ASSESSMENT AND DETECTION

Validation of the Bayesian Alcoholism Test Compared to Single Biomarkers in Detecting Harmful Drinking

Sanne Korzec1, Alex Korzec2,∗, Katherine Conigrave3,4, Janneke Gisolf5 and Boris Tabakof6

1Informatics Institute, University of Amsterdam (UvA), Amsterdam, The Netherlands, 2Department of Psychology, St. Lucas Andreas Hospital, Amsterdam, The Netherlands, 3Drug Health Services, Royal Prince Alfred Hospital, Sydney, Australia, 4Faculty of Medicine, University of Sydney, Sydney, Australia, 5Department of Physiology, Academic Medical Centre, Amsterdam, The Netherlands and 6Department of Pharmacology, University of Colorado, School of Medicine, Anschutz Medical Center, Aurora, CO, USA

∗Corresponding author: A. Korzec, Sint Lucas Andreas Ziekenhuis, Post Box 9423, 1006 AE Amsterdam, The Netherlands. Tel: +31-20-5108562; Fax: +31-20-6126431; E-mail: skorzec@science.uva.nl and a.korzec@slaz.nl

(Received 25 August 2008; first review notified 10 October 2008; in revised form 6 January 2009; accepted 13 February 2009; advance access publication 16 March 2009)

Abstract — Aims: Conventional tests for alcohol dependence often fail to detect hazardous and harmful alcohol use (HHAU) accurately. We previously validated the Bayesian Alcoholism Test (BAT) for the detection of HHAU among males. This uses 15 biochemical and clinical variables, including questionnaire data to calculate the probability of harmful (≥ 80 g alcohol/day) and moderate (≤ 40 g/day) drinking. Here we investigate the BAT’s diagnostic performance when more limited clinical data are available.

Methods: The WHO/ISBRA Collaborative Project recruited subjects from the general community and alcohol dependence treatment services. We analysed data from male drinkers: 318 alcohol dependent, 220 heavy and 712 moderate drinkers. Drinking was assessed using the Alcohol-Use Disorders and Associated Disabilities Interview Schedule. Eight of 15 markers used in the original BAT could be extracted from the WHO/ISBRA database.

Results: Comparing harmful to moderate drinkers, the area under the ROC curve for BAT (0.90) was significantly higher than that for CDT (0.82), GGT (0.77) and AST (0.76). Comparing hazardous to moderate drinkers, the area under the ROC curve for BAT (0.78) was significantly higher than that for AST (0.65) but not significantly higher than that for CDT (0.71) and GGT (0.70). For all 1250 subjects, the amount consumed correlated significantly better with BAT (0.65) than with AST (0.40) alone.

Conclusions: The BAT is more accurate than commonly used single biological markers in detecting harmful alcohol use, even when only half the input requirements are available. Computerized record keeping increases the practicality of use of algorithms in the detection of harmful drinking.

INTRODUCTION

Alcohol problems still often go undiagnosed in primary health care settings and hospitals (Denny et al., 2003; Shourie et al., 2007), despite various clues to diagnosis that exist in the medical record or laboratory results. In past years, various authors have explored computer algorithms to aid the diagnosis of alcohol use disorders (e.g. Lichtenstein et al., 1989). At the time when these algorithms were first described, there was only limited interest in their use, largely because of a perception that they would be too expensive or cumbersome for routine use. With increasing use of electronic medical records and electronically stored laboratory results, the roles of algorithms that make use of existing data warrant re-exploration. Furthermore, the lack of improvement in the rate of clinical diagnosis of alcohol use disorders, despite many years’ efforts at training medical students and doctors, suggests that we need to re-explore ways of better utilizing available clinical information.

Alcohol problems represent a heterogeneous set of disorders. Two overlapping conceptual frameworks are used to categorize them. The first approach comprises the psychiatric diagnoses of alcohol dependence and alcohol abuse (Alcohol Use Disorders: AUD). The second approach examines the amount of alcohol consumed. Various thresholds and terminology have been chosen to define the level of drinking. For example, the term hazardous and harmful alcohol use (HHAU) has been used to describe drinking above selected thresholds (Saunders and Lee, 2000; Conigrave et al., 2002).

Research in enhancing diagnostic accuracy of alcohol use disorders has used decision trees or sequential diagrams. The use of Bayesian networks has the advantage over decision trees of making use of a priori knowledge of the population. Each additional piece of clinical information that is added increases diagnostic accuracy. In contrast, decision trees carry the risk of ending up in an incorrect branch after each decision.

In 2004, a Bayesian expert system was developed (Korzec et al., 2005) for detection of HHAU and not for the broad spectrum of alcoholism. The Bayesian Alcoholism Test (BAT) consists of a graphical structure, the nodes of which represent diseases, symptoms and biochemical tests. An arrow going from disease to symptom or from disease to biochemical test indicates that the symptom or test is dependent on that disease (Fig. 1). Apart from their graphical structure, the Bayesian network also works with conditional probability tables that provide the conditional probability distribution that a certain disease causes different symptoms or biochemical abnormalities. The two kinds of information, graphical and probabilistic are combined and result in a calculated probability that a patient is suffering from that disease. This system was tested in a population of harmful, hazardous and moderate drinkers (Korzec et al., 2005). The BAT proved to be more accurate than traditional single markers such as gamma glutamyltransferase (GGT) and carbohydrate-deficient transferrin (CDT) in detecting harmful alcohol use.

However, it is not always possible to obtain the full range of clinical and biochemical indicators that constitute the BAT. In this study, we explore the diagnostic properties of the BAT in a situation where only partial data are available. The hypothesis investigated in this study is that the BAT is a more accurate tool to detect HHAU than other, currently used single tests such as CDT and GGT, even when only partial input data are available.

© The Author 2009. Published by Oxford University Press on behalf of the Medical Council on Alcohol. All rights reserved
Fig. 1. Network for the Bayesian Alcoholism Test. The *a priori* probabilities for diseases and states (left) are combined with the biochemical (right) and clinical findings (bottom). An arrow going from disease to symptom or biochemical test indicates that the symptom or test is dependent on the disease or state.

**MATERIALS AND METHODS**

**Subjects and selection**

Data were obtained from the WHO/ISBRA Collaborative Project (Conigrave et al., 2002). Subjects in this study were recruited from five countries: Australia, Brazil, Canada, Finland, and Japan, and were aged 18 and over. Subjects were recruited from inpatient alcohol dependence treatment services or from the general community by approach through employers, occupational groups or advertisements. General hospital inpatients were not recruited. Patients resident in alcohol dependence treatment units were interviewed within 72 h of admission. Subjects were excluded if they had major medical or psychiatric disorders other than AUD, if they reported intravenous drug use, were being treated for dependence on a substance other than alcohol or had received disulfiram treatment in the past month. Except for the non-drinking group, patients were excluded if they had not consumed a drink containing alcohol in the past month. Of the 1863 interviewed subjects, 1250 were male. Only men were selected for the current analysis because of their higher number and also because the BAT has so far only been validated for use in men. Of the 1250 males, 57% drank <40 g ethanol per day, 17.6% drank between 40 and 80 g and 25.4% drank >80 g ethanol per day in the last month (number of drinks per 30 days). In this selected population, a total of 43% drank >40 g of ethanol per day in the last month.

For the purposes of this study, subjects consuming >80 g of ethanol per day are defined as harmful users and subjects consuming between 40 and 80 g of ethanol per day as hazardous users (Saunders and Lee, 2000; Conigrave et al., 2002). Subjects drinking ≤40 g of ethanol per day are defined as moderate drinkers and are used as a control group.

**Instruments, blood samples and clinical data**

Drinking quantity and behaviour was assessed using the WHO/ISBRA interview schedule, which was an adaptation of the Alcohol Use Disorders and Associated Disabilities Interview Schedule (AUDADIS) (Grant et al., 1995). As part of this, alcohol use frequency and quantity of beverage-specific alcohol consumption during the past 30 days were assessed. Blood samples were collected for CDT, GGT and aspartate aminotransferase (AST). GGT and AST were assayed by reflectance spectrophotometry using a Vitros 250 Analyser (Ortho Clinical Diagnostics, Rochester, NY, USA). Serum CDT determinations were carried out in duplicate using the CDTect test, where the CDT content is expressed as an absolute amount (in U/litre, where 1 U refers to ~1 ng transferrin) of the transferrin isoforms with a pI > 5.7 (i.e. α-, mono- and part of disialo transferrin) (Conigrave et al., 2002). In previous WHO/ISBRA analyses, GGT and AST levels of >40 U/litre were considered elevated. The BAT cut-off was set to 0.5. Within the BAT, the cut-off values used for GGT, AST and CDT are as in the original BAT, where GGT levels of >65 U/litre and AST levels of >45 U/litre are considered elevated. In both studies, CDT levels of >20 U/litre were considered elevated. Information about diabetes, smoking habits, body mass index (BMI) and whether subjects self-reported an enlarged liver was also collected.

**Bayesian Alcoholism Test**

As described above, the BAT uses conditional probability tables to give the conditional probability distribution that a disease has caused different symptoms and biochemical abnormalities. By using Bayes’ theorems, one can calculate the extent to which the presence of a symptom predicts the presence of
HHAU. For each individual \( i \), BAT returns an output value \( v(i) \): \( \Omega \rightarrow [0–1] \) (\( v \) can take any value out of the set \( \Omega \) ranging from 0 to 1). In the earlier validation study, the cut-off value of BAT for diagnosing HHAU was set at 0.5 (Korzec et al., 2005). A module was added to the existing BAT program to automatically calculate the BAT for large numbers of data. An advantage of the BAT above single diagnostic tests is that it allows combining of results of many tests. The original BAT (BAT15) used 15 clinical and biochemical indicators (GGT, alanine aminotransferase (ALT), AST, ALT/AST ratio, CDT, alkaline phosphatase (AP), mean corpuscular volume (MCV), hepatitis risk, diabetes, BMI, palpable liver, spider naevi, level of response to alcohol (LRA), smoking and responses to the CAGE questionnaire). Seven comparable indicators were present in the WHO/ISBRA dataset (GGT, AST, CDT, diabetes, BMI, self-report of enlarged liver and smoking). The CAGE items were partially present. In the present study, we used the input information that was available; for most subjects, this amounted to 8 clinical and biochemical indicators instead of 15. The resulting BAT output value (BATs) therefore differed from the original (BAT15) output value. Since BAT uses prevalence probabilities for calculating the chance of someone having HHAU, we used the prevalence of HHAU extracted from the dataset.

**Results**

### Sensitivity, specificity and likelihood ratios

Comparing hazardous drinkers with moderate drinkers, the sensitivity of BAT (75.8%) was significantly higher than that of CDT (60.0%) and AST (45.0%), but not GGT (67.3%). The specificity of BAT (90.0%) was comparable to that of CDT (91.9%) and AST (90.1%), and significantly higher than that for GGT (73.7%). The positive and negative likelihood ratios of the BAT (7.6/0.27) were significantly better than those for CDT (6.2/0.43), GGT (4.4/0.58) and AST (5.8/0.66) (Table 2).

Comparing hazardous drinkers with moderate drinkers, the sensitivity of BAT (42.7%) was not significantly more sensitive than CDT (37.1%) and GGT (56.6%) but was significantly more sensitive than AST (23.6%). The specificity of the BAT (90.0%) was comparable to that of CDT (90.0%) and AST (90.1%), but BAT was significantly more specific than GGT (73.7%). The positive and negative likelihood ratios of the BAT (4.3/0.64) were significantly better than those of CDT (2.8/0.87) and AST (2.8/0.87).

### Diagnostic accuracy

Comparing hazardous drinkers with moderate drinkers, BAT correctly identified 241 hazardous drinkers in this dataset and missed 77. BAT correctly identified 641 moderate drinkers and categorized 71 subjects incorrectly. The area under the curve for BAT (0.90) was significantly higher than the curves for CDT (0.82), GGT (0.77) and AST (0.76).

Comparing hazardous drinkers with moderate drinkers, BAT correctly identified 94 hazardous drinkers in this dataset and missed 126. BAT correctly identified 641 moderate drinkers and categorized 71 subjects incorrectly.
Table 2. Sensitivity and specificity and likelihood ratios of BAT8, CDT, GGT and AST for identifying harmful alcohol use (>80 g/day) as compared with controls (<40 g/day)

<table>
<thead>
<tr>
<th>Test</th>
<th>n</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Likelihood ratio+</th>
<th>Likelihood ratio−</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAT8</td>
<td>1030</td>
<td>75.8** (70.8–80.2)</td>
<td>90.0* (87.6–92.0)</td>
<td>7.6*** (6.0–9.6)</td>
<td>0.27*** (0.22–0.33)</td>
</tr>
<tr>
<td>CDT</td>
<td>980</td>
<td>60.0 (54.4–65.4)</td>
<td>91.9 (89.6–93.7)</td>
<td>6.2 (4.8–7.9)</td>
<td>0.43 (0.37–0.49)</td>
</tr>
<tr>
<td>GGT</td>
<td>977</td>
<td>67.3 (61.8–72.4)</td>
<td>73.7 (70.3–76.9)</td>
<td>4.4 (3.4–5.6)</td>
<td>0.58 (0.52–0.65)</td>
</tr>
<tr>
<td>AST</td>
<td>977</td>
<td>45.0 (39.5–50.7)</td>
<td>90.1 (87.6–92.1)</td>
<td>5.8 (4.2–7.9)</td>
<td>0.66 (0.60–0.72)</td>
</tr>
</tbody>
</table>

Bold values: the best result in the table.

*Significant difference BAT8 compared to GGT at the P level < 0.05.
**Significant difference BAT8 compared to CDT and AST at the P level < 0.05.
***BAT8 compared to CDT, GGT and AST. Significant difference at the P level < 0.05.

Values within parentheses: 95% confidence intervals.

Fig. 2. Areas under the curve of BAT8, CDT, GGT and AST in the detection of harmful drinkers (>80 g/day) and moderate drinkers (<40 g/day).

Table 3. Areas under the curve of BAT8, CDT, GGT and AST for harmful drinkers (>80 g/day) and for hazardous drinkers (≥40 g/day and ≤80 g/day) compared with the control group (<40 g/day)

<table>
<thead>
<tr>
<th>Test</th>
<th>AUC comparing harmful users with controls</th>
<th>AUC comparing hazardous use with controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAT8</td>
<td>0.90†† (0.87–0.92) standard error 0.011</td>
<td>0.77† (0.72–0.80) standard error 0.020</td>
</tr>
<tr>
<td>CDT</td>
<td>0.82 (0.79–0.85) standard error 0.016</td>
<td>0.72 (0.67–0.75) standard error 0.021</td>
</tr>
<tr>
<td>GGT</td>
<td>0.77 (0.74–0.80) standard error 0.017</td>
<td>0.70 (0.66–0.74) standard error 0.021</td>
</tr>
<tr>
<td>AST</td>
<td>0.76 (0.72–0.79) standard error 0.018</td>
<td>0.65 (0.61–0.69) standard error 0.021</td>
</tr>
</tbody>
</table>

Bold values: the best result in the table.
†BAT8 compared to AST. Significant difference at the P level < 0.05.
††BAT8 compared to CDT, GGT and AST. Significant difference at the P level < 0.05.

Values within parentheses: 95% confidence intervals.

Fig. 3. Areas under the curve of BAT8, CDT, GGT and AST in the detection of hazardous drinking (40–80 g/day) and moderate drinkers (drinking <40 g/day).

The diagnostic accuracy of BAT compared to CDT was 0.86:0.82. The area under the curve for BAT (0.77) was not significantly higher than that for CDT (0.72) and GGT (0.70) but was significantly higher than for AST (0.65). Figure 2 shows the ROC curve for the detection of harmful from moderate drinkers. Figure 3 shows the ROC curve for the detection of hazardous from moderate drinkers. Table 3 summarizes the areas under the curve of both Figs. 2 and 3.

Correlations
Using pooled data from all 1250 males included in the WHO/ISBRA dataset, the amount of drinking demonstrated a significantly better correlation coefficient with BAT8 0.647 (CI: 0.613–0.678) than with CDT 0.515 (CI: 0.472–0.555), GGT 0.438 (CI: 0.390–0.482) or AST 0.393 (CI: 0.344–0.440) alone (CI: 95%).

DISCUSSION
Background
Currently, single tests to detect HHAU are usually based on questionnaires or biochemical markers, and both have limitations. Current tests based on interview or questionnaires to identify alcoholism have the following limitations. It may be difficult to obtain accurate data on alcohol intake in patients with impaired ability to provide a history, or in situations, such as medicolegal assessments, where there may be a strong motivation to minimize self-report of drinking. Biochemical tests,
like CDT, are often either not sensitive or not specific enough in the detection of HHAU.

The BAT attempts to address both problems. It uses biochemical markers to circumvent problems in the clinical history but allows additional information (e.g. CAGE) to enter into the diagnosis.

The present study confirms previous findings that BAT has significantly better diagnostic properties compared with CDT or GGT in detecting harmful alcohol use (Korzec et al., 2005). It also extends these findings by showing that BAT has better diagnostic parameters for identification of harmful use when only limited clinical input variables are available. BAT is also correlated significantly better with actual alcohol consumed than do CDT, GGT or AST. BAT, on the other hand, is no better than conventional single laboratory markers in the identification of hazardous drinkers.

Limitations

Only 8 out of 15 markers from the original BAT were used. BAT is able to provide responses to queries of the following type: ‘given values obtained for some, but not necessarily all diagnostic tests, what is the probability that a subject has a harmful or hazardous alcohol intake?’ One would expect that deleting important markers from the BAT like ALT and MCV would negatively affect the comparison between BAT and single markers like CDT and GGT. In keeping with this, the sensitivity and the specificity of BAT (75.8%/90.0%) in the current study are not as good as those of BAT (94.0%/97.9%) in the original study (Korzec et al., 2005).

Secondly, one can question how well AUDADIS questions concerning drinking problems as used in the WHO/ISBRA Collaborative Project correspond with the CAGE questions. However, it seems unlikely that the questions used in lieu of the CAGE items (Table 1) would artificially increase the accuracy of the BAT. In contrast, those items were selected in such a way to minimize the risk that the ‘CAGE’ questions would score inappropriately positive. Furthermore, in BAT, the maximum of positive ‘CAGE’ question is 3 whereas it is 4 in BAT.

A third limitation is that in this WHO/ISBRA dataset and in previous studies on the BAT, the prevalence of HHAU is high. The performance of BAT remains to be established in populations with a lower prevalence of HHAU.

Remaining issues

This analysis showed that using an algorithm to combine information both from a limited medical history (diabetes, BMI, self-report of enlarged liver and smoking status and three questions broadly analogous to CAGE items) and from laboratory tests increases the accuracy of detection of harmful alcohol use compared with single laboratory tests. It could be argued that the effort of collecting additional clinical and biochemical indicators in BAT was not worthwhile, considering the relatively small (though statistically significant) increase in accuracy. However, with the advent of computerised medical records, more routinely collected clinical information can potentially be made available for use in diagnostic algorithms without extra effort. These developments in technology warrant re-examination of the use of computerized algorithms to aid detection of HHAU. It is to be hoped that alerting the clinician to clusters of relevant information will raise their rate of diagnosis of alcohol problems, and their provision of treatment.

Further studies could delineate the most parsimonious set of variables from the original BAT that can contribute to diagnostic accuracy.

CONCLUSIONS

The BAT has a greater diagnostic accuracy than commonly used single biological markers in detecting harmful alcohol use, even when only half the input requirements of the original BAT are available. Additional research is needed to further validate and refine the BAT in a female population, in subjects with liver disease and in populations with a low prevalence of HHAU. The advent of computerization of clinic records increases the practicality of using algorithms to raise the clinician’s awareness of the risk of alcohol problems.

REFERENCES