Behavioral Sciences and the Law
Behav. Sci. Law **29**: 116–137 (2011)
Published online 21 May 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/bsl.935

Judging Intoxication

Steve Rubenzer, Ph.D., ABPP*

Judgments of whether a person is intoxicated by alcohol are important in a number of civil and law enforcement settings. This paper reviews how well people are able to make such judgments, the evidence for individual signs of intoxication, several structured rating techniques, and the use of sobriety tests. It is concluded that observers relying on common-sense clues of intoxication have limited ability to assess the blood alcohol concentration (BAC) levels of strangers, particularly below .10%. This generalization holds across professions that might be expected to show greater accuracy. Structured assessment instruments based on observable signs have shown promise but are confounded by the wide variations between casual social drinkers and those that have obtained a high level of tolerance. Among sobriety tests, only NHTSA's Standardized Field Sobriety Tests (SFSTs) have substantial, but seriously flawed, research support. Assessing the sobriety of strangers in the low to moderate BAC ranges without resort to chemical tests remains a daunting task. Copyright © 2010 John Wiley & Sons, Ltd.

Persons in a variety of settings may need to assess whether another person has had too much to drink. Among those who face such decisions are police officers investigating a possible drunk driving offense, social hosts and establishments who serve alcohol, employers concerned that workers have been drinking on the job, and partygoers who are relying on a friend to provide a safe ride home. This paper will review the literature on assessment of alcohol intoxication, emphasizing the problem of *driving under the influence* (DUI) investigations. Readers primarily interested in the National Highway Traffic Safety Administration's (NHTSA) Standardized Field Sobriety Tests are referred to the work of Rubenzer (2007/2008). The current paper will review the ability of several professions to assess intoxication, then will examine structured instruments and sobriety tests, and finally will review the evidence available for individual signs of intoxication. Readers interested in the effect of alcohol on laboratory tasks and discreet physical and cognitive functions are referred to reviews by Holloway (1994, 1995), Kruger (1993), Mokowitz and Robinson (1988), and Moskowitz (2000).

UNSTRUCTURED ASSESSMENTS OF INTOXICATION

Individual studies have assessed the ability of college students, police officers, physicians, bartenders, psychologists, and alcohol counselors to assess alcohol intoxication. In general, none of these groups show a consistent ability to diagnose intoxication without access to specialized tests.

*Correspondence to: Steve Rubenzer, Forensic Sobriety Assessment, LLC, 11914 Astoria, Suite 490, Houston, TX 77089, U.S.A. E-mail: rubenzer.steve@att.net

Copyright © 2010 John Wiley & Sons, Ltd.

McGuire (1986) compared the breath samples of 1115 drivers near busy intersections at three timeslots during weekend nights. Graduate students collected breath samples and classified drivers into one of three categories: *appears sober*, *possibly under the influence*, and *definitely under the influence*. These ratings correlated only .34 with BAC.[1] Raters identified only 20.1% of drivers with BACs over .10% as intoxicated, but the false positive rate was very small at 1.8%. Ninety-three percent of those judged sober had a BAC of .00%. Using the survey data to estimate base rates at sobriety checkpoints, and comparing these figures with arrest rates, the authors estimated that checkpoints were correctly identifying only 20–22% of drivers with a BAC over .10%. In fact, this may be an overestimate, since all arrests were presumed to be valid (not false positives). However, both parts of the study suggested that many legally intoxicated drivers were escaping detection—or at least arrest.

Carroll, Rosenberg, and Funke (1988) examined alcohol and mental health counselors' abilities to detect intoxication of a videotaped 21 year old subject over four levels of BAC. Alcohol counselors were found no more accurate than mental health counselors, and more experienced alcohol counselors were little better than their less experienced colleagues. However, with only one drinking subject, the generalizability of this finding is questionable and in need of replication.

Burns, Nusbaumer, and Reiling (2003) examined the relative weight bartenders put on various signs of intoxication, but did not examine the correlation of signs to BAC. *Slurred speech*, *disturbing others in the bar*, *clumsiness*, *number of drinks served*, and *mood changes* were given the highest relevance ratings, all above 4.0 on a five-point scale. *Redness of face* was given the lowest rating at 2.66.

In the first study involving police, Pagano and Taylor (1970) investigated the ability of 18 officers to judge the intoxication of 36 students dosed to obtain .04% or .10% BAC. Students were asked to assume the role of a spouse being investigated in the aftermath of a domestic squabble in order to prevent self-consciousness regarding intoxication and its display. Officers made three assessments: the first 30 seconds after initial contact, the second five minutes later, and the third 15 minutes later. The third assessment was made after officers completed an alcohol assessment checklist. Officers were asked to estimate BAC by making a check mark on a scale (line) with divisions for different BACs ranging from .00% to .30%. Subjects were breath tested before and after the contact, and each officer assessed both a low and high BAC subject. Mean BAC predictions for the low dose condition were reported to be fairly accurate, but the standard deviation is not provided. In contrast, BAC estimations for subjects in the .10% group were quite low, averaging .058, .065, and .062% for the three assessments. In fact, the BAC estimates were not significantly different between the low and high alcohol dose conditions. Officers were also asked to rate their confidence in their BAC assessment on a five-point Likert scale. For the low dose condition, a mean value of 2.94 was observed compared with 2.76 for high dosage subjects. Confidence ratings tended to increase from the first to the third assessment rating, but neither confidence nor years of police experience were related to accuracy.

Langenbucher and Nathan (1983) conducted three experiments in which they examined in turn the ability of social drinkers, bartenders, and police officers to accurately assess the BAC of drinking subjects. In the first study, 49 social drinkers

[1] While this paper, and many others, refers to blood alcohol concentration, the reader should be aware that in many studies BAC is estimated through breath testing equipment.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

observed two men and two women subjects at .00, .05, and .10% BAC. They were asked to classify the subjects as *slightly intoxicated*, *moderately intoxicated*, or *very intoxicated/legally drunk*, and also to provide ratings of intoxication on a 50-point continuous scale. Observers rated confidence in their assessment on a seven-point Likert scale. They sat in two rows of chairs arranged in a semicircular configuration as they observed the subject during a brief, two to three minute interview. They also observed the subject enter and leave. Only four of 16 classifications were correct, and in all four instances, legally intoxicated targets were not identified as such. Only two of five sober targets were correctly classified (false positive rate = .60) and the other three subjects were rated as either moderately intoxicated or legally drunk. The continuous ratings for legally intoxicated targets were grossly inaccurate, usually underestimated. No relationship between estimation accuracy and observer confidence was found.

In the second study by Langenbucher and Nathan, the authors examined the ability of 12 bartenders with an average of five years experience. In addition to the behavioral observations available in the first study, they observed the subject negotiate a staircase before and after the interview. The bartenders correctly rated targets in only one of four instances (this subject was sober, but the other sober target was rated as moderately intoxicated). Nine out of 12 bartenders indicated that they would continue to serve a subject whose BAC was .11%. Bartenders' estimates were somewhat more accurate than were social drinkers, but no relationship between years of experience as a bartender and BAC estimation accuracy was found.

The final study examined the ability of 30 New Jersey police officers at assessing intoxication in two settings. The laboratory setting was the same as for previous studies. In the other condition, officers approached a subject seated in a car and were allowed to test the sobriety of the target in any way they wished for three minutes, including having the subject exit the vehicle and administration of sobriety tests. Overall, police officers were less accurate estimators of BAC than bartenders. However, five police officers were consistently more accurate than the others as a group. Four of the five were members of the special tactical unit for the apprehension of drunk drivers, and had received more than 90 hours of training in the identification of drunk drivers and administration of sobriety tests. They also averaged 216 alcohol-related arrests compared with 86 for the less accurate officers. All of them accurately classified the sober subjects, but underestimated the BAC of drinking subjects.

The Langenbucher and Nathan study is limited by the very small number of drinking targets examined (four in each study) and incomplete reporting of data. A tabular presentation with sensitivity and specificity statistics would be much more informative and easier to digest.

Grossman et al. (1996) examined the accuracy of police officers' intoxication judgments by comparing records from the trauma center registry and the Washington State Patrol Traffic Accident Database. In this way, they were able to assemble a database of 1336 subjects in which both sobriety assessments and BAC measures were taken post-accident. Unfortunately, little information about the police officers' assessments is provided, and in at least some cases it included sobriety tests and portable breath tests. Subjects were initially classified by police officers into four categories: (a) *had not been drinking* ($n = 746$), (b) *had been drinking, not impaired* ($n = 22$), (c) *had been drinking, sobriety unknown* ($n = 168$), and (d) *had been drinking, impaired* ($n = 568$). For the main analysis, the first two groups were combined into a *not impaired* group and the *had been drinking/sobriety unknown* group (mean BAC = .13%)

Copyright © 2010 John Wiley & Sons, Ltd.
DOI: 10.1002/bsl

was dropped. The authors reported very good diagnostic statistics for a criterion of .10% BAC, with a sensitivity of .91 and specificity of .90. When those subjects who had been administered portable breath tests were omitted from the analysis, the sensitivity was reduced to .74, but specificity was even higher at .97. However, the intoxicated group had a mean BAC of .19%, while the *completely sober* and *drinking but not intoxicated* groups had mean BACs of .02% and .05% respectively. More importantly, only 22 subjects were in the *drinking and not intoxicated* group vs. 746 in the *completely sober* group. Recall also that a substantial ($n = 168$) group was dropped because officers could not make a definite judgment, despite a mean BAC of .13%. Thus, the diagnostic statistics reflect the ability to discriminate very low BAC subjects from very high BAC subjects, and are much less impressive than presented. The authors also reported that, when a criterion of .08% was utilized for dichotomous judgments of intoxication, diagnostic statistics were equivalent to those obtained for .10%. The authors concluded that police officers were able to recognize drunk drivers with a fairly high degree of accuracy when investigating crashes in which the driver is transported to trauma center. However, without knowledge of which sobriety tests were performed, on what number of cases—and the given the virtual absence of subjects in .02–.12% BAC range—the findings are of little clear value.

Brick and Carpenter (2001) examined the accuracy of judgments of intoxication by police officers registered in DUI detection courses. Six individuals, all moderate drinkers, were given alcohol sufficient to obtain BAC levels of .08–.09, .11–.13, and .15–.16%. The instructor engaged them in conversation for 30–60 seconds and these conversations were videotaped and then shown to police officers in random order. Officers were then asked six questions regarding the subject, including whether the person had been drinking, whether they would continue to serve the person as a bartender or social host, and whether the person was too drunk to drive an automobile. Confidence ratings were also obtained for each question, but results are presented only in charts and no standard deviations are provided. Officers were uncertain whether persons in the .08–.09% BAC condition had been drinking, and they actually received substantially higher probability ratings than those in the medium (.11–.13%) BAC condition. Only for subjects with .15–.16% BAC were the majority correctly identified. Officers judged most subjects in the low and medium conditions as able to drive a car and most did not see a problem with serving them further drinks. Police officers were often very confident of their observations and conclusions even when they were incorrect.

Although this article is primarily interested in judging intoxication from the perspective of an observer, several articles have reported the ability of people to judge their own BAC or level of intoxication. Both positive (Bois & Vogel-Sprott, 1974; Russ, Harwood, & Geller, 1986) and negative results (Maisto & Adesso, 1977) have been reported.

To summarize, there is little evidence that police officers, bartenders, mental health professionals, or alcohol counselors can accurately assess intoxication of strangers at moderate levels of intoxication from informal observations. In addition, the limited evidence available suggests that even extensive experience serving drinkers or assessing drunk drivers does not substantially improve this skill without reliance on sobriety or breath tests. Significant numbers of sober or low BAC subjects were identified as intoxicated in several studies, while substantial numbers of legally intoxicated subjects escaped detection. Despite low levels of accuracy, police officers tend to be quite confident in their judgments, and the evidence is consistent in showing little relationship between confidence and accuracy.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

STRUCTURED APPROACHES

Bogen (1927a, 1927b, 1928) appears to have completed the earliest formal investigations of intoxication and its physical and behavioral correlates. Patients suspected of acute alcohol intoxication were rated on a number of physical symptoms, including *odor of alcohol* (OOA), *pupil dilation or constriction*, *ability to stand and walk*, and *speech*. As part of the evaluation, examiners asked patients to walk across a hallway, to stand with feet together and eyes closed without swaying, to touch the tip of the nose with eyes closed, and to repeat the phrase ''Methodist Episcopal'' (Bogen, 1928). Results were presented in terms of percentages of patients showing a given sign at five levels of urine alcohol. Surprisingly, a number of intoxication signs showed a complex relationship with urine alcohol level. For example, *swaying while standing* was present in 46% of those suspected of acute intoxication but with less than 1 mg of alcohol per cc of urine, 81% of those with 1 mg/cc, but only 15% of those with 5 mg/cc. Slurred or confused speech was also considerably less frequent in the highest urine alcohol category than in several lower categories. However, such results may be explained by the fact that substantial numbers of highly intoxicated subjects were unable to stand or speak at all. The author refers to the ''diagnosis'' of intoxication and used it to cross-tabulate results for the various physical symptoms, but there is no description of procedures or variables used to arrive at the diagnosis. No reliability figures are reported for any variables. Widmark (1981) carried out similar studies in Germany shortly after Bogen. A number of subsequent studies in the US (Burns, & Moskowitz, 1977; McKnight, Lange, & McKnight, 1999; McKnight, Langston, McKnight, & Lange, 2002; Sussman, Needalman, & Mengert, 1990) and Finland (Penttilä, Tenhu, & Kataia, 1974) evaluated multiple sobriety tests and indicators of intoxication. The results from both groups of studies will be discussed in the relevant content sections below.

Perham, Moore, Shepherd, and Cusens (2007) had graduate students briefly assess nearly a thousand ''city centre drinkers'' through observation of three signs of intoxication: *glazed eyes*, *slurred speech*, and *staggering gait*. Unlike most such studies, interrater reliability was reported for judgments of intoxication ($\rho = .552$). It should be noted that the Pearson correlation does not take account of differences in means between raters, only rank order, and an intraclass correlation might provide a both lower and more accurate index of agreement (Shrout & Fleiss, 1979). The correlation of the two surveyors' groups with BAC was reported to be $\rho = .556$ and .527, respectively, indicating a moderate relationship.

Several sets of authors have attempted to create behavior- or symptom-based rating instruments for the assessment of intoxication. Three studies examined the ability of physicians to assess intoxication using behavioral checklists or simple clinical status measures. Penner and Coldwell (1958) examined whether physicians' judgments of intoxication based on pulse rate, general appearance, gait, and mental status were related to performance on a closed driving course. Unfortunately, little information is provided about the administration or scoring rules for several sobriety tests (*picking up a coin*, *finger to nose*, *finger to finger*, *nystagmus*) that were administered or how the criterion was assessed. The authors concluded, ''Minor degrees of impairment of coordination are difficult if not impossible to detect unless the normal (performance of the individual) is known'' (p. 796), and reported that one of the two physicians recorded three times as many alterations in coordination as the other physician, whereas their judgments about impairment to drive showed a twofold difference. The authors concluded that their

Copyright © 2010 John Wiley & Sons, Ltd.
DOI: 10.1002/bsl

results showed little agreement between medical examination and driving impairment and that "some other means... should be used to determine alcoholic impairment of driving ability" (p. 800).

Rutherford (1977) examined the presence of various signs of alcohol intoxication in 114 ER patients. For patients with a BAC greater than .10%, those with the highest BAC levels (>.30%) had the highest false negative (failure to identify as intoxicated) rates (20%), while five of six patients in the .001–.099% range were regarded as having not drunk at all. Rutherford also examined the relationship of several indicators to BAC ranges, but provided tallies rather than summary statistics. He noted that red eyes were infrequently observed, possibly because of failure to assess for this indicator. Although 66 of 114 subjects had no BAC, three of them were judged to be have been drinking by clinical staff.

Holt, Stewart, Dixon, Elton, Taylor & Little (1980) recorded observations and judgments of intoxication by two emergency room physicians for 702 consecutive admissions. The physicians recorded four specific symptoms (red eyes, odor of alcohol, slurred speech, and abnormal coordination) and these ratings were tabulated by the results of portable breath readings following the examination. The authors reported few false positive errors but failure to detect 19% of patients with a BAC greater than .08%. The sample included a relatively small portion of subjects with BACs from .001-.08%, a factor that probably increased accuracy and should be recalled when findings from this study are cited below.

Cherpitel et al. (2005), in a WHO collaborative study, examined emergency room diagnoses of alcohol intoxication with BAC. The study's subjects were 4,798 patients from 12 countries presenting at emergency rooms within 6 hours of acute injury. Clinician raters were specially trained physicians and nurses. They were provided a brief form listing nine symptoms of alcohol intoxication (items not disclosed) and asked to provide ratings from *mild alcohol intoxication* to *very severe alcohol intoxication*. There does not appear to have been a category to indicate no alcohol consumption. Raters engaged subjects in a 25 minute interview, which included a number of standardized questions about alcohol use and asked about recent use. The authors reported that 84.6% of patients with a BAC of .06% or more were assessed as intoxicated, whereas 93.4% below this level were deemed not intoxicated. However, a tabulation of results (their Table 1) indicated that only 81.2% of those with a BAC of .000–.059% were judged as not intoxicated, and the discrepancy is not explained. The great majority of the sample (80%) had apparently not consumed alcohol (the article is unclear whether this classification was based on self-report). Accuracy rates were much lower for those

Table 1. Items of the Alcohol Symptom Checklist

Smell of alcohol
Fine motor control (impaired)
Gross motor control (impaired)
Slurred speech
Changed speech volume
Decreased alertness
Sweating
Respiration slow
Sleepiness
Pace of speech
Red eyes

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

who reported drinking within the last six hours, with raw agreement rates falling to below 30% for some groups and averaging 39% overall. In other words, the high accuracy rates reported appear largely due to correct assessments of completely sober subjects. Although there was concern that alcohol-tolerant drinkers might not be detected, in fact the opposite pattern was reported: Such persons were often assessed as intoxicated even at low BAC levels. While BAC correlated .65 with physician classifications into one of four categories defined by BAC (*none* [.00–.059%], *mild* [.06–.099%], *moderate* [.10–.119%], and *severe* [>.20%]), this was probably due to direct queries about consumption, the high proportion of sober subjects, and data collection in a treatment as opposed to law enforcement setting. Collateral data and disclosure from patients are confounds for field studies, and indeed, for nearly all studies to date involving judgments of intoxication.

Several other researchers sought to develop formal instruments for assessment of intoxication based on ratings of physical observations. Simpson-Crawford and Slater (1971) investigated the relation of hypothesized alcohol-related eye symptoms to BAC. These symptoms included *suffusion of conjunctivas* (red eyes), *eyelid drag*, *pupillary light reflex*, *diminished peripheral vision*, and *nystagmus* (a jerking movement of the eyes) when the subject was asked to visually follow a smoothly moving object). Most of the items appear to have been informal and little information is provided to guide scoring of them. Multiple regression of eye symptoms predicted BAC with an R^2 of .85, .81 corrected for shrinkage, among 50 archival cases. Few of the statistical and psychometric details are provided. A factor analysis of the study variables, including BAC, age, and weight, found three factors, two with substantial loadings on BAC. The first factor was defined primarily by nystagmus (.81), reduced peripheral vision (.66), red eyes (.58), and BAC (.55). One of these eye signs, nystagmus, was subsequently incorporated into one of NHTSA's Standardized Field Sobriety Tests, Horizontal Gaze Nystagmus (HGN).

Teplin & Lutz (1985) evaluated 28 previously identified signs of alcohol intoxication among 567 emergency room admissions. Physicians recorded the presence of these signs and took breathalyzer readings. The authors then evaluated each of the signs by multiple criteria, including interrater reliability, the correlation with BAC ($r > .50$), frequency of occurrence (>5% in the total sample), correlation with the total scale ($r > .50$), and first factor loading (>.50). Eleven items survived this process and were retained for the final scale, which was named the Alcohol Symptom Checklist (ASC; see Table 1). The authors reported impressive psychometric qualities for the ASC, including a coefficient alpha of .92, interrater reliability of .93, and correlation with BAC of .84 in a cross-validation sample. Because the majority of subjects had no alcohol in their system at all, the authors reported a separate analysis limited to those that did ($n = 166$). The correlation with BAC was slightly reduced but remained strong ($r = .77$). The correlation with BAC was significantly reduced for those with a heavy drinking history (defined as seven drinks or more per day), falling to .55, as it was for light drinkers ($r = .54$) and those who admitted taking illegal drugs ($r = .49$), but not for those referred for psychiatric evaluation ($r = .85$). The low correlation for light drinkers may have been due to a predominance of .00% BAC in that group, although no such analyses are reported. The authors reported diagnostic statistics for several cut-off scores using BAC criteria of .05% and .10%. Sensitivity for recommended cut-off scores ranged from .86 to .96 and specificities range from .91 to .97, with values for the two BAC levels comparable. The authors reported that the ASC was unable to discriminate between zero and near-zero BAC levels.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

Several additional factors should be noted. There was a very wide range of BACs in the sample, from .00% to about .40%. Although some previous studies showed physicians are not particularly good at estimating BAC, they performed exceptionally well in this study. Using a simple five-point scale, their ratings correlated .90 with BAC—better than did ratings on the ASC. This is likely to be due to a bimodal distribution of BACs, with many subjects into being completely sober or very drunk, or perhaps because patients simply reported to the raters how much they had to drink. Only about 3% of subjects were in the range of .05–.10%, and 15% in excess of .10%. The distribution of BACs above .10% is not given, which is an important omission.

Sullivan, Hauptman, and Bronstein (1987) attempted to cross-validate the ASC with 21 patients with histories of alcohol abuse who were admitted to the emergency department. All had a history of recent alcohol abuse and no recent drug abuse, and patients were excluded if they had a head injury or medical illness that would interfere with their level of consciousness. Most came to the hospital seeking alcohol detoxification, complaining of drinking too much, or asking to see a counselor. Patients had an average BAC of nearly .30%, but nonetheless scored below the recommended cut-off scores on the ASC (mean score = 2.62; recommended cut-off scores range from 4 to 6 for .10% BAC). Further, no significant correlation was found between BAC and ASC scores. Although the small number of subjects would mitigate against finding significant results, correlations ranged from only .18 to .24. It should be noted that, by selecting heavily intoxicated patients for the entire group, there was a restriction of range and subjects likely consumed considerably more alcohol than the heavy drinkers group in the study by Teplin and Lutz. Nonetheless, the work of Sullivan et al. is a serious failure to replicate the ASC among heavy drinkers, and provides empirical evidence to support the notion that many heavy drinkers can show little outward indication of intoxication even at high BAC levels (see also Perper, Twerski, & Wienand, 1986; Sobell, Sobell, & VanderSpek, 1979; Urso, Gavaler, & Van Thiel, 1981).

McKnight, Langston, Marques, and Tippetts (1997); see also McKnight, Langston, Schaefer, & Aasved, 1991; McKnight & Marquis, 1990) evaluated the ability of social hosts and bartenders to accurately classify over a thousand drinkers in natural settings according to their BAC levels. The authors initially assembled 220 signs of alcohol impairment

> from an extensive review of the literature, focus group discussions with wait persons, police, and social host observations of bar patrons in a prior study (McKnight, 1988). A four-person panel of alcohol researchers reduced the list to 166 cues based on reported strength of relationship to blood alcohol level and judged frequency of occurrence in social situations. The remaining cues were then consolidated into the final list types by grouping similar cues. (p. 249)

Raters either used their own, unguided judgment or completed a checklist of 22 symptoms. The average BAC level associated with onset of these symptoms (see Figure 1) was determined during an pilot study involving 149 drinkers, four observers. Drinkers consumed alcohol in private homes and, importantly, were known to the raters prior to drinking. For drinkers assessed in a public establishment, the drinkers were ordinary patrons who became subjects when servers considered them impaired and asked them for breath samples.

During the evaluation stage of the study, servers observed either small groups (four to six) or large groups (25–35) of guests. Servers were blind to the number of drinks consumed by the drinkers, as these were served by research assistants out of their sight.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

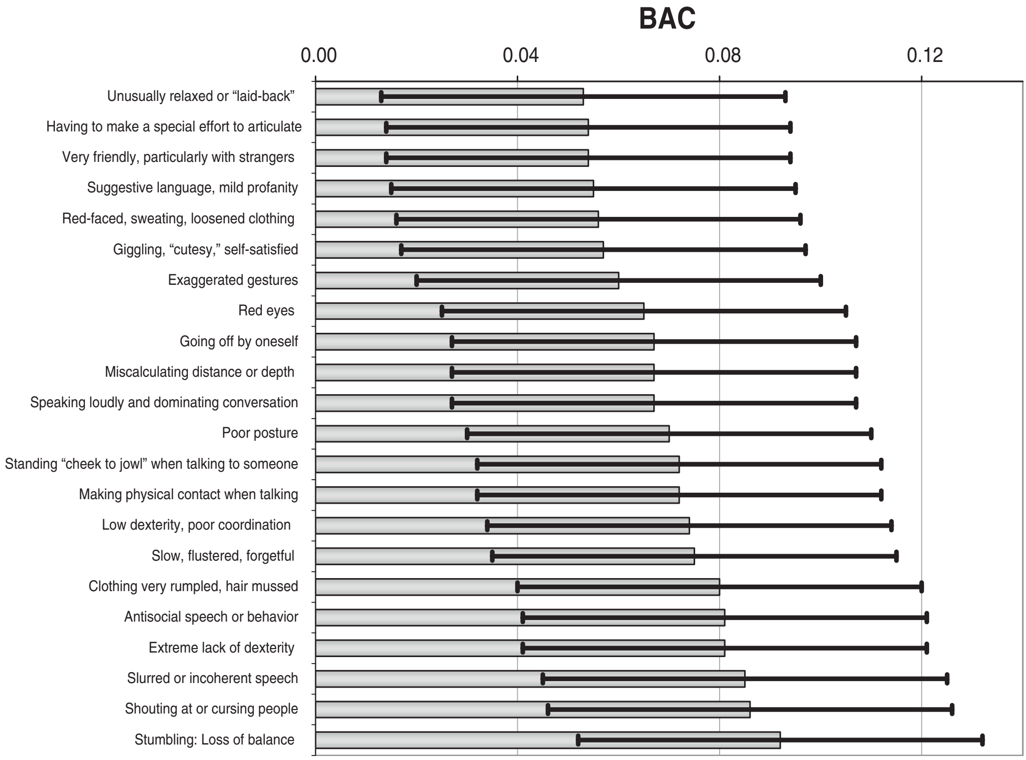

Figure 1. Alcohol-related signs in McKnight et al. (1997), average BAC at which they were first observed, and error bars indicating one standard deviation (SD; theoretically 68% of the population) based on an average SD of .04% BAC ($n = 962$).

This was not true for bartenders, however, and no separate analysis was performed for these two quite different groups. Social hosts were able to accurately classify subjects as below or above .04% BAC, achieving *d*' values from 1.5 to 2.0. There was a small advantage for those raters who used the alcohol symptom assessment guide over those that did not in the small social group condition. Hosts using the symptom guide achieved a true positive rate (sensitivity) of 81.4%, with a false positive rate of 24.3%. Raters were less successful at distinguishing those above or below .08%, the current legal standard in nearly all US jurisdictions, particularly in public drinking establishments where the subjects were strangers. Nonetheless, substantial discrimination was achieved ($d' = 1.2$–1.4), although sensitivity fell to as low as .33, with an average near .50. Thus, raters did better at discriminating drinkers above and below the .04% threshold than above or below 08%.

Perhaps of greatest interest was the ordering of alcohol-related symptoms, which implies the possibility of a Guttman-type scale, based on nearly a thousand subjects. At the lowest level, *unusually relaxed or "laid back"* was first observed at a mean BAC of .053%, while *stumbling: loss of balance* had a mean onset of .092%. Standard deviations are not given for individual cues, but they reportedly ranged from .032 to .047% BAC. Thus, a confidence interval embracing only one standard deviation shows that the overlap among symptoms is too great to provide any fine discrimination, as shown in Figure 1, using .04% as the SEM for all cues. The error bars indicate about 68% of the population: 95% error bars would be twice as large.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

INDICATIONS OF DRIVING IMPAIRMENT

Harris (1980) identified driving errors as indications of intoxication. In the first part of the study, 1,288 DWI arrest reports from nine different police agencies were examined, which contained a total of 3,658 visual detection cues such as weaving, stopping in a traffic lane, etc. The second part of the study involved "ride alongs" with police officers engaged in general patrol and DWI enforcement. A total of 643 detection events were recorded and a total of 1,681 cues observed. An average of 2.6 cues per event were observed, similar to the average of 2.8 cues per arrest reported in the national sample of arrest reports. In the ride along study, 38% of subjects had a BAC of equal to or greater than .10%, while the BAC of 63% of the sample was above or equal to .05%. Based on the data obtained, the author created probability values to indicate the likelihood that a driver had a BAC above the given criteria if a particular cue was present. Some of these achieved fairly high values (stopping without cause, .69; following too closely, .62; turning with a wide radius, .60), but these represented modest increases over the base rate (.38) and show limited diagnosticity. The author then offered a drunk driver detection guide, which listed various cues and "the percentage of nighttime drivers with BAC equal to or greater than .10" (p. 729). However, such estimates may be biased. Although research assistants accompanied officers and every car that displayed driving errors was supposed to be stopped, the fact that the average case had over two and a half errors suggests that stops were made more selectively. This cannot be assumed to be a fair sample of nighttime drivers, who may show isolated driving errors, and the probabilities reported are likely overestimates. For example, an officer may observe momentary weaving of the driver, but after further observation, decide that a stop is not warranted. Drivers that passed this subsequent assessment phase would not be included in Harris's sample, and are less likely to be impaired than those that do not. Finally, the figures cited represent the *positive predictive power* (PPP) of each sign. PPP will always exceed the base rate and the observed PPP depends a great deal on it. Thus, the probability that a driver who repeatedly swerves is intoxicated may be 80% when the base rate is 50%, but only 8% when the base rate is 5%.

Stuster (1997) conducted a similar series of studies, but unlike the work of Harris (1980), all drivers in the exploratory study showing any indication of driving impairment were asked to provide a breath sample. In the subsequent preliminary and validation studies, data collection forms completed by the police provided the data. Although instructed otherwise, it is likely that such forms were only generated when an officer observed sufficient cause to make a stop, thus creating the same likely bias as in the work of Harris. Stuster provided PPP values for various signs, including those that do not involve driving and would be observed in the course of a preliminary DWI investigation, such as slurred speech or poor balance. Generally, these values were high and exceeded the base rate, which varied from 2.4 to 36% across the various data pools, and were surprisingly consistent given a 15-fold difference in base rate across studies. Unfortunately, other diagnostic statistics (e.g., sensitivity, false positive rates) were not reported. When calculated, diagnosticity indices from the preliminary and validation studies exceeded those from other studies investigating judgments of intoxication (including the data in the work of Stuster collected by researchers rather than police) by a factor of 10–100. False positives were virtually non-existent. In addition, the primary criterion in the validation study, *DWI arrest*, was problematic for several reasons. Some jurisdictions in the Stuster study enforced a strict no-alcohol law for underage drivers.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

Further, the arrest decision reflected the summary judgment of the officer, not BAC, and the predictors undoubtedly influenced the arrest decision. Because the values in the latter two studies appear to be outliers, they will not be cited in the following sections on individual cues of intoxication.

BEHAVIORAL/OBSERVABLE INDICATIONS OF INTOXICATION

This section will review the empirical evidence for various single signs or symptoms of alcohol intoxication, focusing on those that have been the subject of three or more studies. Such signs are often relied upon by police officers investigating possible DUI events (National Highway Traffic Safety Administration, 2006). However, the reader should recall the limitations inherent in such studies (availability of collateral information, observations of other indications of intoxication, and skewed sample distributions), and realize that the figures reported may be affected by such factors.

Red Eyes

Redness of the eyes is one of the signs most frequently associated with drinking, and one that frequently appears on DUI arrest reports. Teplin & Lutz (1985) included this sign among their 11 indicators of intoxication after their analysis found it met all their inclusion requirements, including substantial correlation with BAC, correlation with the total scale, frequency, and substantial loading on the first factor for their instrument. McKnight et al. (1997) reported that red eyes first appeared at an average BAC of .065% among nearly a thousand social drinkers in natural environments, although as discussed earlier there was a wide range among individuals. McKnight et al. (1999) reported a small correlation ($r = .19$) of red eyes with BAC among drinkers dosed to achieve .00%–.12% in a laboratory setting. McKnight et al. (2002) reported that having red eyes correlated .42 with BAC in the laboratory portion of the study ($N = 35$), but that this correlation was obtained entirely below .04% BAC. Thus, in this study red eyes appeared to be one of the few sensitive indictors of modest alcohol consumption. However, this finding appears to contradict the earlier findings of McKnight et al. (1997) based on a much larger sample. Holt et al. (1980) reported that red eyes was an effective diagnostic sign for distinguishing those who had consumed alcohol vs. those who had not (sensitivity = .67, false positive rate = .04), producing a likelihood ratio (LR) of 18.6 (LR = sensitivity / false positive rate). For patients above or below .08%, the LR was reduced (LR = 11.2), suggesting this cue functioned better as an indicator of alcohol consumption than for intoxication. The sensitivity and false positive rate were .75 and .07, respectively. Finally, Stuster (1997) reviewed indictors with promise to detect low BACs, but chose not to include red eyes because he deemed the judgment too subjective.

In conclusion, there is considerable evidence that alcohol produces red eyes and most studies suggest that its incidence increases as BAC rises. However, the available data are conflicting about the BACs at which it is likely to first appear and to correlate with BAC. The study by McKnight, Langston, Marques, & Tippetts (1997) appears to be the most reliable guide at present.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

Odor of Alcohol (OOA)

Alcohol itself has no odor, and the means by which observers conclude that someone has alcohol on their breath is not understood. Bogen (1928) reported that, among 250 persons suspected of acute alcohol intoxication, the percentage with OOA increased from 40 to 82% as the amount of alcohol in 1 cc of urine increased from 0 mg to 1 mg (the reason for the high rate of positive findings at 0 mg is not discussed), with 100% of subjects having OOA at 5 mg per cc of urine. Widmark (1932) reported that police physicians did not observe OOA in subjects with BACs below .06%. Proportions of suspects with OOA increased as BAC increased linearly with BAC, but did not reach 100% until subjects were over .261% BAC. Compton (1985) reported that 7% of drivers at .00% BAC were positive for OOA, while only 39% of drivers between .05 and .09% BAC and 61% of those between .10 and .15% were. Variability among officers was reported to be large, but no formal interrater reliability figures were reported. Holt et al. (1980) reported that OOA was the "most useful" clue they examined and it produced impressive diagnostic statistics (sensitivity = .77, false positive rate = .01, LR = 64.3) for separating those who had consumed any alcohol vs. those that had not. For those above vs. below .08%, the corresponding figures were sensitivity = .86, false positive rate = .06, and LR = 14.5. Stuster (1997) reported that 74% of drivers stopped in their initial (ride along) study who had OOA were above a BAC of .08%, slightly more than double the base rate in the study (36%).

Odor of alcohol also survived the rigorous criteria imposed by authors of the Alcohol Symptom Checklist, and it was reported to correlate .79 with BAC, the highest of any cue examined. However, the large percentage of completely sober subjects may have inflated the correlation observed, and in field DUI assessment the distinction will usually be between those who have drunk some alcohol and those who have drunk too much.

Moskowitz, Burns, & Ferguson (1999) carried out the most rigorous study on this topic. Drinking subjects were physically separated from raters by a large curtain. Subjects were instructed to blow through a plastic tube that passed through a slit in the curtain so that visual observations of behavior or physical appearance could not influence their judgments. Although the authors assured that "the only cue presented to the officer would be odor" (p. 176), it appears from the design description that sounds and time of day could have provided potential cues. Fourteen subjects were tested in four repeated trials over a four hour period, and 20 Los Angles police officers, all NHTSA-trained Drug Recognition Experts,[2] served as judges. Subjects were tested 30 minutes after finishing their drinks. Officers were asked to judge whether the odor of alcohol was present or not, and if so, its intensity. The authors reported that a chi square test of the relation between BAC and odor strength did not reach statistical significance. However, re-analysis for this paper with the Mantel–Haenszel chi square linear-by-linear association test, which is appropriate for tables with ordered categories, did show a significant association (5.468, $df = 1, p = .019$).[3] When OOA was judged "strong," 30 of 33 such subjects had BACs over .08%, while no subject produced a strong OOA with a BAC below .04%. Strong OOA identified 23% of subjects over .08% BAC while the

[2] The use of the term Drug Recognition Expert is not meant to convey support for the DRE program, as much of the research supporting it has not been published in peer-reviewed journals and the accuracy levels reported appear to depend very much on the subject openly disclosing to the officer what drugs were consumed.
[3] Scott Millis consulted on review of this data and performed this analysis.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

false positive rate was 6%, producing a quite respectable LR of 3.95. "Moderate" OOA had no diagnostic value. Of 164 observations at .00% BAC, 26 (15.9%) were false positives and an additional 18 (11%) of observations produced an "uncertain" judgment. Overall accuracy rate for all subjects fell to 55% in the third observation session, after many subjects consumed a lunch consisting of pizza, salad, and corn chips. Officers were completely unable to determine beverage type, which included beer, wine, vodka, and bourbon. Given this finding and the fact alcohol itself has no odor, the authors suggested that the perception of odor of alcohol may be due to a metabolite of alcohol, a conclusion dismissed by at least one toxicologist (J. Booker, personal communication, 2007).

In sum, a number of studies have reported a substantial correlation with BAC and the relative absence of detectable OOA at low BAC levels, notwithstanding the lack of a satisfactory scientific explanation. This clue deserves further investigation, but is limited by interference from food consumption and the fact that the officer's assessment cannot be recorded for independent evaluation by the trier-of-fact.

Distortions of Speech

Research on the effect of alcohol on speech was recently reviewed by Cutler and Henton (2004), who summarized

> All researchers report that the most noticeable difference is a slower rate of speech in the intoxicated (Lester & Skousen, 1974; Johnson, Pisoni & Bernacki, 1990; Behne, Rivera & Pisoni, 1991; Hollien & Martin, 1996; Chin, Large & Pisoni, 1997). Fundamental frequency (pitch) has been reported to rise with intoxication (Hollien & Martin, 1996), but also simply to become more variable (Behne & Rivera, 1990; Chin, Large & Pisoni, 1997). Differences in voice quality after drinking are also noticeable to trained observers (Künzel, Braun & Eysholdt, 1992), but again, individual differences are so great that it is not possible to predict the specific effects of alcohol on phonation. Place of articulation may be affected, with more posterior articulations being favored (Behne & Rivera, 1990), presumably because opening the mouth and articulating clearly demands more energy than an intoxicated talker can muster (p. 38).

Bogen (1928) reported that the incidence of *slurred speech* and *confused speech* actually decreased as blood urine levels went from 0 to 5 mg/cc, but this was confounded by the increase in the percentage of subjects who were *unable to speak* at levels of 4 and 5 mg/cc. However, no clear increase in prevalence of slurred speech was observed from 0 to 3 mg/cc either. Widmark (1981) reported that no subjects showed *stammering speech* below .10%, with an approximately linear increase from this BAC level to .32%. Penttilä et al. (1974) reported that *quality of speech* had very little correlation with BAC level below .15%, but did correlate .36 with BAC among all subjects. Holt et al. (1980) concluded that slurred speech was not particularly helpful in assessing intoxication in ER patients. Nonetheless, this cue performed well in separating those above and below .08%, with a LR of 11.6. Sensitivity was moderate at .59 with a false positive rate of only .05, but most subjects defined as sober had BACs of 0.00%. Stuster (1997) reported that 89% of drivers stopped for suspicion of DWI who displayed slurred speech exceeded .08%, compared with 36% of the total sample. Perham et al. (2007) reported that men who showed slurred speech had an average BAC of .135%, whereas women who had slurred speech had an average BAC of .161%. McKnight et al. (1997) reported a mean onset of slurred speech at .085% in a large, natural setting sample, while Teplin & Lutz (1985) found that this cue met all their inclusion criteria for the ASC.

Copyright © 2010 John Wiley & Sons, Ltd.

Klingholz, Penning, and Liebhardt (1988) examined the ability of an automated statistical program based on several objective measures of speech (e.g., fundamental frequency, signal-to-noise ratio) to classify 11 speakers at BAC levels from .05 to .15%. The authors reported high levels of accuracy (95–100%), but cautioned that these figures were dependent on having baseline measures, as individuals differed markedly. In another portion of the study, 12 speech therapists listened to 30 second samples of speech and made judgments about whether the speaker was intoxicated. When speech samples were presented in random order, the accuracy rate was only 54%. When listeners were given paired samples from the same subject (one while sober and one while intoxicated), accuracy rose to 61.1%. Accuracy was considerably better (82%) for subjects over .10% BAC.

Pisoni and Martin (1989) examined the ability of college students and police officers to detect speech abnormalities (primarily *slurring*) associated with alcohol consumption. Eight male university students served as speakers. They were directed to read 34 sentences chosen to contain one or two key words likely to be compromised by alcohol and constructed to represent a range of articulation difficulty. Each speaker was recorded speaking the sentences when sober and after dosing with alcohol. Drinking speakers were at .10–.19% BAC when tested. Twenty-one university students served as raters in the first experiment and were asked to choose which of two sentences was spoken under intoxication. A mean of 73.8% accuracy was observed across all speakers and raters. The raters' confidence in their assessments was reported to be related to their accuracy, but no statistics were presented. In the second study, raters heard and judged 24 of the original 34 sentences, one at a time, rather than making a comparison between sober and intoxicated conditions as in the first study. Fourteen Indiana State Troopers and 30 students served as raters. The police officers were significantly more accurate than students ($d' = .79$ versus .60) but accuracy rates were low for both groups (64.7% and 61.5%, respectively). Even when raters expressed the highest levels of confidence ("5" on a 1–5 scale where "5" was labeled "most confident"), they were not more than 75% accurate. The authors also examined acoustic waveforms from sober and intoxicated speech samples and reported that differences could be detected in a number of speech variables, including phonemes, pitch, and speed.

In sum, there are clear *group* differences in speech variables intoxicated and sober speakers. However, judgments about individuals, particularly at low BAC levels and without knowledge of their sober speech parameters, are likely to be of modest accuracy and subject to considerable overconfidence.

Impaired Walking/Gait

Before the development of formal sobriety tests such as the Walk and Turn, subjects in intoxication studies were sometimes instructed to simply walk normally across a room. *Staggered gait* did not show a clear relationship to urine alcohol concentration in the study by Bogen (1928), but Widmark (1981) reported a more or less linear increase in incidence of impaired subjects from .08% (6%) to .34% BAC (100%). Penttilä et al. (1974) reported that *walking along a line*, *walking with eyes open*, *walking with eyes closed*, and *gait in turning* all showed little correlation with BAC under .151% ($r = .10–.24$), but substantial correlations when all cases (BAC .00 to >.24%) were considered ($r = .43–.55$). Although Teplin and Lutz did not specifically examine walking, they reported that

Copyright © 2010 John Wiley & Sons, Ltd.
DOI: 10.1002/bsl

impairment of *gross motor control* met all their criteria for items in the ASC, while McKnight et al. (1997) reported that the item *clumsy* had an average BAC at onset of .074%. Perham et al. (2007) reported a much higher average onset BAC value for *staggering gait* (.186%), possibly because more severe impairment was required to score this cue than in the other studies.

In sum, impaired walking is generally related to increasing levels of BAC, although, depending on stringency of scoring, impairment may not be apparent until subjects substantially exceed current legal limits.

SOBRIETY TESTS

Two large studies have examined multiple, formal sobriety tests with the goal of creating a battery for use in law enforcement. Penttilä et al. (1974) studied 495 suspected drunk drivers in Finland who were examined by physicians following detention. In some European countries, there is no statutory BAC limit and judgments of intoxication are made by physicians based on a physical examination. The authors conducted extensive statistical analyses but did not report interrater reliabilities. They did report that more objective tests and indicators performed better, and recommended several possible sobriety test batteries based on multiple regression analysis. With the exception of a *collecting small objects* (on the ground) test for subjects under .126%, only nystagmus tests were useful for subjects under .15%.

Burns and Moskowitz (1977) examined 15 candidate sobriety tests, although most of these were subject to only rational analysis or pilot study. Six were selected for formal evaluation among 238 volunteer, drinking subjects. They were evaluated by police officers and research assistants. Based on multiple regression, a Standardized Field Sobriety Test (SFST) battery was created based on the Walk and Turn, One Leg Stand, and Alcohol Gaze Nystagmus tests. These three tests were reported to yield judgments as accurate as the six tests. Subsequently, Tharp, Burns, and Moskowitz (1981) refined the three tests and added data on their interrater reliability and relation to BAC.

Rubenzer (2007/2008) exhaustively reviewed the literature on the NHTSA SFSTs, which have been the subject of considerable research, although much of it is unpublished. Although most such studies provide superficially supportive evidence, Rubenzer noted that none of the studies were conducted truly blind, that potentially confounding factors such as fatigue, anxiety, and age had not been investigated, and that there is a lack of adequate, broadly based norms. While these tests have undergone much more investigation and development than other tests in the literature, all have utilized BAC as the criterion, not impaired ability to drive. When introduced as evidence in trial, sobriety tests are used to establish mental or physical impairment, and there is little empirical evidence to support this use. In the next paragraph, findings from studies that used BAC as a criterion are summarized.

The *Horizontal Gaze Nystagmus* test, which looks for jerky eye movements as the subject follows a smoothly moving stimulus or fixates on a stimulus displaced from the midline of the face, correlated best with BAC, averaging .65 across nine studies (range .51–.77). For .08% BAC, an average sensitivity of .88 (range .79–.98) was observed and an average LR of 3.6 (range 2.3–6.6) based on six studies. The average false positive rate was approximately .28 (range .13–.37). However, a recent study by one of the test's creators, not included in the previous analysis, found a false positive rate

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

of .67, which remained fairly constant across variations in test administration (Burns, 2007). Rubenzer and Stevenson (2010) reviewed vision science and application issues that are relevant to HGN's use in field sobriety testing.

Walk and Turn performance also correlated substantially with BAC (mean $r = .55$, range .37–.61, four studies), but it appears that its cut-off score (two or more "clues") is set too low, particularly for older, heavier, and physically inactive or compromised subjects. Only two, unpublished, studies (McKnight and Langston, 1993; McKnight & Langston, 1993, cited by McKnight, Langston, Lange, & McKnight, 1995; Stuster & Burns, 1998) reported diagnostic statistics at .08% BAC, yielding a modest mean LR of 1.9 and false positive rate of .37.

The third SFST, *One Leg Stand*, had a lower average correlation with BAC ($r = .45$, range = .16–.60, six studies) than Walk and Turn, but surprisingly strong diagnostic discriminative power at .08% (LR = 3.7) and .10% (LR = 4.3). Like Walk and Turn, there are extremely limited data available at .08%. Based on two studies, average sensitivity was .69 and the false positive rate .25. Data from Burns & Dioquino (1997) suggest a substantial false positive rate, but incomplete reporting prevents a precise analysis.

The following section reviews the limited empirical evidence for non-SFST sobriety tests that have been the subject of significant empirical evaluation.

The Romberg/Rhomberg

The Romberg test was developed in 1840 as a test of ataxia. The Romberg sign was positive if the person performed adequately with eyes open, but poorly with eyes closed. It was designed as a test of the person's proprioception (sense of body position and movement), and not a test of cerebellar function (Khasnis and Gokula (2003). The cerebellum is involved in much coordinated large muscle movement and is one of the areas most affected by moderate doses of alcohol. In the standard Romberg, the subject stands with his feet together, arms at the side, with the head tilted back and eyes closed. Indications of impaired balance, such as swaying or falling, are observed. In some variations, the subject is also asked to estimate the passage of 30 seconds while performing the physical portion of the test.

The first evaluation of the Romberg as a sobriety test located was the investigation of Bogen (1928). He did not directly report findings from the Romberg, but reported that *sway while standing* showed an erratic pattern across five levels of urine alcohol level. Widmark (1981) used a variation with the subject placing one foot in front of the other. He found that 17% of subjects failed at .08% or below, while the percentage rose steadily as BAC rose to .32%. Penner and Coldwell (1958), on the other hand, reported virtually no correlation of Romberg performance with BAC. Penttilä et al. (1974) found the Romberg with eyes open to be the best sobriety test when the entire BAC range (.00–.24 + %) of subjects was considered ($r = .59$), but only a modest correlation with BAC for subjects under .15% ($r = .31$). Findings for the Romberg with eyes closed were slightly lower. In the 1977 NHTSA laboratory study (Burns & Moskowitz, 1977), the Romberg was identified as one of the most promising tests in the second pilot study on 30 subjects, but was not examined in the main part of the study. Seidl, Müller, and Reinhardt (1994) reported that the Romberg was the most sensitive test of balance problems examined, and in contrast to most previous findings showed sensitivity at low

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

BAC levels. Stuster (1997) reported that 81% of drivers suspected of DWI who showed swaying and balance problems (but who were not formally assessed on the Romberg) had BACs above .08%, compared with 36% of the total sample.

A major limitation of the Romberg when used as a police sobriety test is the lack of a standardized scoring protocol. Most early studies reported no information at all on how it was scored, while Burns and Moskotwitz (1977) used vertical stripes on the wall of the testing room to help estimate degree of body sway. However, the report is silent about how wide the strips were and precisely how they were used. Scoring of the Romberg has been automated and computerized in laboratory applications since the 1980s (Bhattacharya, Morgan, Shukla, Ramakrishanan, & Wang, 1987; Black, Wall, Rockette, & Kitch, 1982). While some researchers have devised means of objectively quantifying body sway, others qualify the shift in body's center of gravity. Bhattacharya et al. (1987) reported that a laboratory balance test was sensitive to low levels of alcohol (.015% and .03%) for females, but not males. While some of these studies (e.g., Black et al.) provide basic norms, they are expressed in terms of electronically recorded mean displacement of the center of gravity and are of no use in interpreting performance on the traditional Romberg used as a sobriety test. In a review of studies that examined alcohol and body sway, Moskowitz (2000) concluded that body sway was apparent in 50% or more of subjects who exceeded .06%, but it appears that the review only utilized studies that found impairment, seemingly introducing a selection bias.

The Romberg is used occasionally as an outcome measure in tests of drug effects (e.g., Bramness, Skurtveit, & Mørland, 2003), and a variation was investigated as a measure the effects of marijuana by Papfotiou (2001). It is also part of the Drug Recognition Program, where the subject's accuracy in estimating thirty seconds is scored rather than body sway. While some research has examined the effect of alcohol and marijuana on estimation of time or distance traveled in a car, the procedures used are substantially different than used in sobriety testing and the affects of alcohol are not robust (Bech, Rafaelsn, & Rafaelsen, 1973; Tinklenberg, Kopell, Melges, & Hollister, 1972; Tinklenberg, Roth, & Kopell, 1976).

In sum, there is no one Romberg test, as some of the variations completely alter the nature of the test. The standard Romberg, as described above and used as a measure of balance, is affected by alcohol, and performance tends to show deterioration as BACs exceed moderate levels. However, findings are inconsistent about its sensitivity to low BAC levels, and there is a lack of adequate standardization and norms. Time estimation is not a part of the traditional Romberg test and there is very little evidence to support its use in assessing sobriety.

Finger to Nose

Like the Romberg, *Finger to Nose* (FN) is used in some studies of drug effects. Formal evaluations are sparse, and there is considerable variation in scoring across studies. Burns and Moskowitz (1977) conducted a pilot study with 30 subjects, and found an increase from less than one error per subject at .00% BAC to an average of 4.05 errors at .10%. FN correlated .42 with BAC in the main study. Sussman et al. (1990) found that adding FN and *Finger to Finger* to previous observations and tests actually decreased accuracy, dropping the correlation with BAC from $r = .446$ to .414. McKnight et al. (1999) reported observing a trivial correlation of $r = .05$ with BAC

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

in the laboratory portion of that study. Papfotiou (2001) found FN somewhat sensitive to marijuana, but scored six aspects of the test other than accuracy touching the nose, and did not score this latter cue.

Finger to Finger (FF)

The prototypical FF is performed with the thumb touching each finger tip, beginning with the index finger and proceeding to the little finger and then back. However, there are a number of similar tests and they should not be equated. As with the Romberg and FN, FF is used in drug toxicity studies and as an informal neurological test.

Widmark (1981) did not specify how his version of FF was performed, but reported that errors showed a linear increase with BAC up to .32% BAC, with no false positives below .08%. However, sensitivity remained below .52 until .18% was exceeded. Penttilä et al. (1974) reported that a different finger touching task[4] showed little correlation with BAC under .151% or over .24% BAC, although there was a modest correlation ($r = .36$) for subjects across the full range of BAC. Burns and Moskowitz (1977) reported that sober subjects in the pilot study averaged .60 error points, while those at .10% BAC showed 4.60. Although FF was apparently examined in the main part of the Burns and Moskowitz study, its correlation with BAC is mysteriously missing from the results reported. In addition, the instructions to give this test appear incorrect: the subject is instructed to touch the thumb to each finger, counting 1–2–3–4–5–5–4–3–2–1. Since the hand only has four fingers, this description would appear to be in error.

Teplin & Lutz (1985) reported that the ability presumably underlying FF, *fine motor control*, met all the inclusion criteria for the ASC. The authors do not report what procedures were used to assess fine motor control. Sussman et al. (1990), cited in the previous section, reported that FF and FN together actually decreased accuracy of BAC estimations from other observations. McKnight et al. (1999) found that FF errors correlated .21 with BAC in the laboratory study (subjects dosed to a maximum of .12%), with similar or lower correlations for other tasks involving the fingers. McKnight et al. (2002) found that two more complicated finger sobriety tests (*finger sequences* and *finger spelling*) were not significantly related to BAC, although subjects in this study were dosed only to .08%, resulting in a quite restricted range.

Saying the Alphabet

This venerable test has been the subject of few formal investigations. Sussman et al. (1990) reported that, when administered along with Hand Pat (see description below), observers' accuracy (correlation with BAC) increased from those based just on informal observations ($r = .387$ to .446). However, the difference is not large and may be attributable to Hand Pat. McKnight et al. (1999) reported a small correlation of *Alphabet* performance with BAC in the laboratory study ($r = .20$), but subjects here were all under .12%. In the field study, saying the alphabet showed a large effect size of $d' = 1.40$, with a sensitivity of .77 and specificity of .745 (false positive rate = .255).

[4] Subjects were asked to touch the tips of their two index fingers together.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

Hand Pat

Hand Pat consists of "clapping the hands alternately with palm and back of hand, while counting one–two" (McKnight et al., 1999, p. 149). Formal studies are very limited. As reported above, Sussman et al. found that the combination of Hand Pat and Saying the Alphabet showed incremental validity over other sobriety tests. McKnight et al. found Hand Pat to correlate only .17 with BAC in the laboratory portion of their study, although subjects were limited to BACs below .12%. In the field study, HP performed quite well (sensitivity = .923; specificity = .723; $d' = 2.0$). However, it did not lead to an improvement in predicting BAC beyond what could be achieved by HGN alone. It was nonetheless recommended by the authors because it has substantial face validity as a measure of impairment, whereas HGN does not.

SUMMARY

This review finds that judging low to moderate levels of intoxication in strangers is a difficult task. A variety of professions that might be expected to show substantial skill assessing intoxication do not. No behavioral or physical sign has emerged that is consistently related to a specific level of BAC without large variation among individuals, with the possible exception of nystagmus. Acquired tolerance has confounded attempts to create observational scales, despite the sophisticated approach of Teplin & Lutz (1985), and led to paradoxical findings: Alcoholics can appear unimpaired at BACs that could be fatal to many drinkers (Perper et al., 1986; Sullivan et al., 1987; Urso et al., 1981), yet appear more impaired than social drinkers at low BACs (Burns & Moskowitz, 1977; Cherpital et al., 2005; Honkanen, 1977). More attention should be paid to the difficulties of assessing intoxication in this high risk group. Finally, no study to date has examined whether individuals show a characteristic pattern of alcohol-related symptoms. Nearly all studies to date have referenced signs of intoxication to BAC, usually as estimated by breath test, not to behavioral or physical impairment.

The non-SFST sobriety tests reviewed were all marked by serious problems. The Romberg and Finger to Nose have markedly different administration and scoring schemes across studies. All of the sobriety tests have been the subject of remarkably few published, peer-reviewed studies, and most show weak correlations with BACs below .15%. None have been validated as measures of impairment or have even rudimentary norms. Among the tests reviewed, Hand Pat and Saying the Alphabet appear promising but require considerable further research. Like virtually all sobriety tests other than nystagmus, they show substantial correlations with BAC only across a large range of BAC and do not appear able to gauge BAC levels below .10%.

Several suggestions can be made for future research based on the findings reported above. Since sobriety tests and individual behavioral signs will likely be cited as evidence of intoxication in court, they should be validated against established measures of cognitive, physical, or driving impairment as well as BAC. Data for drinkers who have developed tolerance should be collected and analyzed separately from those of social drinkers. Complete diagnostic statistics should be reported for multiple BAC levels (e.g., .04%, .08%) and impairment criteria (e.g., mild, moderate, severe) and for varying base rates. Correlations among various indicators should be reported and, where sample size permits, factor analyses and tests of incremental validity performed.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

Researchers should also consider evidence that alcohol intoxication may be multifactorial (Mundt, Perrine, & Searles, 1997; Simpson-Crawford & Slater, 1971), and it may be premature to focus on a single dimension. Finally, none of the studies establishing correlations between physical and behavioral signs have been conducted blind, so other observations may bias observations and inflate the correlations among the signs and with BAC.

Advocates of lower legal BAC limits argue that even very low levels of alcohol substantially impair driving and the abilities that underlie driving (Moskowitz, 2000). If this proposition is accurate, it should be possible to observe such impairment in stressful and demanding social situations (investigations of DUI) and on sobriety tests. So far, there is little evidence that most existing sobriety tests are up to the task.

REFERENCES

Bech, P., Rafaelsen, L., & Rafaelsen, O. J. (1973). Cannabis and alcohol: Effects on estimation of time and distance. *Psychopharmacologia*, *32*, 373–381.

Behne, D. M., & Rivera, S. M. (1990). Effects of alcohol on speech: Acoustic analysis of spondees. *Research on Speech Perception*, Progress Report, 16 (pp. 263–291). Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Behne, D. M., Rivera, S. M., & Pisoni, D. B. (1991). Effects of alcohol on speech: Durations of isolated words, sentences, and passages in fluent speech. *Research on Speech Perception*, Progress Report, 17 (pp. 285–301). Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Bhattacharya, A., Morgan, R., Shukla, R., Ramakrishana, H. D., & Wang, L. (1987). Non-invasive estimation of afferent inputs for postural stability under low levels of alcohol. *Annals of Biomedical Engineering*, *15*, 533–550.

Black, F. O., Wall, C., Rockette, H. E., & Kitch, R. (1982). Normal subject postural sway during the Romberg Test. *American Journal of Otolaryngology*, *3*, 309–318.

Bogen, E. (1927a). The diagnosis of drunkenness—A quantitative study of acute alcoholic intoxication. *California and Western Medicine*, *26*(6), 778–783.

Bogen, E. (1927b). Drunkenness: A quantitative study of acute alcoholic intoxication. *Journal of the American Medical Association*, *89*(18), 1508–1511.

Bogen, E. (1928). Drunkenness: A quantitative study of acute alcoholic intoxication. *American Journal of the Medical Sciences*, *176*(2), 153–167.

Bois, C., & Vogel-Sprott, M. (1974). Discrimination of low blood alcohol in self-titration skills in social drinkers. *Quarterly Journal of Alcohol*, *35*, 86–97.

Bramness, J. G., Skurtveit, S., & Mørland, J. (2003). Testing for benzodiazepine inebriation—Relationship between benzodiazepine concentration and simple clinical tests for impairment in a sample of drugged drivers. *European Journal of Clinical Pharmacology*, *59*(8/9), 593–601.

Brick, J., & Carpenter, J. A. (2001). The identification of alcohol intoxication by police. *Alcoholism: Clinical and Experimental Research*, *25*(6), 850–855.

Burns, E. D., Nusbaumer, M. R., & Reiling, D. M. (1993). Think they're drunk?: Alcohol servers and the identification of intoxication. *Journal of Drug Education*, *33*(2), 177–186.

Burns, M. (2007). *The robustness of the horizontal gaze nystagmus test* (Tech. Rep.). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Burns, M., & Dioquino, T., (1997). *A Florida validation study of the standardized field sobriety test (SFST) battery*. Jacksonville, FL: Florida Department of Health.

Burns, M., & Moskowitz, H. (1977). *Psychophysiological tests for DWI arrest* (DOT-HS-802-424). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Carroll, N., Rosenberg, H., & Funke, S. (1988). Recognition of intoxication by alcohol counselors. *Journal of Substance Abuse Treatment*, *5*(4), 239–246.

Cherpitel, C., Bond, J., Ye, Y., Room, R., Poznyak, V., Rehm, J., & Peden, M. (2005). Clinical assessment compared with breathalyser readings in the emergency room: Concordance of ICD-10 Y90 and Y91 codes. *Emergency Medicine Journal*, *22*, 689–695.

Chin, S. B., Large, N. R., & Pisoni, D. B. (1997). Effects of alcohol on words in context: A first report. In *Research on spoken language processing* (Progress Rep. 21) (pp. 403-420). Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Compton, R. P. (1985). *Pilot test of selected DWI detection procedures for use at sobriety checkpoints* (DOT-HS-806-724). Washington, DC: National Highway Traffic Safety Administration.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

Cutler, A., & Henton, C. G. (2004). There's many a slip 'twixt the cup and the lip. In H. Quené & V. van Heuven (Eds.), *On speech and language: Studies for Sieb G. Nooteboom* (pp. 37–46). Utrecht, The Netherlands: Netherlands Graduate School of Linguistics (LOT).
Grossman, D. C., Mueller, B. A., Kenaston, T., Salzberg, P., Cooper, W., & Jurkovich, G. J. (1996). The validity of police assessment of driver intoxication in motor vehicle crashes leading to hospitalization. *Accident Analysis and Prevention*, *28*(4), 435–442.
Harris, D. H. (1980). Visual detection of driving while intoxicated. *Human Factors*, *22*(6), 725–732.
Hollien, H., & Martin, C. A. (1996). Conducting research on the effects of intoxication on speech. *Forensic Linguistics: The International Journal of Speech, Language and the Law*, *3*, 107–127.
Holloway, F. A. (1994). *Low-dose alcohol effects on human behavior and performance: A review of the post-1984 research* (Medicine Technical Report DOT/FAA/AM-94/24). Washington, DC: Department of Transportation, Federal Aviation Administration, Office of Aviation.
Holloway, F. A. (1995). Low-dose alcohol effects on human behavior and performance. *Alcohol, Drugs and Driving*, *11*(1), 39–56.
Holt, S., Stewart, I. C., Dixon, J. M., Elton, R. A., Taylor, T. V., & Little, K. (1980). Alcohol and the emergency service patient. *British Medical Journal*, *281*, 638–640.
Honkanen, R. (1977). Records based on clinical examination as an indicator of alcohol involvement in injuries at emergency stations. *Scandinavian Journal of Social Medicine*, *5*, 91–95.
Johnson, K., Pisoni, D. B., & Bernacki, R. H. (1990). Do voice recordings reveal whether a person is intoxicated? A case study. *Phonetica*, *47*, 215–237.
Khasnis, A., & Gokula, R. (2003). Romberg's test. *Journal of Postgraduate Medicine*, *49*(2), 169–172.
Klingholz, F., Penning, R., & Liebhardt, E. (1988). Recognition of low-level alcohol intoxication from the speech signal. *Journal of the Acoustical Society of America*, *84*, 929–935.
Kruger, H.-P. (1993). Effects of low alcohol doses: A review of the literature. In: H. -P. Utselmann, G. Berghaus, G. Kroj, (Eds.) *Alcohol, drugs and traffic safety, T-'92: Proceedings of the 12th International Conference on Alcohol, Drugs, and Traffic Safety*. Cologne: Verlag TUV Rheinland.
Künzel, H. J., Braun, A., & Eysholdt, U. (1992). *Einfluss von alcohol auf sprache und stimme*. Heidelberg: Kriminalistik.
Langenbucher, J., & Nathan, P. E. (1983). Psychology, public policy, and the evidence for alcohol intoxication. *American Psychologist*, *38*(10), 1070–1077.
Lester, L., & Skousen, R. (1974). The phonology of drunkenness. In A. Bruck, R. Fox, & M. W. LaGaly (Eds.), *Papers from the parasession on natural phonology* (pp. 233–239). Chicago, IL: Chicago Linguistic Society.
Maisto, S. A., & Adesso, V. J. (1977). Effects of instructions and feedback on blood alcohol level discrimination training in nonalcoholic drinkers. *Journal of Consulting and Clinical Psychology*, *45*, 625–636.
McGuire, F. L. (1986). The accuracy of estimating the sobriety of drinking drivers. *Journal of Safety Research*, *17*, 81–85.
McKnight, A. J. (1988). *Development and field test of a responsible alcohol service program* (National Public Services Research Ansated Final Rep. Contract No. DTNH22-91-C-05109). Washington, DC: National Highway Traffic Safety Administration.
McKnight, A. J., Lange, J. E., & McKnight, A. S. (1999). Development of a standardized boating sobriety test. *Accident Analysis and Prevention*, *31*(12), 147–152.
McKnight, A. J., Langston, E. A., Lange, J. E., & McKnight, A. S. (1995). *Development of standardized field sobriety tests for lower BAC limits* (Report prepared under NHTSA Contract No. DTNH 22-92-C-00700). Washington, DC: US Department of Transportation, National Highway Traffic Safety Administration.
McKnight, A. J., Langston, E. A., Marques, P. R., & Tippetts, A. S. (1997). Estimating blood alcohol from observable signs. *Accident Analysis and Prevention*, *29*(2), 247–255.
McKnight, A. J., Langston, E. A., McKnight, A. S., & Lange, J. E. (2002). Sobriety tests for low alcohol blood concentrations. *Accident Analysis and Prevention*, *34*(3), 305–311.
McKnight, A. J., Langston, E. A., Schaefer, J. M., Aasved, M. J. (1991). Field validation of behavioral cues to detect alcohol intoxication (Final Report Contract No. DTNH22-89-07392). Washington, DC: National Highway Traffic Safety Administration.
McKnight, A. J., & Marques, P. R. (1990). *Host and server determination of alcohol intoxication level* (Final Rep.DOT HS 807 639). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
Moskowitz, H. (2000). *A review of the literature on the effects of low doses of alcohol on driving-related skills* (DOT HS 809 028). Washington, DC: National Highway Traffic Safety Administration.
Moskowitz, H., Burns, M., Ferguson, S. (1999). Police officers' detection of breath odors from alcohol ingestion. *Accident Analysis and Prevention*, *31*, 175–180.
Moskowitz, H., & Robinson, C. D. (1988). *Effects of low doses of alcohol on driving-related skills: A review of the evidence* (DOT HS 807 280). Washington, DC: National Highway Traffic Safety Administration.
Mundt, J. C., Perrine, M. W., & Searles, J. S. (1997). Individual differences in alcohol responsivity: Physiological, psychomotor and subjective response domains. *Journal of Studies on Alcohol*, *58*(2), 130–140.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl

National Highway Traffic Safety Administration, U.S., Dept. of Transportation. (1984/2006). *DWI detection and standardized field sobriety testing, participant manual* (HS 178 R2/00). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Pagano, M. R., & Taylor, S. P. (1979). Police perceptions of alcohol intoxication. *Journal of Applied Psychology*, *10*(2), 166–174.

Papfotiou, K. (2001). *Evaluation of efficacy of sobriety testing to detect blood cannabis content and impaired driving ability*. Doctoral dissertation. Retrieved November 9, 2007, from http://adt.lib.swin.edu.au/uploads/approved/adt-VSWT20050323.083420/public/02whole.pdf

Penner, D. W., & Coldwell, B. B. (1958). Car driving and alcohol consumption: Medical observations on an experiment. *Canadian Medical Association Journal*, *79*, 793–800.

Penttilä, A., Tenhu, M., & Kataia, M. (1974). Clinical examination for intoxication in cases of suspected drunken driving: II. *Reports from Talja, The Central Organization for Traffic Safety in Finland*, *15*, 1–78.

Perham, N., Moore, S. C., Shepherd, J., & Cusens, B. (2007). Identifying drunkenness in the night-time economy. *Addiction*, *102*(3), 377–380.

Perper, J. A., Twerski, A., & Wienand, J. W. (1986). Tolerance at high blood alcohol concentrations: A study of 110 cases and review of the literature. *Journal of Forensic Sciences*, *31*(1), 212–221.

Pisoni, D. B., & Martin, C. S. (1989). Effects of alcohol on the acoustic phonetic properties of speech: Perceptual and acoustic analysis. *Alcohol: Clinical and Experimental Research*, *13*(4), 577–587.

Rubenzer, S. J. (2007/2008). The Standardized Field Sobriety Tests: A review of scientific and legal issues. *Law and Human Behavior 32*(4), 293–313. DOI: 10.1007/s10979-007-9111-y.

Rubenzer, S. J., & Stevenson, S. (2010). Horizontal Gaze Nystagmus: A review of vision science and application issues. *Journal of Forensic Sciences*, *55*(2), 394–409.

Russ, N. W., Harwood, M. K., & Geller, E. S. (1986). Estimating alcohol impairment in the field: Implications for drunken driving. *Journal of Studies on Alcohol*, *47*(3), 237–240.

Rutherford, W. H. (1977). Diagnosis of alcohol ingestion in mild head injuries. *Lancet*, *1*, 1021–1023.

Seidl, S., Müller, M., & Reinhardt, G. (1994). Regulation of equilibrium in acute alcoholic intoxication with the craniocorpography method. *Blutalkohol*, *31*(3), 186–195.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *8*(2), 420–428.

Simpson-Crawford, T., & Slater, S. W. (1971). Eye signs in suspected drinking drivers: Clinical examination and relation to blood alcohol. *New Zealand Medical Journal*, *74*(471), 92–96.

Sobell, M. B., Sobell, L. C., & VanderSpek, R. (1979). Relationships among clinical judgment, self-report and breath analysis measures of intoxication in alcoholics. *Journal of Consulting and Clinical Psychology*, *47*, 204–206.

Stuster, J. (1997). *The detection of DWI at BAC's below 0.10* (Report prepared for the US Department of Transportation, Contract No. DTNH-22-93-C-08187). Santa Barbara, CA: Anacapa Sciences.

Sullivan, J. B., Hauptman, M., & Bronstein, A. C. (1987). Lack of observable intoxication in humans with high plasma alcohol concentrations. *Journal of Forensic Sciences*, *32*(6), 1660–1665.

Sussman, E. D., Needalman, A., & Mengert, P. H. (1990). *An experimental evaluation of a field sobriety test battery in the marine environment* (Report prepared under Contract No. DOT-CG-D-04-90). Washington, DC: US Department of Transportation, US Coast Guard.

Teplin, L. A., Lutz, G. W. (1985). Measuring alcohol intoxication: The development, reliability and validity of an observational instrument. *Journal of Studies on Alcohol*, *46*(6), 459–466.

Tharp, V., Burns, M., & Moskowitz, H. (1981). *Development and field test of psychophysical tests for DWI arrest* (DOT-HS-805-864). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Tinklenberg, J. R., Kopell, B. S., Melges, F. T., & Hollister, L. E. (1972). Marijuana and alcohol: Time production and memory functions. *Archives of General Psychiatry*, *27*, 812–815.

Tinklenberg, J. R., Roth, W. T., & Kopell, B. S. (1976). Marijuana and alcohol: Differential effects on time perception, heart rate, and subjective response. *Psychopharmacology*, *49*, 275–279.

Urso, T., Gavaler, J. S., & Van Thiel, D. H. (1981). Blood ethanol levels in sober alcohol users seen in an emergency room. *Life Sciences*, *23*, 1053–1056.

Widmark, E. M. P. (1981). *Principles and applications of medicolegal alcohol determination* (translation from the original 1932 German edition edited by R. C. Baselt) Davis, CA: Biomedical Publications.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/bsl