Alcohol and Alcoholism Vol. 49, No. 4, pp. 423–429, 2014
Advance Access Publication 15 April 2014
doi: 10.1093/alcalc/agu016

How Accurate Are Blood (or Breath) Tests for Identifying Self-Reported Heavy Drinking Among People with Alcohol Dependence?

Nicolas Bertholet[1,*], Michael R. Winter[2], Debbie M. Cheng[3,4], Jeffrey H. Samet[3,5] and Richard Saitz[2,6]

[1]Alcohol Treatment Center, Department of Community Medicine and Health, Lausanne University Hospital, Lausanne, Switzerland, [2]Data Coordinating Center, Boston University School of Public Health, Boston, MA, USA, [3]Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Boston University and Boston Medical Center, Boston, MA, USA, [4]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA, [5]Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, USA and [6]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

*Corresponding author: Alcohol Treatment Center, Department of Community Medicine and Health, Lausanne University Hospital, Beaumont 21b, P2, 02, 1011 Lausanne, Switzerland. E-mail: Nicolas.Bertholet@chuv.ch

(*Received* 28 *August* 2013; *first review notified* 18 *November* 2013; *in revised form* 6 *March* 2014; *accepted* 18 *March* 2014)

Abstract — Aims: Managing patients with alcohol dependence includes assessment for heavy drinking, typically by asking patients. Some recommend biomarkers to detect heavy drinking but evidence of accuracy is limited. **Methods:** Among people with dependence, we assessed the performance of disialo-carbohydrate-deficient transferrin (%dCDT, ≥1.7%), gamma-glutamyltransferase (GGT, ≥66 U/l), either %dCDT or GGT positive, and breath alcohol (> 0) for identifying 3 self-reported heavy drinking levels: any heavy drinking (≥4 drinks/day or >7 drinks/week for women, ≥5 drinks/day or >14 drinks/week for men), recurrent (≥5 drinks/day on ≥5 days) and persistent heavy drinking (≥5 drinks/day on ≥7 consecutive days). Subjects ($n = 402$) with dependence and current heavy drinking were referred to primary care and assessed 6 months later with biomarkers and validated self-reported calendar method assessment of past 30-day alcohol use. **Results:** The self-reported prevalence of any, recurrent and persistent heavy drinking was 54, 34 and 17%. Sensitivity of %dCDT for detecting any, recurrent and persistent self-reported heavy drinking was 41, 53 and 66%. Specificity was 96, 90 and 84%, respectively. %dCDT had higher sensitivity than GGT and breath test for each alcohol use level but was not adequately sensitive to detect heavy drinking (missing 34–59% of the cases). Either %dCDT or GGT positive improved sensitivity but not to satisfactory levels, and specificity decreased. Neither a breath test nor GGT was sufficiently sensitive (both tests missed 70–80% of cases). **Conclusions:** Although biomarkers may provide some useful information, their sensitivity is low the incremental value over self-report in clinical settings is questionable.

INTRODUCTION

Managing patients with chronic alcohol dependence includes assessment for heavy drinking and relapse, typically achieved by asking patients (McLellan *et al.*, 2000; McLellan, 2002; Saitz *et al.*, 2008). Biomarkers have been suggested as a way to detect heavy drinking and relapse (a return to heavy drinking after a period with no heavy use) (Neumann and Spies, 2003).

There are various biomarkers of alcohol use. Alcohol can be measured in the blood, breath or urine. Because of its short half-life of alcohol, the window of assessment for alcohol in these samples is brief, and therefore these tests may miss drinking that is not very recent. Nevertheless, these measures may be useful because they are not invasive and can provide 'point of service' rapid results. Other options include measures of tissue damage and/or physiological responses caused by alcohol. These indirect measures of alcohol use usually have longer windows of assessment but are likely to be influenced by factors other than alcohol use itself.

To be useful to clinicians in real life situations, any blood test or biomarker must have adequate accuracy, reflected by its sensitivity and specificity, and adequate positive and negative predictive values.

Carbohydrate-deficient transferrin (CDT) is considered the most specific biomarker used clinically for detecting heavy alcohol use and monitoring abstinence during treatment (Salaspuro, 1999; Anton, 2001; Arndt, 2001; Anton *et al.*, 2002; Golka and Wiese, 2004; Schwan *et al.*, 2004; Hannuksela *et al.*, 2007). Its performance appears to be less affected by liver diseases than are liver enzyme markers used to detect heavy alcohol consumption (Anttila *et al.*, 2003; Fleming *et al.*, 2004; Arndt *et al.*, 2006; Bortolotti *et al.*, 2006). There is also a dose–response relationship between alcohol use and CDT levels (Sillanaukee *et al.*, 2000; Schellenberg *et al.*, 2005). Burke and colleagues found that a 10% change in CDT had 70% sensitivity and 80% specificity for detecting changes of at least 2 drinks/day in men (Burke *et al.*, 1998). CDT has also been considered for detecting relapse or heavy drinking among people with alcohol dependence (Allen *et al.*, 2001). One study suggests that a rise in CDT level can herald a relapse self-reported by the patient, and therefore be of clinical interest to detect relapse (Mitchell *et al.*, 1997). Disialo-carbohydrate-deficient transferrin (measured as %dCDT) is the CDT isoform reported to be the most specific for heavy alcohol use.

Another commonly used test in clinical practice to detect heavy drinking is the serum gamma-glutamyl-transferase (GGT). GGT appears to be less sensitive and specific than CDT but, because it is relatively independent of CDT, combining the two tests may improve sensitivity (Conigrave *et al.*, 2002; Neumann and Spies, 2003; Hannuksela *et al.*, 2007).

Nevertheless, there is a lack of data on the operating characteristics of biomarkers outside of test validation studies, in which participants are recruited based on their alcohol consumption to establish the tests operating characteristics, or studies conducted among patients following specialized treatment programs. These studies provide crucial information on biomarkers operating characteristics but are less informative with respect to broader clinical samples such as patients with alcohol dependence seen in primary health care settings. Knowing the operating characteristics of biomarkers to detect any heavy use among people with dependence that are not in treatment *per se*, a situation encountered in general health care settings, would be useful to clinicians.

In the present study, we assessed the operating characteristics of these tests for identifying heavy drinking among people with dependence: disialo-carbohydrate-deficient transferrin

(%dCDT), gamma-glutamyltransferase (GGT), %dCDT and GGT together, and breath alcohol.

METHODS

We used cross-sectional data from 402 adults with alcohol dependence and current heavy drinking at study entry. This study is a secondary analysis of data collected for the Addiction Health Evaluation And Disease management (AHEAD) study (Saitz *et al.*, 2013). The AHEAD study aimed to evaluate the effectiveness of a chronic disease management program providing integrated health services for alcohol and other drug dependence in primary care. Participants were recruited between September 2006 and September 2008 primarily from a free-standing residential detoxification unit in Boston, MA, USA, as well as from self- and physician referrals from a large urban teaching hospital and through local advertisements. None of the participants entered the study through the acute internal medicine hospital, but two participants were referred to enroll in this outpatient study after a hospitalization. Participants were eligible if they were 18 years of age or older, fluent in English or Spanish, had alcohol or drug dependence, reported drinking ≥4 standard drinks for women and ≥5 standard drinks for men at least twice, or ≥15 drinks per week for women and ≥22 drinks per week for men in an average week in the past 30 days or recent illicit drug use, provided 2 contacts to assist with follow-up, had no plans to move from the local area within a year of screening and a score >21 on Mini-Mental State Examination (i.e. no serious cognitive impairment) (Folstein *et al.*, 1975; Smith *et al.*, 2006). People were excluded if they were not able to provide informed consent, were pregnant or had breath alcohol >100 mg/dl. The study was approved by the Institutional Review Board at Boston University Medical Campus (BUMC IRB). All participants were referred to primary care and received information to give to their primary care provider.

For the present study we used data from AHEAD participants who reported drinking ≥4 standard drinks for women or ≥5 standard drinks for men at least twice, or ≥15 drinks per week for women and ≥22 drinks per week for men in an average week in the past 30 days and had alcohol dependence [determined using the Composite International Diagnostic Interview Short Form (CIDI-SF)] at study entry. Other baseline instruments included: the Addiction Severity Index (ASI) (McLellan *et al.*, 1992), the Short From Health Survey (SF-12v2) (Ware *et al.*, 1996) and the Katz comorbidity questionnaire (Katz *et al.*, 1996).

Assessments of alcohol use by interview and biomarkers were performed at the study follow-up visits. Follow-up visits took place 3, 6 and 12 months after the baseline assessment but blood tests for biomarkers were only performed at the 6-month visit (or at the 12-month visit if missed at the 6-month visit). Past 30-day alcohol consumption was determined using the Timeline Followback, a validated structured calendar method (Sobell and Sobell, 1995), at the same visit as the biomarker testing. The Timeline Followback is a widely used and accepted reference standard, particularly when administered by trained personnel and with assurances of confidentiality. For substance use, the Timeline Followback has demonstrated good test-retest reliability (Sobell *et al.*, 1986, 2001; Fals-Stewart *et al.*, 2000; Carey *et al.*, 2004). Its agreement with other self-report measures of alcohol use is good (Grant *et al.*, 1995; Seale *et al.*, 2006; Johnson-Greene *et al.*, 2009). It has been shown to be correlated with reports by patient's collaterals (Sobell and Sobell, 1992). The present study includes only those participants with assessments of alcohol use by both interview and biomarker.

Biomarkers

CDT was measured as %dCDT using a cutoff point of ≥1.7% to define a positive test. The assay, based on high-pressure liquid chromatography with spectrophotometric detection, focuses solely on one isoform, the disialo-transferrin. We used cutoff values recommended by the Medical University of South Carolina, Institute of Psychiatry, Clinical Neurobiology Laboratory, Charleston, SC (where the samples were analyzed). GGT positive cutoff point was defined as: ≥66 U/l (the upper limit of normal in the laboratory where the samples were tested). We also used a combination of the two tests (either %dCDT or GGT or both positive). A positive breath alcohol test was >0. We also measured aspartate aminotransferase (AST) and alanine aminotransferase (ALT) and computed the AST/ALT ratio, a potential marker of alcohol intake (but one that has been questioned since it may also indicate advanced alcoholic liver disease rather than heavy drinking), and used two different cutoffs: AST/ALT>1.5, AST/ALT>2.0 (Salaspuro, 1987; Nyblom *et al.*, 2004).

Primary analyses

We assessed the operating characteristics of %dCDT (≥1.7), GGT (≥66 U/l), a combination of %dCDT-GGT (either or both positive), and breath alcohol (>0) for identifying the following three self-reported alcohol use levels over the past 30 days (determined using the Timeline Followback as the reference standard):

(1) any heavy drinking (≥4 drinks in a day or >7 drinks/week for women, ≥5 drinks in a day or >14 drinks/week for men),
(2) recurrent heavy drinking (≥5 drinks in a day on at least 5 days)
(3) persistent heavy drinking (≥5 drinks in a day on at least 7 consecutive days).

Sensitivity, specificity and corresponding exact 95% confidence intervals were calculated for each of the four biomarker tests to detect the three heavy alcohol use levels.

Receiver operating characteristic (ROC) curves were estimated to summarize the accuracy of %dCDT, GGT, as well as to evaluate the optimal cutoff points for distinguishing subjects with any, recurrent and persistent self-reported heavy drinking. Optimal cutoff points were determined by maximizing the combination of sensitivity and specificity. The area under the ROC curve (AUC) was used to quantify the overall accuracy of each test (1.0 represents a perfect test and 0.5 represents a test that cannot discriminate better than chance). Positive and negative predictive values (PPV, NPV), and likelihood ratio positive and negative (LR+, LR−) were also estimated along with 95% confidence intervals (calculated using published formulas) (Altman and Gardner, 1992).

Secondary analyses

Sensitivity and specificity of AST/ALT (with two cutoffs: >1.5 and >2.0) were estimated for comparison with %dCDT and GGT. ROC curves were estimated to summarize the accuracy of AST/ALT.

Because gender and BMI may affect the performance of biomarkers, the analyses were repeated stratified by gender and BMI (≤30 vs. >30).

All analyses were performed using SAS version 9.1.3 (Cary, NC, USA).

Table 1. Characteristics of study participants

Baseline sample characteristics ($n = 402$)	
Age, mean (SD)	40.0 (9.8)
Female, *n* (%)	100 (24.9%)
Homelessness (1+ night/past 3 months), *n* (%)	259 (64.4%)
Unemployed, *n* (%)	173 (43.0%)
Marital status, married, *n* (%)	25 (6.2%)
Education level (<High School, HS, >HS)	23.6%/48.8%/27.6%
Race ethnicity, *n* (%)	
Non-Hispanic white	160 (39.8%)
Non-Hispanic black	152 (37.8%)
Hispanic	59 (14.7%)
Non-Hispanic other	31 (7.7%)
Alcohol use, past 30 days	
Number of drinks per week, median (25th, 75th)	61.7 (25.9; 119.7)
Drinks/drinking day, median (25th, 75th)	14.6 (9.0; 23.4)
Maximum drinks in 1 day, median (25th, 75th)	24.0 (14.0; 39.0)
Percent days abstinent, median (25th, 75th)	25.0 (0.0; 66.7)
Any binge drinking, *n* (%)	400 (99.5%)
Current tobacco use, *n* (%)	360 (89.6%)
Heroin, cocaine or marijuana use, past 30 days, *n* (%)	344 (85.6%)
HCV infection[a]	159 (39.9%)
SF-12 Physical Component Summary, mean (SD)	41.5 (8.4)
SF-12 Mental Component Summary, mean (SD)	30.5 (9.8)
Addiction Severity Index—alcohol, mean (SD)	0.6 (0.3)
Addiction Severity Index—drugs, mean (SD)	0.3 (0.2)
Katz comorbidity scale, lifetime, mean (SD)	0.8 (1.1)
Alcohol use at follow-up ($n = 374$)[b]	
Number of drinks per week, median (25th, 75th)	2.5 (0.0; 25.2)
Drinks/drinking day, median (25th, 75th)	4.5 (0.0; 11.0)
Maximum drinks in one day, median (25th, 75th)	6.0 (0.0; 14.0)
Percent days abstinent, median (25th, 75th)	93.3 (56.7; 100.0)
Any binge drinking, *n* (%)	202 (54.0%)

SF-12: Short Form Health Survey v2.
[a]HCV status was based on antibody testing and/or viral load testing results. The majority of the sample (401/402, 99.8%) was antibody tested, and of those, 47.7% (84/176) who tested seropositive had results confirmed with HCV viral load testing.
[b]i.e. 6 or 12 month depending when the tests were available for each participant.

RESULTS

Based on the study aims (i.e. detecting heavy drinking among patients with alcohol dependence), we selected among AHEAD study participants ($n = 569$) those who had alcohol dependence and heavy drinking (past 30 days) at baseline ($n = 402$). Baseline characteristics of this study's analytic sample ($n = 374$) as well as alcohol use data at follow-up (i.e. 6 or 12 month depending when the tests were available for each participant) are presented in Table 1.

By self-report reference standard, the prevalence of any, recurrent and persistent heavy drinking at follow-up was 54.1 33.9, and 16.8%, respectively.

Primary analyses

The estimated sensitivity, specificity, positive predictive and negative predictive values of %dCDT (at cutoff point of ≥1.7%), GGT (for ≥66 U/l), combination test (either or both positive), and breath alcohol (>0) to identify any, recurrent, and persistent self-reported heavy drinking are presented in Table 2; %dCDT alone had higher sensitivity than GGT and breath alcohol test for each alcohol use level. Nevertheless, the test missed 34–59% of the cases. The combination of %dCDT and GGT appeared to improve sensitivity but specificity decreased.

Likelihood ratios positive and negative are presented in Table 3 (a likelihood ratio positive is the probability of a person who has the disease testing positive divided by the probability of a person who does not have the disease testing positive. A likelihood ratio negative is the probability of a person who has the disease testing negative divided by the probability of a person who does not have the disease testing

Table 2. Performance of biomarkers to detect three self-reported heavy alcohol use levels

Target condition and test	*N*	Sensitivity % (95% CI)[a]	Specificity	PPV	NPV
Any heavy drinking					
%dCDT	356	40.7 (33.7, 48.0)	95.7 (91.3, 98.3)	91.9 (83.9, 96.7)	57.4 (51.3, 63.4)
GGT	361	20.0 (14.6, 26.3)	86.1 (79.9, 91.0)	62.9 (49.7, 74.8)	47.8 (42.0, 53.7)
Either %dCDT+ or GGT+	358	52.6 (45.3, 59.8)	82.3 (75.6, 87.8)	77.9 (69.8, 84.7)	59.5 (52.8, 65.9)
Breath alcohol	375	19.9 (14.6, 26.1)	96.6 (92.7, 98.7)	87.0 (73.7, 95.1)	51.1 (45.5, 56.6)
Recurrent heavy drinking					
%dCDT	356	52.5 (43.2, 61.7)	90.3 (85.7, 93.7)	73.3 (62.6, 82.2)	78.9 (73.5, 83.6)
GGT	361	19.3 (12.7, 27.6)	83.9 (78.6, 88.3)	37.1 (25.2, 50.3)	67.9 (62.3, 73.2)
Either %dCDT+ or GGT+	358	60.0 (50.7, 68.8)	75.2 (69.2, 80.6)	55.0 (46.0, 63.7)	78.9 (73.0, 84.0)
Breath alcohol	375	26.0 (18.5, 34.7)	94.4 (90.9, 96.9)	69.6 (54.3, 82.3)	72.3 (67.2, 77.1)
Persistent heavy drinking					
%dCDT	356	66.1 (52.6, 77.9)	84.2 (79.5, 88.1)	45.4 (34.6, 56.5)	92.6 (88.8, 95.4)
GGT	361	29.3 (18.1, 42.7)	85.1 (80.6, 89.0)	27.4 (16.9, 40.2)	86.3 (81.9, 90.0)
Either %dCDT+ or GGT+	358	74.6 (61.6, 85.0)	70.9 (65.4, 76.0)	33.6 (25.6, 42.4)	93.4 (89.3, 96.3)
Breath alcohol	375	30.7 (19.6, 43.7)	91.4 (87.7, 94.2)	41.3 (27.0, 56.8)	86.9 (82.8, 90.4)

Any heavy drinking was defined as ≥4 drinks per occasion or >7 drinks per week for women, ≥5 drinks per occasion or >14 per week for men. Recurrent heavy drinking was defined as ≥5 drinks per day on at least 5 days over the past 30 days. Persistent heavy drinking was defined as ≥5 drinks per day on at least 7 consecutive days over the past 30 days.
[a]Exact 95% CI (confidence interval).

Table 3. Likelihood ratios positive and negative for three self-reported heavy alcohol use levels

Target condition and test	Likelihood ratio+ (95% CI)	Likelihood ratio– (95% CI)
Any heavy drinking		
%dCDT	9.42 (4.48, 19.84)	0.62 (0.55, 0.70)
GGT	1.44 (0.90, 2.31)	0.93 (0.85, 1.02)
%dCDT+ or GGT+	2.97 (2.08, 4.25)	0.58 (0.49, 0.68)
Breath alcohol	5.77 (2.51, 13.28)	0.83 (77.0, 89.4)
Recurrent heavy drinking		
%dCDT	5.39 (3.53, 8.23)	0.53 (0.43, 0.64)
GGT	1.20 (0.75, 1.91)	0.96 (0.87, 1.07)
%dCDT+ or GGT+	2.42 (1.86, 3.16)	0.53 (0.42, 0.67)
Breath alcohol	4.68 (2.60, 8.45)	0.78 (0.70, 0.87)
Persistent heavy drinking		
%dCDT	4.18 (3.03, 5.75)	0.40 (0.28, 0.58)
GGT	1.97 (1.22, 3.20)	0.83 (0.70, 0.99)
%dCDT+ or GGT+	2.56 (2.03, 3.23)	0.36 (0.23, 0.56)
Breath alcohol	3.55 (2.11, 5.97)	0.76 (0.64, 0.90)

Any heavy drinking was defined as ≥4 drinks per occasion or >7 drinks per week for women, ≥5 drinks per occasion or >14 per week for men; Recurrent heavy drinking was defined as ≥5 drinks per day on at least 5 days over the past 30 days; Persistent heavy drinking was defined as ≥5 drinks per day on at least 7 consecutive days over the past 30 days.
CI, confidence interval.

negative); %dCDT appeared to have the best likelihood ratio positive for each alcohol use level, followed next by breath alcohol. A likelihood ratio positive >5 has been used to indicate a moderate to large increase in the odds of the disease if the test is positive and, similarly, a likelihood ratio negative <0.2 suggests a moderate to large decrease in the odds of the disease if the test is negative (Jaeschke *et al.*, 1994). Two tests, %dCDT (9.47) and breath alcohol (5.85), had estimated likelihood ratios positive >5 for detecting any heavy drinking, although the lower bounds of the 95% confidence intervals were below the threshold. For the detection of recurrent heavy drinking, only %dCDT had an estimated likelihood ratio positive >5 (5.41). No test had a likelihood ratio positive >5 for persistent heavy drinking. Across the three different levels of heavy drinking, no test had a likelihood ratio negative <0.2.

We assessed whether or not false positives for %dCDT, GGT and AST/ALT were corroborated by the other tests. Among the false positive biomarker cases, most (41/50, 82%) were not corroborated by the other tests: among the false positive CDT cases ($n = 7$), 1 case was corroborated by a positive GGT test; among the false positive GGT cases ($n = 23$), 1 case was corroborated by a positive CDT test; 2 cases were corroborated by a positive breath alcohol content test; among the false positive AST/ALT >1.5 cases ($n = 17$), 1 case was corroborated by a positive CDT test, 1 case was corroborated by a positive breath alcohol content test, and 2 cases were corroborated by a positive GGT test; among the false positive AST/ALT >2.0 cases ($n = 3$), 1 case was corroborated by a positive GGT test.

Optimal cutoff points and test accuracy for %dCDT and GGT

The ROC curves for %dCDT gave the following optimal cutoff points for any, recurrent and persistent self-reported heavy drinking, respectively: 1.5% (sensitivity: 50.5%; specificity: 90.1%); 1.3% (sensitivity: 75.8%; specificity: 69.1%); 1.4% (sensitivity: 81.4%; specificity: 69.7%). The estimated area under the curve (AUC) suggests fair to good accuracy of the test.

Table 4. Optimal cutoff points with sensitivity and specificity, and area under the curve (AUC) for %dCDT, GGT and AST/ALT for three self-reported heavy alcohol use levels

	Optimal cutoff point	Sensitivity for optimal cutoff point (%)	Specificity for optimal cutoff point (%)	AUC
Any heavy drinking				
%dCDT	1.5%	50.5	90.1	0.77
GGT	24 IU/l	71.8	48.8	0.61
AST/ALT	1.08	56.1	59.0	0.59
Recurrent heavy drinking				
%dCDT	1.3%	75.8	69.1	0.80
GGT	27 IU/l	75.6	54.1	0.64
AST/ALT	1.08	60.2	56.3	0.59
Persistent heavy drinking				
%dCDT	1.4%	81.4	69.7	0.81
GGT	40 IU/l	55.2	70.3	0.66
AST/ALT	1.21	54.1	67.2	0.61

Any heavy drinking was defined as ≥4 drinks per occasion or >7 drinks per week for women, ≥5 drinks per occasion or >14 per week for men; Recurrent heavy drinking was defined as ≥5 drinks per day on at least 5 days over the past 30 days; Persistent heavy drinking was defined as ≥5 drinks per day on at least 7 consecutive days over the past 30 days.

Optimal cutoff points for GGT were: 24 IU/l (sensitivity: 71.8%; specificity: 48.8%); 27 IU/l (sensitivity: 75.6%; specificity: 54.1%); 40 IU/l (sensitivity: 55.2%; specificity 70.3%). The estimated area under the curve (AUC) suggests poor accuracy of the test.

Estimated area under the curve, optimal cutoff points and corresponding sensitivity and specificity for %dCDT and GGT are presented in Table 4.

Secondary analyses

AST/ALT

Sensitivity (95% confidence interval, CI) for AST/ALT >1.5 was 14.8% (10.1%; 20.6%), 16.3% (10.2%; 24.0%) and 21.3% (11.9%; 33.7%) to detect any, recurrent and persistent self-reported heavy drinking. Corresponding specificity (95% CI) was 89.8% (84.1%; 93.9%), 89.2% (84.5%; 92.8%) and 89.1% (85.0%; 92.4%), respectively.

Sensitivity (95% CI) for AST/ALT >2 was 3.1% (1.1%; 6.5%), 2.4% (0.5%; 7.0%) and 4.9% (1.0%; 13.7%) to detect any, recurrent and persistent self-reported heavy drinking. Corresponding specificity was 98.2% (94.8%; 99.6%), 97.5% (94.6%; 99.1%) and 98.0% (95.7%; 99.3%).

The ROC curves for the AST/ALT ratio gave the following optimal cutoff points for any, recurrent and persistent self-reported heavy drinking, respectively: 1.08 (sensitivity: 56.1%; specificity: 59.0%); 1.08 (sensitivity: 60.2%; specificity: 56.3%); 1.21 (sensitivity: 54.1%; specificity: 67.2%). The estimated area under the curve (AUC) suggests unacceptable accuracy of the test.

Estimated area under the curve, optimal cutoff points and corresponding sensitivity and specificity for the AST/ALT ratio are presented in Table 4.

Stratified analyses by gender and BMI

Stratified analyses by gender and BMI are presented in Table 5. In general, sensitivity appeared slightly better for men compared with women. Larger differences were observed

Table 5. Performance of biomarkers to detect three self-reported heavy alcohol use levels, stratified by gender and BMI

Gender		Male			Female	
Target condition and test	*N*	Sensitivity % (95% CI)	Specificity % (95% CI)	*N*	Sensitivity % (95% CI)	Specificity % (95% CI)
Any heavy drinking						
%dCDT	274	42.4 (34.6, 50.5)	94.8 (89.1, 98.1)	82	33.3 (18.6, 51.0)	97.8 (88.5, 99.9)
GGT	277	23.3 (16.9, 30.6)	82.2 (74.1, 88.6)	84	5.6 (0.7, 18.7)	95.8 (85.8, 99.5)
%dCDT+ or GGT+	276	56.3 (48.2, 64.2)	78.0 (69.4, 85.1)	82	36.1 (20.8, 53.8)	93.5 (82.1, 98.6)
Breath alcohol	280	20.9 (14.9, 27.9)	95.7 (90.3, 98.6)	95	15.8 (6.0, 31.3)	98.3 (90.6, 100.0)
Recurrent heavy drinking						
%dCDT	274	53.3 (43.3, 63.1)	89.9 (84.4, 94.0)	82	46.7 (21.3, 73.4)	91.0 (81.5, 96.6)
GGT	277	20.2 (13.0, 29.2)	78.6 (71.8, 84.5)	84	13.3 (1.7, 40.5)	97.1 (89.9, 99.7)
%dCDT+ or GGT+	276	60.9 (50.9, 70.3)	70.2 (62.7, 76.9)	82	53.3 (26.6, 78.7)	88.1 (77.8, 94.7)
Breath alcohol	280	26.4 (18.3, 35.9)	93.7 (89.0, 96.8)	95	23.5 (6.8, 49.9)	96.2 (89.2, 99.2)
Persistent heavy drinking						
%dCDT	274	66.7 (52.1, 79.2)	82.5 (76.9, 87.3)	82	62.5 (24.5, 91.5)	89.2 (79.8, 95.2)
GGT	277	30.0 (17.9, 44.6)	81.1 (75.3, 85.9)	84	25.0 (3.2, 65.1)	97.4 (90.8, 99.7)
%dCDT+ or GGT+	276	74.5 (60.4, 85.7)	65.8 (59.2, 71.9)	82	75.0 (34.9, 96.8)	86.5 (76.6, 93.3)
Breath alcohol	280	29.6 (18.0, 43.6)	89.8 (85.1, 93.4)	95	37.5 (8.5, 75.5)	95.4 (88.6, 98.7)
BMI		BMI ≤30			BMI >30	
Target condition and test		Sensitivity % (95%CI)	Specificity % (95%CI)		Sensitivity % (95%CI)	Specificity % (95%CI)
Any heavy drinking						
%dCDT	255	46.7 (38.2, 55.4)	94.9 (89.3, 98.1)	95	27.3 (16.1, 41.0)	97.5 (86.8, 99.9)
GGT	258	19.9 (13.5, 27.6)	83.6 (75.8, 89.7)	97	19.3 (10.1, 31.9)	97.5 (86.8, 99.9)
%dCDT+ or GGT+	257	56.9 (48.2, 65.4)	79.2 (70.8, 86.0)	95	41.8 (28.7, 55.9)	95.0 (83.1, 99.4)
Breath alcohol	265	21.7 (15.2, 29.6)	98.4 (94.4, 99.8)	103	14.8 (7.0, 26.2)	95.2 (83.8, 99.4)
Recurrent heavy drinking						
%dCDT	255	61.5 (50.1, 71.9)	88.9 (83.3, 93.2)	95	32.4 (18.0, 50.0)	93.1 (83.3, 98.1)
GGT	258	19.5 (11.6, 29.7)	82.4 (75.9, 87.7)	97	18.9 (8.0, 35.2)	91.7 (81.6, 97.2)
%dCDT+ or GGT+	257	67.5 (56.3, 77.4)	73.0 (65.8, 79.4)	95	43.2 (27.1, 60.5)	84.5 (72.6, 92.7)
Breath alcohol	265	30.5 (20.8, 41.6)	96.2 (92.3, 98.5)	103	17.1 (7.2, 32.1)	93.6 (84.3, 98.2)
Persistent heavy drinking						
%dCDT	255	70.7 (54.5, 83.9)	80.8 (74.9, 85.9)	95	55.6 (30.8, 78.5)	92.2 (83.8, 97.1)
GGT	258	30.0 (16.6, 46.5)	83.9 (78.4, 88.6)	97	27.8 (9.7, 53.5)	91.1 (82.6, 96.4)
%dCDT+ or GGT+	257	78.1 (62.4, 89.4)	67.1 (60.4, 73.4)	95	66.7 (41.0, 86.7)	83.1 (72.9, 90.7)
Breath alcohol	265	39.0 (24.2, 55.5)	92.9 (88.7, 95.9)	103	14.3 (3.1, 36.3)	90.2 (81.7, 95.7)

Any heavy drinking was defined as ≥4 drinks per occasion or >7 drinks per week for women, ≥5 drinks per occasion or >14 per week for men; Recurrent heavy drinking was defined as ≥5 drinks per day on at least 5 days over the past 30 days; Persistent heavy drinking was defined as ≥5 drinks per day on at least 7 consecutive days over the past 30 days.

by BMI status: %dCDT and breath alcohol appeared to be more sensitive among individuals with BMI ≤ 30 compared with those with BMI > 30. Tests were generally more specific for women and for people with BMI > 30.

DISCUSSION

We aimed at investigating the accuracy and operating characteristics of blood and breath tests (i.e. biomarkers) for detecting heavy drinking in individuals with alcohol dependence who were followed for 6 to 12 months after being enrolled with current heavy drinking. We used self-reported validated calendar measures of alcohol use from the Timeline Followback as the reference standard. Among individuals with alcohol dependence, %dCDT yielded the best performance. The estimated likelihood ratio positive for %dCDT suggests that a positive test may be associated with a moderate to large increase in the odds of any heavy drinking and recurrent heavy drinking in this population. Nevertheless, at the recommended cutoff point for %dCDT, almost 60% of individuals reporting heavy drinking amounts were not identified. The sensitivity of the tests were marginally better when %dCDT and GGT results were combined: when %dCDT was combined with GGT, 25% of individuals reporting drinking five or more drinks on at least 7 consecutive days in the past month were not identified. Specificity appeared to decrease for the combined test compared with either test alone. Therefore, while a positive test might be useful to detect or herald relapse (Mitchell *et al.*, 1997), the %dCDT test will miss most cases, making the test of little use in clinical practice.

The breath alcohol test and GGT missed 70–80% of the cases. Both GGT and breath alcohol test detect approximately only one in five cases of heavy drinking. As such, both tests were very insensitive for detecting any heavy drinking reported by the TLFB. Breath alcohol is generally assumed to be poor for detecting drinking over time (since it only detects recent drinking). GGT is thought to be useful for detecting heavy drinking even when drinking has not occurred in recent hours. Combining these two tests would not be much more clinically useful (since at best, assuming that they never detect the same patients, they would detect two in five cases).

The specificity of the studied biomarkers was adequate, but, given the prevalence of drinking in the population, negative predictive values indicate the biomarkers were clinically useful only for the highest drinking category. In addition, the

challenge for clinicians is to detect individuals who are drinking heavily and therefore sensitivity is more critical, which is different from other testing situations where specificity is key (i.e. forensic testing).

Optimal cutoffs for %dCDT were roughly similar to the recommended cutoffs by the testing laboratory. For GGT, optimal cutoffs were very different from cutoffs used clinically. The optimal cutoff points (24, 27 and 40 U/l for any, recurrent and persistent self-reported heavy drinking) are below what is considered abnormal in a clinical setting (66 U/l) and general accuracy of the test was limited, making it inadvisable for clinicians to choose to use GGT by itself to detect heavy drinking.

In general, in ROC analyses, areas under the curve were *fair* to *good* for %dCDT but not for other tests. For example the performance of AST/ALT was close to what we would expect if the test performed no better than chance.

Observed sensitivity and specificity for %dCDT were in the mid-upper range of those reported in previous studies conducted in various populations and with various methods of testing (Koch *et al.*, 2004). In their systematic review, Koch *et al.* reported on two CDT assays (CDTect and CDTTriTIA) with sensitivity ranging from 10 to 85% and specificity from 77 to 100%. As shown in other studies, combining CDT and GGT increased sensitivity (Sillanaukee and Olsson, 2001). The simple combination of the two tests (either or both test positive) allows sensitivities in the 50–75% range (depending on the targeted drinking level). Nevertheless, the combination of the two tests, in terms of accuracy, does not appear sufficiently valuable.

The present study has limitations. The reference standard used in the present study should not be considered equivalent to a short clinical assessment of alcohol use or to self-report in response to unstructured questions The TLFB is an extensive procedure that is unlikely to be implemented in clinical practice. Therefore our findings do not necessarily inform how laboratory tests compare to the usual shorter and likely less reliable questions used in clinical practice to identify heavy drinking. Nevertheless, the TLFB has shown good agreement with shorter self-report instruments of alcohol use (Seale *et al.*, 2006; Johnson-Greene *et al.*, 2009). In the present study, the TLFB was used as a reference standard to understand the potential clinical utility of laboratory tests.

Also, a strong effort was made to insure confidentiality. This is appropriate when studying sensitivity and other operating characteristics, but may differ from what would occur in clinical practice. Therefore our results should not be interpreted as a comparison between biomarkers and self-report as it is likely to occur in day-to-day clinical practice. In the present study, we have a real world clinical population in terms of their risk for heavy drinking, but with the advantage of the research situation with respect to measuring the reference standard for heavy drinking (validated calendar method, assurances of confidentiality). Therefore, the research situation maximizes the likelihood of accurate reporting of heavy drinking. This is a strength of the study, since it removes one of the main justifications for testing (i.e. the fact that patient might not report accurately their drinking) and gives information on how reliable biomarkers are when accuracy of self-report is maximized. In some cases, biomarkers were positive when the self-report was not. We do not have enough information to determine whether or not these tests were false elevations due to something else or whether they allowed identifying heavy drinking in subjects who did not self-report it. Numbers were small and in most cases, one positive biomarker result was not corroborated by the other tests. Despite the limitation in being able to interpret reasons for false positives and false negatives, the main focus and findings of this paper remain highly clinically relevant—in informing the question of how well can laboratory tests can detect reference standard self-reported heavy drinking.

Another limitation is that we did not test all possible biomarkers, such as phosphatidylethanol (PeTH) and ethylglucuronide (EtG), tests that are not in routine clinical use (Helander *et al.*, 2012; Kharbouche *et al.*, 2012). These biomarkers are promising but appropriate cutoffs are still being determined. There is hope that they may be better in detecting heavy drinking, though further research is needed—the biggest challenge being whether they can distinguish heavy drinking from any drinking. Despite limitations, we think our study provides important insights for clinicians about widely available tests: from a clinical point of view, even though %dCDT alone is not very sensitive, it can perform adequately. When facing a positive %dCDT test, positive predictive values were acceptable to good for the two lowest levels of drinking. Due to the lower prevalence of the highest drinking category, positive predictive value was 45%. Therefore, %dCDT may have better predictive value for detecting heavier levels of drinking than the ones reported here but the positive predictive value will be greatly affected due to the very limited prevalence of heaviest drinking levels, even in a population that consisted of individuals with quite severe alcohol dependence. Furthermore, most would agree that it is important to be able to detect all of the three heavy drinking levels we defined in this study when managing patients with alcohol dependence.

In conclusion, biomarkers have been recommended for assessing return to heavy drinking in people with alcohol dependence in treatment. However, the tests do not appear to perform sufficiently well enough as stand-alone tests. In our data, even if we speculate that all the participants who had a positive %dCDT test and reported no heavy use were not reporting their alcohol use (i.e. making the assumption that the test has NO false positives), only 7/356 additional subjects would have been picked up by the test. Therefore, our findings do not support their use at this time.

A carefully taken history of alcohol use is likely to provide more useful information than lab testing, even though, as stated in the limitations paragraph, clinicians are not expected to use a TLFB procedure with their patients. Also the question remains as to how to intervene clinically among those reporting no heavy use whose tests are positive. Challenging patient reports is delicate and can negatively impact the doctor–patient relationship, especially since false positive blood tests are possible and even likely. As a result we think that biomarkers have potential on occasion to bring useful additional information to clinicians in the detection of heavy drinking among patients with dependence, but this is likely to be the exception rather than the rule. Such use should occur with discussion with the patient in the context of a supportive doctor–patient relationship, where discrepancies between self-report and lab-tests results can be discussed. With regard to research settings, the biomarkers we have studied do not appear to have sufficient diagnostic accuracy for identifying heavy alcohol use to be used as stand-alone tests.

Funding — The AHEAD study was funded by grant R01 AA010870 from the National Institute on Alcohol Abuse and Alcoholism and grant R01 DA010019 from the National Institute on Drug Abuse.

Conflict of interest statement. None declared.

REFERENCES

Allen JP, Litten RZ, Fertig JB *et al*. (2001) Carbohydrate-deficient transferrin: an aid to early recognition of alcohol relapse. *Am J Addict* **10**(Suppl):24–8.

Altman DG, Gardner MJ. (1992) Confidence intervals for research findings. *Br J Obstet Gynaecol* **99**:90–1.

Anton RF. (2001) Carbohydrate-deficient transferrin for detection and monitoring of sustained heavy drinking. What have we learned? Where do we go from here? *Alcohol* **25**:185–8.

Anton RF, Lieber C, Tabakoff B. (2002) Carbohydrate-deficient transferrin and gamma-glutamyltransferase for the detection and monitoring of alcohol use: results from a multisite study. *Alcohol Clin Exp Res* **26**:1215–22.

Anttila P, Jarvi K, Latvala J *et al*. (2003) A new modified gamma-% CDT method improves the detection of problem drinking: studies in alcoholics with or without liver disease. *Clin Chim Acta* **338**:45–51.

Arndt T. (2001) Carbohydrate-deficient transferrin as a marker of chronic alcohol abuse: a critical review of preanalysis, analysis, and interpretation. *Clin Chem* **47**:13–27.

Arndt T, Meier U, Nauck M *et al*. (2006) Primary biliary cirrhosis is not a clinical condition for increased carbohydrate-deficient transferrin: experience with four independent CDT analysis methods. *Clin Chim Acta* **372**:184–7.

Bortolotti F, De Paoli G, Tagliaro F. (2006) Carbohydrate-deficient transferrin (CDT) as a marker of alcohol abuse: a critical review of the literature 2001–2005. *J Chromatogr B Analyt Technol Biomed Life Sci* **841**:96–109.

Burke V, Puddey IB, Rakic V *et al*. (1998) Carbohydrate-deficient transferrin as a marker of change in alcohol intake in men drinking 20 to 60 g of alcohol per day. *Alcohol Clin Exp Res* **22**:1973–80.

Carey KB, Carey MP, Maisto SA *et al*. (2004) Temporal stability of the timeline followback interview for alcohol and drug use with psychiatric outpatients. *J Stud Alcohol* **65**:774–81.

Conigrave KM, Degenhardt LJ, Whitfield JB *et al*. (2002) CDT, GGT, and AST as markers of alcohol use: the WHO/ISBRA collaborative project. *Alcohol Clin Exp Res* **26**:332–9.

Fals-Stewart W, O'Farrell TJ, Freitas TT *et al*. (2000) The timeline followback reports of psychoactive substance use by drug-abusing patients: psychometric properties. *J Consult Clin Psychol* **68**:134–44.

Fleming MF, Anton RF, Spies CD. (2004) A review of genetic, biological, pharmacological, and clinical factors that affect carbohydrate-deficient transferrin levels. *Alcohol Clin Exp Res* **28**:1347–55.

Folstein MF, Folstein SE, McHugh PR. (1975) 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**:189–98.

Golka K, Wiese A. (2004) Carbohydrate-deficient transferrin (CDT) —a biomarker for long-term alcohol consumption. *J Toxicol Environ Health B Crit Rev* **7**:319–37.

Grant KA, Tonigan JS, Miller WR. (1995) Comparison of three alcohol consumption measures: a concurrent validity study. *J Stud Alcohol* **56**:168–72.

Hannuksela ML, Liisanantti MK, Nissinen AE *et al*. (2007) Biochemical markers of alcoholism. *Clin Chem Lab Med* **45**:953–61.

Helander A, Peter O, Zheng Y. (2012) Monitoring of the alcohol biomarkers PEth, CDT and EtG/EtS in an outpatient treatment setting. *Alcohol Alcohol* **47**:552–7.

Jaeschke R, Guyatt GH, Sackett DL. (1994) Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* **271**:703–7.

Johnson-Greene D, McCaul ME, Roger P. (2009) Screening for hazardous drinking using the Michigan Alcohol Screening Test-Geriatric Version (MAST-G) in elderly persons with acute cerebrovascular accidents. *Alcohol Clin Exp Res* **33**:1555–61.

Katz JN, Chang LC, Sangha O *et al*. (1996) Can comorbidity be measured by questionnaire rather than medical record review? *Med Care* **34**:73–84.

Kharbouche H, Faouzi M, Sanchez N *et al*. (2012) Diagnostic performance of ethyl glucuronide in hair for the investigation of alcohol drinking behavior: a comparison with traditional biomarkers. *Int J Legal Med* **126**:243–50.

Koch H, Meerkerk GJ, Zaat JO *et al*. (2004) Accuracy of carbohydrate-deficient transferrin in the detection of excessive alcohol consumption: a systematic review. *Alcohol Alcohol* **39**:75–85.

McLellan AT. (2002) Have we evaluated addiction treatment correctly? Implications from a chronic care perspective. *Addiction* **97**:249–52.

McLellan AT, Kushner H, Metzger D *et al*. (1992) The Fifth Edition of the Addiction Severity Index. *J Subst Abuse Treat* **9**:199–213.

McLellan AT, Lewis DC, O'Brien CP *et al*. (2000) Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. *JAMA* **284**:1689–95.

Mitchell C, Simpson D, Chick J. (1997) Carbohydrate deficient transferrin in detecting relapse in alcohol dependence. *Drug Alcohol Depend* **48**:97–103.

Neumann T, Spies C. (2003) Use of biomarkers for alcohol use disorders in clinical practice. *Addiction* **98**(Suppl 2):81–91.

Nyblom H, Berggren U, Balldin J *et al*. (2004) High AST/ALT ratio may indicate advanced alcoholic liver disease rather than heavy drinking. *Alcohol Alcohol* **39**:336–9.

Saitz R, Larson MJ, Labelle C *et al*. (2008) The case for chronic disease management for addiction. *J Addict Med* **2**:55–65.

Saitz R, Cheng DM, Winter M *et al*. (2013) Chronic care management for dependence on alcohol and other drugs: the AHEAD randomized trial. *JAMA* **310**:1156–67.

Salaspuro M. (1987) Use of enzymes for the diagnosis of alcohol-related organ damage. *Enzyme* **37**:87–107.

Salaspuro M. (1999) Carbohydrate-deficient transferrin as compared to other markers of alcoholism: a systematic review. *Alcohol* **19**:261–71.

Schellenberg F, Schwan R, Mennetrey L *et al*. (2005) Dose-effect relation between daily ethanol intake in the range 0–70 grams and %CDT value: validation of a cut-off value. *Alcohol Alcohol* **40**:531–4.

Schwan R, Albuisson E, Malet L *et al*. (2004) The use of biological laboratory markers in the diagnosis of alcohol misuse: an evidence-based approach. *Drug Alcohol Depend* **74**:273–9.

Seale JP, Boltri JM, Shellenberger S *et al*. (2006) Primary care validation of a single screening question for drinkers. *J Stud Alcohol* **67**:778–84.

Sillanaukee P, Olsson U. (2001) Improved diagnostic classification of alcohol abusers by combining carbohydrate-deficient transferrin and gamma-glutamyltransferase. *Clin Chem* **47**:681–5.

Sillanaukee P, Massot N, Jousilahti P *et al*. (2000) Dose response of laboratory markers to alcohol consumption in a general population. *Am J Epidemiol* **152**:747–51.

Smith KL, Horton NJ, Saitz R *et al*. (2006) The use of the mini-mental state examination in recruitment for substance abuse research studies. *Drug Alcohol Depend* **82**:231–7.

Sobell LC, Sobell MB. (1992) Timeline follow-back. A technique for assessing self-reported alcohol consumption. In Litten AE (ed). *Measuring Alcohol Consumption: Psychosocial and Biochemical Methods*. Totowa, NJ: Humana Press, 41–72.

Sobell LC, Sobell MB. (1995) Alcohol Timeline Followback (TLFB) Users' Manual. Toronto, Canada: Addiction Research Foundation.

Sobell MB, Sobell LC, Klajner F *et al*. (1986) The reliability of a timeline method for assessing normal drinker college students' recent drinking history: utility for alcohol research. *Addict Behav* **11**:149–61.

Sobell LC, Agrawal S, Annis H *et al*. (2001) Cross-cultural evaluation of two drinking assessment instruments: alcohol timeline followback and inventory of drinking situations. *Subst Use Misuse* **36**:313–31.

Ware J, Jr, Kosinski M, Keller SD. (1996) A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* **34**:220–33.