)***
Age groups (ref: 18–29):				
30–54 years		−0.94 (0.20)***	−1.04 (0.22)***	−1.06 (0.21)***
55+ years		−0.66 (0.33)**	−0.62 (0.34)*	−0.64 (0.34)*
BMI (ref: normal):				
Underweight		−0.32 (0.40)	−0.28 (0.42)	−0.26 (0.42)
Overweight		0.19 (0.20)	0.30 (0.21)	0.27 (0.21)
Obese		0.83 (0.30)***	0.78 (0.33)**	0.69 (0.31)**
Ethnicity (ref: white):				
Black		−0.53 (0.49)	−0.94 (0.50)*	−0.98 (0.50)**
Asian-Indian		−0.95 (0.38)**	−0.35 (0.40)	−0.39 (0.40)
Other		0.45 (0.38)	0.33 (0.39)	0.31 (0.39)
Smoking status (ref: non-smoker):				
Ex-smoker		0.21 (0.28)	−0.22 (0.30)	−0.23 (0.30)
Current smoker		1.29 (0.21)***	0.68 (0.23)***	0.67 (0.23)***
Lifetime alcohol dependence (Yes)			1.96 (0.22)***	1.97 (0.22)***
AUC	0.82	0.86	0.89	0.89

*Significant at 10%; **significant at 5%, ***significant at 1%. [a]Instead of GGT and CDT being entered into the equation separately, the combination GGTxCDT² is entered. AUC, area under the curve; BMI, body mass index; CDT, carbohydrate-deficient transferrin; GGT, γ-glutamyl transferase.

contribution of clinical information given the pooled average site effect.

For men (Table 3), including significant clinical variables (ethnicity, age, BMI and smoking status) increases the AUC from 0.82 (lnCDT + lnGGT) to 0.86. Adding information on past alcohol dependence further increases the AUC to 0.89. Use of a combination test variable ($GGT \times CDT^2$) provides as good prediction (AUC 0.89) as inclusion of separate lnGGT and lnCDT variables. For women (Table 4), smoking status was the only non-alcohol-related clinical variable that was significant. Including smoking status increased the AUC from 0.78 (lnGGT alone) to 0.83. Adding information on past alcohol dependence gave an AUC of 0.91.

Figure 1 shows the ROC curves for CDT and GGT, singly, in combination, and with added clinical data for men and for women. From the curves we see that, while the combination of lnCDT and lnGGT offers a greater AUC for men than the individual markers, sensitivity was only moderate (between 48% and 76%) for specificity between 90% and 75%. It is of interest to note that, for women, the ROC for GGT and lnCDT + lnGGT cross but the AUC for these alternative 'tests' are equivalent. Looking at the points on the ROC, it appears that using CDT and GGT in combination may have better diagnostic performance than GGT alone at higher levels of specificity. This is an area worthy of further study. The figure also clearly demonstrates the increase in the AUC as covariate information is added.

Estimates of the post test probability of problem drinking

To estimate post-test probabilities, firstly, only the marker level was included for the overall model. Secondly, post-test probabilities were also computed for two contrasting groups of subjects using complex models that included clinical information, based on the coefficients in the right hand column of Tables 3 and 4. This was done separately for men and for women. For women, the highest risk group was

Table 4. Three logistic regression models and areas under the curve based on lnGGT for women ($n = 526$) in a study on combining carbohydrate-deficient transferrin and gamma-glutamyltransferase to increase diagnostic accuracy for problem drinking

Parameter	lnGGT only Coefficient (st error)	lnGGT + smoking status Coefficient (st error)	lnGGT + smoking + lifetime alcohol dependence. Coefficient (st error)
Constant	–5.59*** (0.06)	–6.83*** (0.72)	–7.54*** (0.85)
lnGGT	1.21 (0.17)***	1.21 (0.18)***	1.07 (0.20)***
Smoking status (ref: non-smoker)			
Ex-smoker		0.57 (0.49)	–0.83 (0.55)
Current smoker		1.96 (0.33)***	0.95 (0.39)**
Lifetime alcohol dependence (Yes)			3.03 (0.39)***
AUC	0.78	0.83	0.91

*Significant at 10%; **significant at 5%; ***significant at 1%. BMI, ethnicity and age were not found significant and so are excluded from the models.
AUC, area under the curve; BMI, body mass index; CDT, carbohydrate-deficient transferrin; GGT, γ-glutamyl transferase.
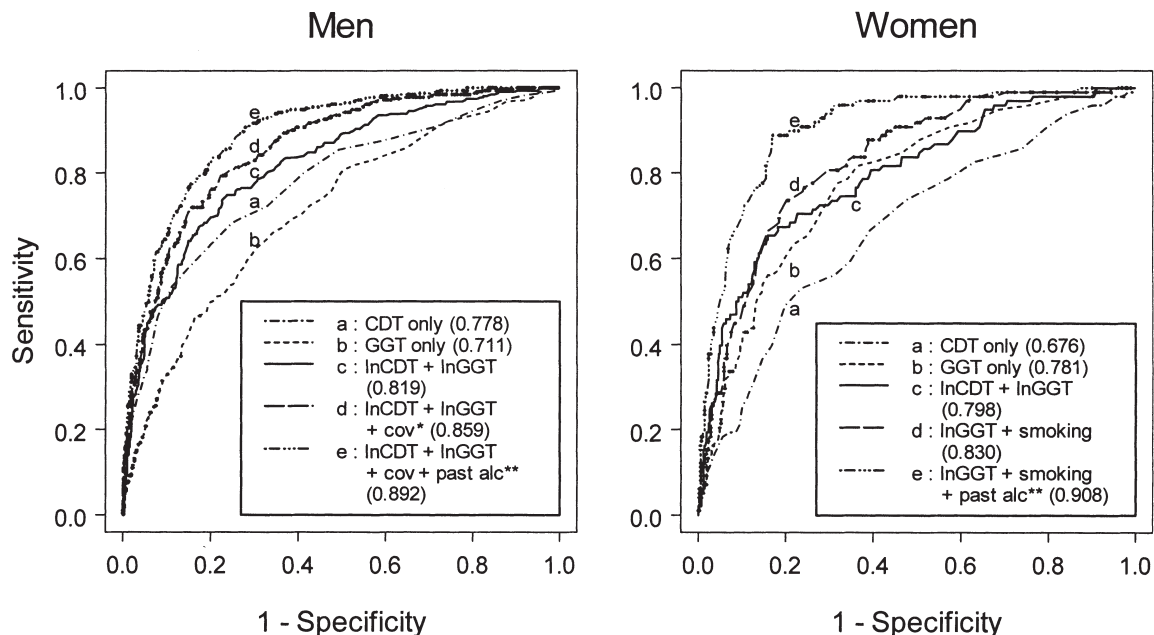

Fig. 1. Comparisons between areas under the curve (AUC) based on combinations of markers with and without clinical information in a study on combining carbohydrate-deficient transferrin and gamma-glutamyltransferase to increase diagnostic accuracy for problem drinking. (AUC are included in parentheses.) *cov refers to BMI, ethnicity, age and smoking status; **past alc refers to past DSM-IV alcohol dependence.

comprised of smokers with a history of alcohol dependence while the lowest risk group was comprised of ex-smokers without a history of alcohol dependence.

Figure 2a shows that for a female smoker with a history of alcohol dependence, with a pre-test probability of current problem drinking of 53%, a GGT value of 40 U/l corresponds to a 60% chance of being a problem drinker. For an ex-smoker with no history of alcohol dependence, with a pre-test probability of current problem drinking of 2%, the same GGT value corresponds to a 3% chance of being a problem drinker. As expected, more extreme values of GGT led to a greater difference between the pre- and post-test probabilities.

For men, there were no study participants in the model-based highest and lowest risk groups, so we chose other high and low risk groups which had appreciable numbers of participants. For the higher risk group, we chose white smokers aged 18–29 years, of normal weight, with a history of alcohol dependence ($n = 57$). For the lower risk group, we chose white non-smoking men aged 18–29 years, of normal weight, with no history of alcohol dependence ($n = 46$). These two groups had higher and lower pre-test probabilities (80.7 and 4.4%, respectively) than the overall model (30%). The post-test probability model was presented in Fig. 2b.

Figure 2b shows that for an 18–29-year-old normal-weight man with a GGT of 67 U/l and a CDT of 30 U/l (GGT*CDT^2 = 60 000) the chance of being a problem drinker is 90% for a smoker with a history of alcohol dependence and 37% for a non-smoker with no history of alcohol-dependence. Figure 2b also illustrates the variation in probability of problem drinking in both the high-risk and low-risk groups as values for the CDT- and GGT-based algorithm vary.

DISCUSSION

The results of our discriminant analysis support the findings of Sillanaukee and Olsson (2001) that using the levels of CDT and GGT in combination may provide a better indicator for the diagnosis of problem drinking (defined as consumption of ≥60 g ethanol per day), than either test alone. In this study, the quantitative results were combined, not just qualitative findings (i.e. whether or not the result was above the reference range). Our more detailed analysis showed that this gain in diagnostic performance is more pronounced in men. Overall, GGT alone provides an equally effective indicator of problem drinking in women compared to its use in combination with CDT. There may be benefit in adding CDT to GGT in women when specificity is above 75%. However, at this level of specificity, sensitivity of the combination of tests is low (<75%). A simple graphic method was produced to help clinicians estimate probabilities for consumption of 60 g or more per day, and for integrating clinical information with biochemical test results. However, the generalizability of this method needs to be studied further.

It should be noted that even when combining markers to maximize diagnostic accuracy, the blood tests still achieved no higher than 75% sensitivity for specificities of >75%. However, in situations where clinical history is unreliable or unobtainable, laboratory markers provide a valuable aid to diagnosis. The current analysis has shown that whatever clinical information is available (e.g. age, BMI, ethnicity and smoking status in men,

and smoking alone in women), combined with the test results can significantly improve diagnostic accuracy. In clinical practice, an estimate of the probability of problem drinking will usually precede the laboratory test results, and this generally helps determine the pre-test probability. If tests are used as a screening instrument without assessment by a clinician, then further information about the patient will improve upon the performance of tests alone. In our study, we produced several post-test probability plots which demonstrated how clinical information can contribute to the predicted probabilities given the same marker values, and marker values can change the clinically based pre-test probability estimates. The method presented here for computing the post-test probability of problem drinking provides a formula that can be adopted by clinicians. Alternatively, our graphs (Figs 2a,b) allow clinicians to estimate post-test probabilities given their own estimate of pretest probability for a particular patient and the marker results.

Although the regression coefficients estimated by discriminant analysis in this study differ from those given by Sillanaukee and Olsson (2001), and also differ from those estimated using logistic regression, the AUC associated with these three predictive functions using CDT + GGT were almost indistinguishable when applied to the current sample. This suggests stability for the combination rules derived from these two studies and derived using these two methods. The stability is explained by the fact that the ratio of the two regression coefficients for CDT and GGT was similar for the three models, and it is this ratio which determines the AUC of an ROC curve (Pepe and Thompson, 2000). The similarity of the ratio of the coefficients across studies indicates stability in the relative contribution of the two markers to the predictive function. This may, in part, be due to the similarity of the sampling strategy used in the two studies. Both used stratified recruitment targeting certain levels of alcohol consumption, rather than random or consecutive sampling.

It is important to note that the reference standard used in this study (self-reported alcohol consumption) is itself subject to error. This is the case for most studies on drinking. If different studies are based on the same reference standard (e.g. a set level of excessive alcohol consumption) but have different degrees of measurement error and differences in underlying population prevalence of excessive consumption, the regression coefficients and AUC from these studies could differ considerably (Irwig et al., 2002). The effect is likely to be an underestimation of the AUC (Walter et al., 1999), although the relative contribution of the markers may be similar. Each regression coefficient will be downward-biased, with the extent of this bias depending on the degree of the misclassification of the reference variable. This may partly explain why the absolute values of regression coefficients and AUC differed between the Sillanaukee study and this one, but the combination rules (understood as the ratio of the two coefficients) were remarkably similar.

Although Sillanaukee and Olsson (2001) considered including gender as a covariate in the discriminant function, the possible modification effect of gender when using markers in combination was not examined. Our results indicate that gender is indeed an effect modifier and the advantage of combining the markers was more evident among men than among women. For women, GGT alone provided the most efficient biochemical method overall for diagnosing problem

(a)
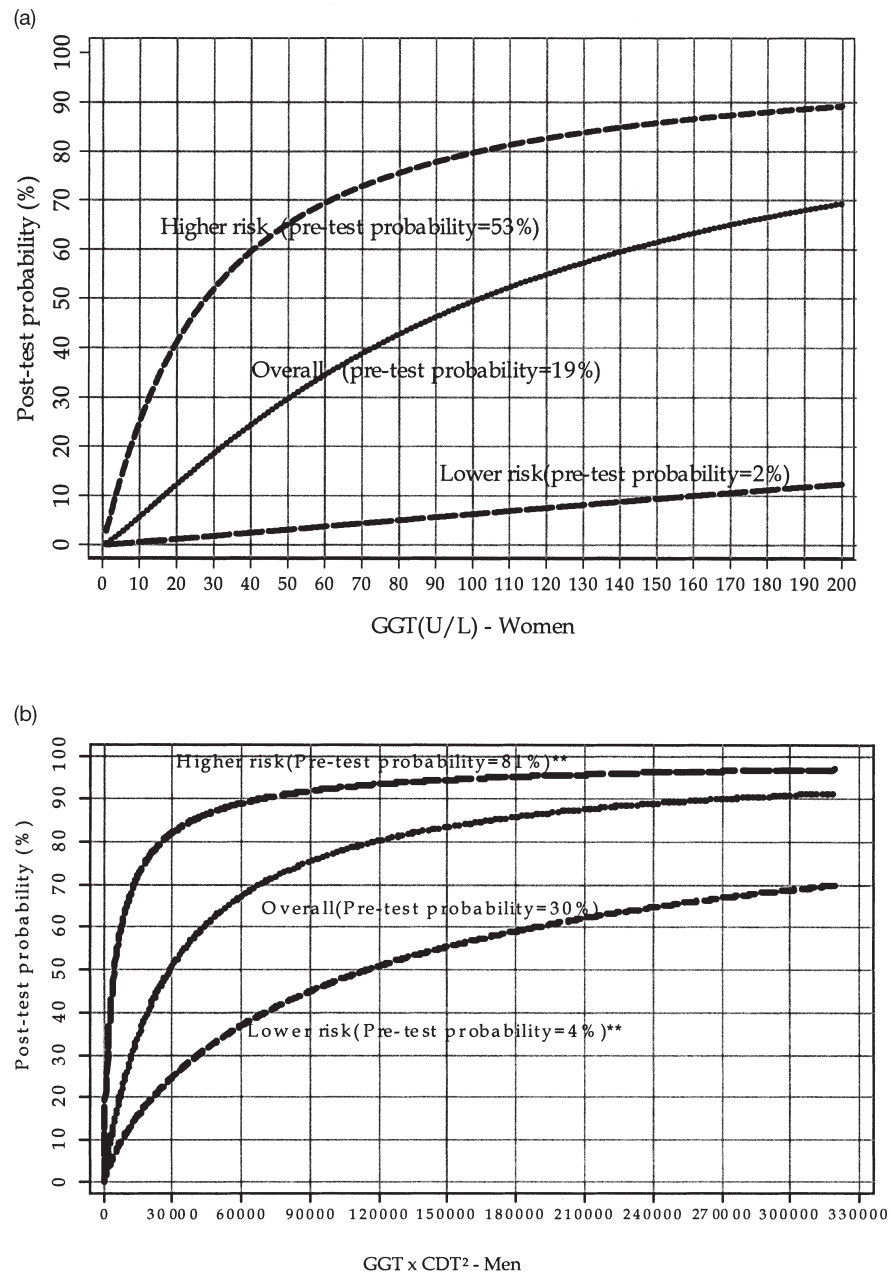

GGT(U/L) - Women

(b)

GGT x CDT² - Men

Fig. 2. Study on combining carbohydrate-deficient transferrin and gamma-glutamyltransferase to increase diagnostic accuracy for problem drinking. (a) The predicted post-test probability for the overall sample, and for higher and lower risk groups for women (higher risk group: smokers with alcohol dependence history; lower risk group: ex-smokers without dependence history). (b) The predicted post-test probability for the overall sample, higher and lower risk groups for men (higher risk: aged 18–29 years, normal weight, white, smoker with alcohol dependent history; lower risk: aged 18–29 years, normal weight, white, non-smoker with no alcohol dependent history).

drinking. This finding is important as it may avoid unnecessary cost and inconvenience in performing multiple tests in women. In contrast to the current results, Anton and Moak (1994), found that a result above the reference range for either CDT or GGT was more effective in detecting alcohol dependence in women than using either marker alone (sensitivity 72%, specificity 90% for the combined test). In that study 18 female alcohol dependent subjects were compared against 21 controls who were 'either minimal alcohol users or complete abstainers' (Anton and Moak, 1994). In the current study there

is a wider range of intermediate drinking patterns. The study of Huseby *et al.* (1997) on the combination of CDT and GGT results did not have sufficient numbers of female subjects to allow firm conclusions on gender differences. It is important to note that in the current study the area under the whole ROC curve was used as the criterion for selecting the best algorithm. Further research may be needed to ascertain whether the partial AUC (e.g. the area underneath the ROC curve when specificity is larger than a certain value, say, 0.8) would be a better criterion for deriving the algorithm (Pepe and Thompson, 2000).

Adding clinical data greatly increased the predictive accuracy of the final diagnostic algorithm. However, it is important to note that these results may have limited generalizability. The sampling frame used in this study targeted community based and alcohol treatment services patients. Subjects were volunteers, in many cases recruited by advertisement or convenience sampling, whereas the ideal would be patients in whom the issue of chronic problem drinking is a clinically relevant question. A stratified sampling frame was employed to ensure adequate numbers of subjects across age, sex and alcohol consumption categories. Hence, the incremental value of the various combinations of clinical information may not be applicable in a population based screening programme or other clinical settings (Irwig *et al.*, 2002). Similarly, one would expect the diagnostic accuracy of individual markers to vary across settings. However, the utility of combining markers is likely to be generalizable from one setting to another under the assumption that the correlation between markers remains constant in both the diseased and non-diseased groups.

The assay used for measurement of CDT (CDTect) has recently become unavailable. However the commercial kit which is currently widely used, Biorad %CDT, has been reported to have very similar performance ($r = 0.85$) to that of CDTect (Scouller *et al.*, 2000; Anton *et al.*, 2001). We do not have data on total transferrin and so cannot comment on effects of unusually high or low transferrin levels. As analyses were restricted to those without self-reported liver disease, and all subjects were apparently healthy (apart from alcohol withdrawal in the dependent patients), extreme values of transferrin are unlikely. Furthermore, we note that our findings of lesser benefit of the addition of CDT to GGT in women than in men, are consistent with findings of reduced accuracy of the %CDT methods in women (Helander *et al.*, 2001). Accordingly, the findings of the present study on the principle of combining values of GGT and CDT, and considering other patient characteristics are likely to be transferable to other assay methods.

We have provided results for all possible pairs of markers for the whole sample because the 'best' combination may depend on factors other than statistical significance. For example, cost, ease of implementation, and individual clinical factors may affect the clinical utility of each of these combinations. There were relatively small differences in diagnostic accuracy (as measured by AUC) among three different linear combinations of markers (0.81, 0.77 and 0.75 for lnCDT + lnGGT, lnCDT + lnAST, lnGGT + lnAST, respectively). Although the lnCDT + lnGGT combination was statistically the best combination, the increase in diagnostic accuracy was relatively small, and this choice of tests should be open to reconsideration if other clinical or cost factors outweigh this limited benefit. As an increasing amount of clinical data is entered into the model, the differences between the test combinations become smaller. For example, if past history of alcohol dependence is entered along with sex, age, BMI, ethnicity and smoking, then all three possible combinations of markers have an AUC of 0.89–0.90. It is important to recall that this study excluded persons with self reported liver disease. The use of biological markers, or their combinations, to detect drinking in patients with known liver disease requires further study.

In summary, we found that the value of combining lnCDT and lnGGT was particularly evident in men. A simple linear combination of the markers (in logarithmic values) worked well and the inclusion of clinical information significantly contributed to an accurate diagnosis. The combination rules were consistent across two studies and the three methods used. We present a simple graphical method which can be used by the clinician to integrate clinical and biochemical data to estimate the probability of drinking 60 g ethanol or more per day. This method warrants evaluation in other populations.

REFERENCES

Anton, R. F. and Moak, D. H. (1994) Carbohydrate-deficient transferrin and γ-glutamyltransferase as markers of heavy alcohol consumption: gender differences. *Alcohol: Clinical and Experimental Research* **18**, 747–754.

Anton, R. F., Dominick C., Bigelow, M., Westby, C. with the CDTect Research Group (2001) Comparison of Bio-Rad %CDT TIA and CDTect as laboratory markers of heavy alcohol use and their relationships with gamma-glutamyltransferase. *Clinical Chemistry* **47**, 1769–1775.

Conigrave, K. M., Saunders, J. B. and Whitfield, J. B. (1995) Diagnostic tests for alcohol consumption. *Alcohol and Alcoholism* **30**, 13–26.

Conigrave, K., Degenhardt, L., Whitfield, J. B., Saunders, J. B., Helander, A. and Tabakoff, B. *et al.* (2002) CDT, GGT and AST as markers of alcohol use: The WHO/ISBRA collaborative project. *Alcohol: Clinical and Experimental Research* **26**, 332–339.

DeLong, E., DeLong, D. and Clarke-Pearson, D. (1988) Comparing the areas under two or more correlated receiver operating characteristics curves: a non-parametric approach. *Biometrics* **44**, 837–845.

Grant, B. F., Harford, T. C., Dawson, D. A., Chou, P. S. and Pickering, R. P. (1995) The Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): reliability of alcohol and drug modules in a general population sample. *Drug and Alcohol Dependency* **39**, 37–44.

Hartz, A., Guse, C. and Kajdacsy-Balla, A. (1997) Identification of heavy drinkers using a combination of laboratory tests. *Journal of Clinical Epidemiology* **50**, 1357–1368.

Helander, A. and Tabakoff, B. (1997) Biochemical markers of alcohol use and abuse: experiences from the Pilot Study of the WHO/ISBRA Collaborative Project on state and trait markers of alcohol. International Society for Biomedical Research on Alcoholism. *Alcohol and Alcoholism* **32**, 133–144.

Helander, A., Fors, M. and Zakrisson, B. (2001) Study of Axis-Shield new %CDT immunoassay for quantification of carbohydrate-deficient transferrin (CDT) in serum. *Alcohol and Alcoholism* **36**, 406–412.

Huseby, N.-E., Nilssen, O. and Kanitz, R.-D. (1997) Evaluation of two biological markers combined as a parameter of alcohol dependency. *Alcohol and Alcoholism* **32**, 731–737.

Irwig, L., Bossuyt, P., Glasziou, P., Gatsonis, C. and Lijmer, J. (2002) Evidence base of clinical diagnosis: Designing studies to ensure that estimates of test accuracy are transferable. *British Medical Journal* **324**, 669–671.

Pepe, M. S. and Thompson, M. (2000) Combining diagnostic test results to increase accuracy. *Biostatistics* **1**, 123–140.

Scouller, K., Conigrave, K. M., Macaskill, P., Irwig, L. and Whitfield, J. B. (2000) Should we use carbohydrate deficient transferrin instead of gamma-glutamyltransferase for detecting problem drinkers? A systematic review and meta-analysis. *Clinical Chemistry* **46**, 1894–1902.

Shaper, A. G., Pocock, S. J., Ashby, D., Walker, M. and Whitehead, T. P. (1985) Biochemical and haematological response to alcohol intake. *Annals of Clinical Biochemistry* **22**, 50–61.

Sillanaukee, P. (1992) The diagnostic value of a discriminant score in the detection of alcohol abuse. *Archives of Pathology and Laboratory Medicine* **116**, 924–929.

Sillanaukee, P. and Olsson, U. (2001) Improved diagnostic classification of alcohol abusers by combining carbohydrate-deficient transferrin and gamma glutamyltranseferase. *Clinical Chemistry* **47**, 681–685.

Statacorp (2001) Stata Statistical Software. College Station, Stata Corporation, Texas.

Su, J. Q. and Liu, J. S. (1993) Linear combinations of multiple diagnostic markers. *Journal of the American Statistical Association* **88**, 1350–1355.

Walter, S. D., Irwig, L. and Glasziou, P. (1999) Meta-analysis of diagnostic tests with imperfect reference standards. *Journal of Clinical Epidemiology* **52**, 943–951.

World Health Organization (1999) WHO/ISBRA Study on biological state and trait markers of alcohol use and dependence: progress report. Social Change and Mental Health. Substance Abuse Department, World Health Organization, Geneva.